# Non-canonical roles of mammalian heterochromatin protein 1 (HP1) homologs

**Inauguraldissertation**

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

**Veronika Ostapcuk**

aus der Tschechischen Republik

Basel, 2018

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von

Prof. Dr. Marc Bühler und Prof. Dr. Wolf Reik.

Basel, den 14. November 2017

Prof. Dr. Martin Spiess

Dekan

“When I was twelve, one of my friends bet another friend a bag of sweets that I would never come to anything. I don't know if this bet was ever settled, and if so, which way it was decided."

*Stephen Hawking*

# *Table of contents*

# *Summary*

The human body is composed of hundreds of different tissues, all containing same genetic information, yet defined by unique gene expression patterns. In order to achieve this, multicellular organisms evolved regulatory mechanisms that go beyond the mere DNA sequence. Unlike in prokaryotes, where DNA is freely accessible to the transcriptional machinery, DNA in eukaryotic cells is wrapped around histone proteins, forming a structure called chromatin. Importantly, chromatin can be regulated by various post-translational modifications of the histone proteins. Such modifications are generally assumed to directly affect the compaction of chromatin and/or act as a recruitment platform for chromatin factors that relax (activate) or condense (repress) chromatin. One of the repressive histone marks, histone 3 lysine 9 methylation (H3K9me), is recognised by members of the HP1 family and/or by other proteins that have the unique ability to spread along chromatin, compacting it, and ultimately forming large inaccessible domains referred to as heterochromatin. Over the past decade, however, the view of HP1s as rigid silencers has been gradually challenged, as it was found that heterochromatic regions produce RNA, and that HP1s are highly mobile molecules. In addition, HP1 proteins were shown to associate with RNAs, and to also associate with chromatin lacking the H3K9me mark. These findings raised several fundamental questions: Is HP1 activity regulated by RNA? How are HP1 proteins recruited to sites lacking H3K9me, and what is their role at those sites?

One aim of my PhD project was to elucidate potential roles of RNA in modulating HP1 activity. For this, I used biochemical methods to dissect RNA binding properties of mammalian HP1 proteins *in vitro.* My work revealed that one of the HP1 homologs, HP1α, interacts with RNA when bound to H3K9me-marked nucleosomes. The physiological role of such interaction could be stabilisation of binding to heterochromatin, or alternatively, eviction from heterochromatin. The major goal of my PhD project, however, was to investigate the mechanism of HP1 recruitment to chromatin lacking the H3K9me mark. To do so, I was using mouse embryonic stem cells (mESCs) as a model system. I have exploited recent advances in genome editing/CRISPR-Cas9 to delete or endogenously tag individual HP1 homologs or various combinations thereof. Genome- and proteome-wide studies subsequently revealed a novel protein complex, which I dubbed "ChAHP". ChAHP contains two HP1 homologs (HP1β and/or HP1γ), the transcription factor Adnp, and the chromatin remodeller Chd4. In collaboration with the group of Nicolas Thomä, we have reconstituted

the ChAHP complex *in vitro* from insect cells and dissected the individual interactions. Together with my proteomics experiments in mESCs, this revealed Adnp as a core bridging module interacting with Chd4 and HP1s. Using ChIP-sequencing, we identified over 15 000 ChAHP-bound genomic sites. Importantly, these sites are devoid of H3K9me2 or H3K9me3. Instead, the complex is targeted via a highly conserved DNA motif recognized by Adnp, and deletion of Adnp or of the DNA motif depletes HP1 binding at the ChAHP sites. In addition, deletion of Adnp or HP1s leads to derepression of lineage-specifying genes bound by ChAHP. However, unlike in case of canonical HP1 silencing, which involves H3K9me and results in formation of a broad heterochromatic domain, ChAHP silencing occurs locally by restricting access to its sites. I propose that this prevents other regulators, including transcriptional activators, from accessing the corresponding DNA sites. Finally, my results provide first insights into the molecular mechanism of a disease that is associated with mutations in the *ADNP* gene, Helsmoortel-Van der Aa syndrome. Mutant Adnp found in Helsmoortel-Van der Aa patients fails to interact with HP1 proteins, and therefore cannot target HP1s to the chromatin. In summary, my work revealed that HP1 proteins can be recruited to genomic loci in a DNA sequence-specific, H3K9 methylation-independent, manner via an interaction with the transcription factor Adnp, and I demonstrated that H3K9 methylation, unlike in canonical silencing, is not required for repression of ChAHP target genes.

# *Introduction*

# 1. Milestones in the history of chromatin biology

All the information necessary for an organism to build, maintain and replicate itself is encoded in its DNA sequence, the genome. The rise in complexity of organisms was accompanied by the expansion of their genomes. The prokaryotic genome of *Escherichia coli* encodes approximately 4 000 genes in 4.6 mega base pairs (Mbp), the unicellular eukaryotic genome of *Schizosaccharomyces pombe* encodes nearly 5 000 genes in 13.8 Mbp, and the human genome encodes 20 000 to 25 000 genes in 3 200 Mbp (Blattner et al., 1997; Venter et al., 2001; Wood et al., 2002). The expansion in genome size, however, does not linearly correlate with an increase in the number of the genes. The protein coding genes of *S. pombe* occupy 60% of its genome, 57% if we exclude introns, but only about 25% of the human genome comprises protein-coding genes, with only 1.1% encoding exons. Higher eukaryotes, and in particular mammals, accumulated large amounts of non-coding DNA consisting mostly from simple repeats, pseudogenes, tandem repeats like centromeres and telomeres, and transposon-derived repeats, with the last class alone taking up more than 45% of the human genome (Lander et al., 2001). The evolutionary benefits and roles of these elements are still under investigation, but what is clear is that organisms with such complex genomes also had to evolve complex regulatory mechanisms to control transcriptional noise. Uncontrolled integration of transposons across the genome would be detrimental to genome stability and survival of the organism, and potentially the whole species. In addition to preventing unwanted transcription, multicellular organisms must regulate differential expression in individual cells during development, and subsequently preserve the acquired tissue-specific expression pattern. How can this be achieved without changing the sequence of DNA in individual cells and tissues? How can a single genotype give rise to different phenotypes?

In the 1940s, Conrad Waddington introduced the term epigenetics. He defined epigenetics as ‘‘the branch of biology which studies the causal interactions between genes and their products which bring the phenotype into being’’ (Waddington, 1942). In the original sense of his definition, epigenetics referred to all changes in environment and pathways that modulated the interpretation of the genotype, and therefore lead to the development of the phenotype, e.g. the process of development of the fertilized zygote into the mature organism. During the following decades however, the rapid development of biological techniques accompanied by an immense increase in the knowledge of molecular mechanisms regulating

gene expression, has led scientists to narrow down the original definition to "the study of mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in DNA sequence" (Riggs and Porter, 1996). This definition was further refined when an additional requirement was added – the induced change has to be self-propagated even in the absence of the initial signal in order to be considered epigenetic (Berger et al., 2009; Ptashne, 2007).

What are those epigenetic changes? In 1950 Ellen Stedman and Edgar Stedman isolated small basic proteins, histones, from the nuclei of red blood and liver cells, finding that the histones varied between the two cell types. They proposed that histones could act as gene suppressors with cell specific variation in their composition (Stedman and Stedman 1950). It took another 24 years to find that histones form an oligomer around which about 200 bp of DNA is bound (Kornberg and Thomas 1974). Based on this work, Roger Kornberg further hypothesized that a histone oligomer wrapped in DNA forms a repeating unit which is flexibly connected to the other units, like "beads on a string", and that this is a general mechanism of DNA storage in eukaryotic cells (Kornberg, 1974). One can easily envisage that histones may act as a steric barrier making DNA transcriptionally less accessible. However, where does the regulation come from? At the turn of the 21$^{st}$ century, there were many publications identifying covalent posttranslational modifications of histones. To date, over a hundred distinct modifications have been described, and it became clear that certain modifications correlate with actively transcribed regions while others with silent regions (review in Zhao and Garcia 2015). Nowadays, histone modifications and their corresponding modifiers and readers are widely referred to as "epigenetic". This, however, does not necessarily satisfy the above definitions of epigenetics by Arthur Riggs and his colleagues, since we often cannot distinguish between DNA-sequence-independent propagation of histone marks over cell divisions, and reestablishment of marks after cell division by a DNA-sequence-dependent machinery. Consequently, there is a wide array of opinions in the field about the correct use of the term "epigenetic". In 2007, Adrian Bird attempted to resolve this confusion by refining the definition of epigenetics to "the structural adaptation of chromosomal regions so as to register, signal or perpetuate altered activity states" (Bird, 2007). Nevertheless, many scientists, and amongst them Mark Ptashne, insist on the use of the term epigenetics only in the context of memory, arguing that histone modifications and their associated factors lack specificity, and as such, cannot be self-perpetuating (Ptashne, 2013). Ultimately, of course, it

is up to individual scientists to choose the definition that fits them best; this work is going to respect the one by Adrian Bird.

Another type of epigenetic mark, which seems to better satisfy all of the above definitions, is DNA methylation. It was first described in 1975 by two parallel works, both suggesting a repressive role of DNA methylation in X chromosome inactivation (Holliday, R. & Pugh, 1975; Riggs, 1975). The epigenetic transmission of DNA methylation across the cell divisions was confirmed shortly after by another study (Bird, 1978), and today it is one of the best documented epigenetic marks (reviewed in (Bird 2002; Jones 2012)). It was later realised, however, that another epigenetic regulator – non-coding RNA – is responsible for the initiation and spreading of X chromosome inactivation. Permanent silencing of one of the two X chromosomes occurs early during female mammalian development. It is initiated by the expression of a long non-coding RNA, X-inactive-specific transcript (Xist). The Xist RNA then binds the whole X chromosome from which it is transcribed, acting as a recruiter of the silencing machinery, with the final repressive state being permanently maintained by DNA methylation and histone-modification mediated silencing (Lee et al. 1996; Penny et al. 1996; for review Galupa and Heard 2015).

In summary, the establishment and maintenance of tissue specific gene expression patterns is achieved through the coordinated action of epigenetic regulators and DNA-sequence specific factors. It remains difficult, however, to make a general statement about the hierarchy of individual factors, as we are often faced with the "causality dilemma". The further chapters summarize the current knowledge of the field relevant to my PhD project.

## 2. Chromatin – structure and function

The genome of eukaryotes is organized into a dynamic structure called chromatin – a complex of DNA, proteins and RNA. There are several benefits of storing genetic information in the form of chromatin: it mediates the spatially efficient storage of DNA, creates a steric barrier for the transcriptional machinery, and can be differentially modified helping to impose a silent or active state over the corresponding DNA (Bannister and Kouzarides, 2011; Lorch et al., 1987). Traditionally, chromatin has been divided into two subtypes based on different cytological staining characteristics, euchromatin and heterochromatin. Euchromatin becomes less compact during interphase, and its staining appears light; heterochromatin stays tightly compacted and, consequently, its staining appears much darker (Fig. 1). Today we know that euchromatin consists of transcriptionally active genes and their regulatory elements, and heterochromatin consists of transcriptionally silent regions. Heterochromatin is further subdivided into permanently silent "constitutive heterochromatin", which mostly contains repetitive sequences, and temporarily silent "facultative heterochromatin", with mostly repressed genes and their regulatory elements. The constitutive heterochromatin plays a crucial role in genome stability and organization, and secures proper chromosome segregation during cell division. (reviewed in Trojer and Reinberg 2007)

**Figure 1. Electron micrograph of a bone marrow plasma cell.**

Two distinct regions can be distinguished in the nucleus of eukaryotic cells, the compact heterochromatin – dark regions, and accessible euchromatin – light regions. Source: UCSF, Office of Educational Technology, Cell Structure lab

These historical definitions, however, do not sufficiently satisfy our increasing knowledge of the molecular signatures of different chromatin environments, and alternatives were therefore proposed. For instance, Filion and colleagues performed genome-wide mapping of 53 chromatin-associated proteins in *Drosophila melanogaster*, and based on their analysis proposed five different chromatin types, three repressive and two active, each defined by a unique combination of chromatin factors (Filion et al., 2010).

The basic repeating unit of chromatin is a nucleosome. The nucleosome consists of eight histone proteins that form an octamer, around which 146 bp of DNA is wrapped in 1.65 turns. The octamer consists of the core histones H3, H4, H2A and H2B, which form one tetramer H3-H4 and two heterodimers H2A-H2B (Kornberg and Thomas, 1974; Luger et al., 1997) (Fig. 2). Comparative studies revealed that core histones are the most evolutionary conserved proteins in eukaryotes, with >90% sequence identity between the yeast and human orthologs (Malik and Henikoff, 2003). This is presumed to be due to the severe structural constraints associated with the assembly of histones into a nucleosome (Luger et al., 1997).

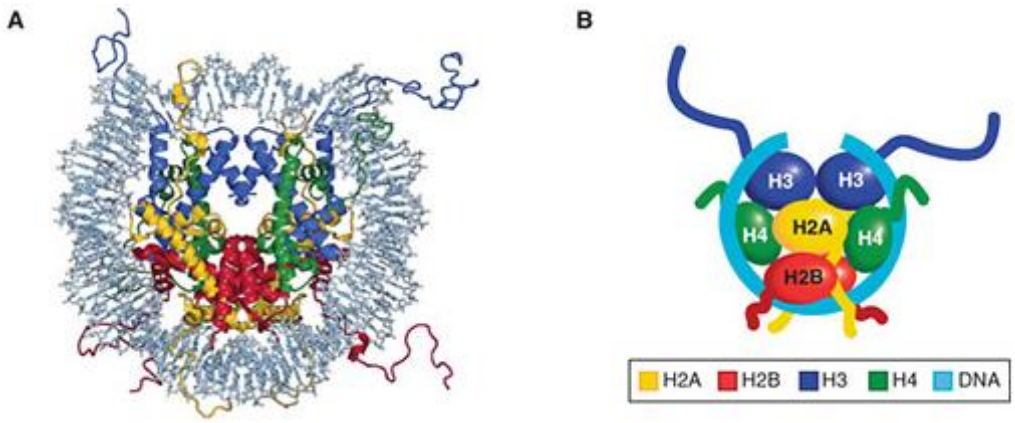



**Figure 2. The structure of the nucleosome.**

(A) A 2.8A structure of a core nucleosome reconstituted from *Xenopus laevis* solved by Karolin Luger and her colleagues (Luger et al. 1997).

(B) A schematic representation of a nucleosome depicting core octamer, consisting of H3-H4 tetramer and two H2A-H2B dimers, wrapped around 147bp of DNA. The globular domains of histones reside within the core of nucleosome, while N-terminal flexible tails of histones protrude from the surface of the nucleosome.

(A, B) Structural images were adapted from (Luger et al., 1997), and reprinted from (Allis et al., 2015) with the permission from Cold Spring Harbor Laboratory Press ©.

In the most open form of chromatin, individual nucleosomes are connected by 10-80 bp of free linker DNA, forming the aforementioned "beads on a string" fibre (Fig. 3). The individual nucleosomes can be brought closer together, and therefore make DNA less accessible and transcriptionally inactive, if the linker DNA is bound by histone H1, with the resulting structure called the "30 nm" fibre (Fig. 3; Robinson and Rhodes 2006; Song et al. 2014). The level of compaction is also strongly affected by histone modifications; charged modifications alter the accessibility by counteracting the negative charge of DNA, while other modifications create specific binding sites for effector proteins that have the ability to compact or relax the fibre (Kouzarides, 2007). The chromatin fibre can be further folded and coiled, creating so called higher-order chromatin structures (300-700 nm). This seems to be facilitated by intra- or inter-chromosomal looping, and anchoring chromatin to the nuclear membrane, possibly even in a physiologically relevant organized manner – clustering either the silent or active chromatin together (Cavalli and Misteli, 2013; van Steensel and Belmont, 2017). The highest level of chromatin compaction, in which the entire chromosomes are organized into compact structures visible under the light microscope, occurs only during cell division (Fig. 3).

Unlike core histones, histone H1 does not seem to display strong evolutionary conservation across eukaryotes. The fission yeast lacks H1 completely (Godde and Widom, 1992), and the budding yeast has only a weakly conserved homolog, Hho1p (31% identical to human H1) (Harshman et al., 2013; Patterton et al., 1998). Hhop1 is not essential and its deletion has no significant effect on global chromatin organization or expression (Hellauer et al., 2001; Patterton et al., 1998). In *Caenorhabditis elegans*, the single H1 homolog is already required for proper germline development and silencing of reporter genes (Jedrusik and Schulze, 2001). Finally, mammals have multiple H1 isoforms and the deletion of single genes has no significant phenotype, most likely due to partial redundancy, but deletion of at least three homologs severely impacts mouse development and chromatin organization (Fan et al., 2003).

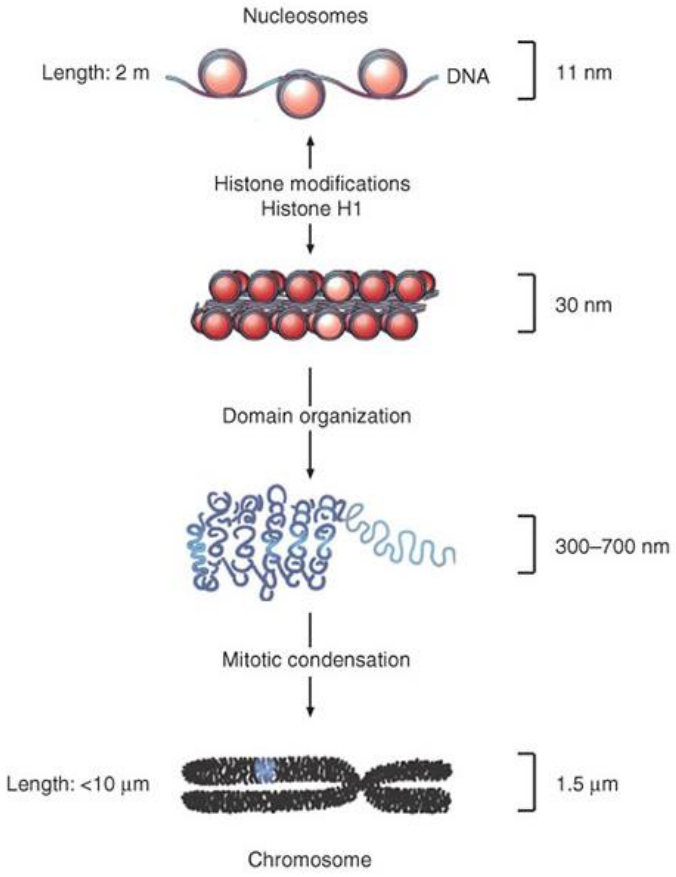



**Figure 3. The organization of chromatin – from nucleosomes to chromosome.**

The 11 nm chromatin fibre consists of individual nucleosomes connected by a free linker DNA. The 30 nm fibre is more compact structure, in which linker DNA is bound by histone H1. The 300-700 nm structure is further folded and looped form of chromatin, creating so called higher-order chromatin. The 1.5 um structure, the chromosome, is the most condensed form of chromatin present in the cell only during cell division. Adapted from (Felsenfeld and Groudine, 2003).

## 2.1. Histone variants

Metazoan genomes encode several histone isoforms. The genes encoding canonical histones, which act in general genome packaging, are clustered together in one array, and are transcribed during the S phase for rapid deposition on newly replicated DNA. Non-canonical histones, histone variants, are typically found individually in the genome, and neither their transcription nor their deposition is synchronised with replication. Variant histones exist for H3, H2A and H2B; they usually vary only in few amino acids compared to the canonical form, but these small differences significantly affect nucleosome assembly and impact the dynamics of the bound chromatin region (Zlatanova et al., 2009). Their roles include

transcription regulation, DNA repair, chromosome segregation, X chromosome inactivation and sperm chromatin condensation (Talbert and Henikoff, 2010, 2016).

The interplay between canonical histones and histone variants is well documented for H3. Mammals have two canonical H3 isoforms – H3.1 and H3.2, and three known variants – H3.3, cenH3 (in mouse and human called CENP-A), and H3t (Talbert and Henikoff, 2010). The canonical H3 is incorporated into chromatin during replication, or alternatively during DNA repair, by a histone chaperone complex called Chromatin assembly factor 1 (CAF1) (Polo et al., 2006; Tagami et al., 2004). When genes become transcriptionally active, canonical H3 is exchanged for H3.3 variants by another complex, Histone regulator 1 (HIRA) (Tagami et al., 2004). Surprisingly, H3.3 can also be assembled into repressed regions like telomeres, mediated by Death-domain-associated-protein – Alpha-thalassemia/mental-retardation-syndrome-X-linked complex (DAXX–ATRX) (Goldberg et al., 2010). The H3t variant is testis specific, and CENP-A is a centromere-specific variant. CENP-A is essential for kinetochore assembly and, consequently, chromosome segregation (Amor et al., 2004). Interestingly, CENP-A is deposited irrespectively of centromeric DNA sequence, as evidenced by so-called neocentromeres – spontaneously formed centromeres at ectopic sites that do not contain any centromeric sequences (Fukagawa and Earnshaw, 2014; Voullaire et al., 1993).

The known mammalian variants for H2B are testis specific. H2A variants are H2A.X, macro-H2A, H2A.Z, H2A.B and H2A.L, and they govern a wide range of functions. The H2Y.X variant was shown to function in double-strand break repair, macro-H2A is a variant specific to inactivated X chromosome, and H2A.Z is often found at promoters of active genes (Bönisch and Hake, 2012; Talbert and Henikoff, 2010). In summary, histone variants provide an additional level of chromatin regulation and specification, and possibly help impose epigenetic states over the corresponding regions.

## 2.2. Histone modifications

Histones are small highly basic proteins consisting of an alpha helical globular domain and a flexible N-terminal “tail”. The globular domains of individual histones form the core of the nucleosome, and the tails protrude from the surface of the core (Fig. 1). The tails of histones

are subject to extensive post-translational modification (PTM), and some PTMs occur at the globular domains as well. The most frequent and best characterized modifications are acetylation, methylation, phosphorylation and ubiquitination; the function and significance of many of the other PTMs remain to be further investigated (Fig. 3; Bannister and Kouzarides, 2011; Tan et al., 2011; Zhao and Garcia, 2015). The tendency of histone modifications to occur at specific residues led David Allis and Brian Strahl to propose the concept of a "histone code" – a unique combination of distinct histone modifications acts as an information centre signalling to downstream factors and pathways, with a different code resulting in a different outcome (Strahl and Allis, 2000). As a result, PTMs can either solidify the inherently repressive nature of chromatin, or alternatively, overcome it and create transcriptionally permissive chromatin.

The following chapters focus only on modifications of histone H3 as this is the most relevant to my PhD project.

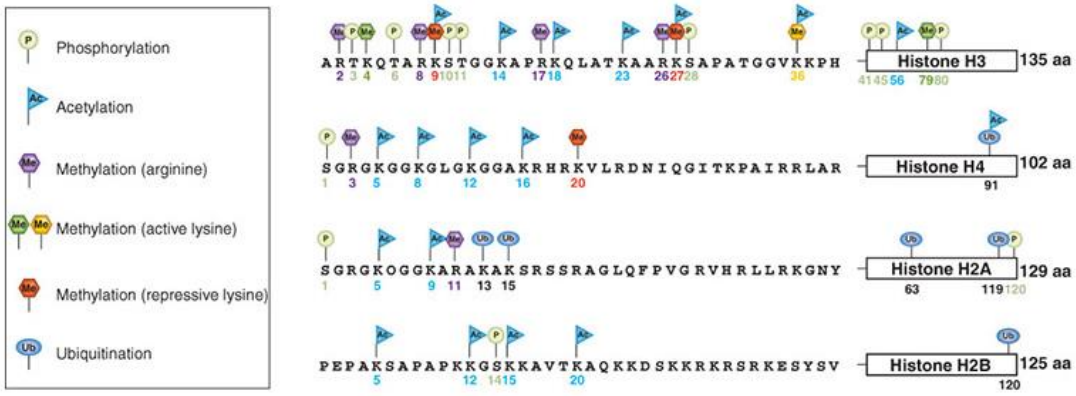



**Figure 4. The posttranslational modifications of core histone tails**

The N-terminal tails of histones are subject to diverse posttranslational modifications (PTM); PTMs occur in the globular domain as well (boxed), some of which are indicated. In general, active marks include acetylation (turquoise Ac flag), arginine methylation (purple Me hexagon), and sometimes lysine methylation, such as H3K4 (green Me hexagon) and H3K36 (yellow Me hexagon) Repressive marks include methylation at H3K9, H3K27, and H4K20 (red Me hexagon). Reprinted from (Allis et al., 2015) with the permission from Cold Spring Harbor Laboratory Press ©.

### 2.2.1. Histone acetylation

The fact that histones and their modifications do not simply serve as structural scaffolds was shown in 1964 in a pioneering study by Allfrey *et al.*, which investigated the role of histone acetylation in the regulation of RNA synthesis (Allfrey et al., 1964). In this work, an *in vitro* transcription assay was used to test the effect of no histones, histones or increasingly acetylated histones on DNA-dependent-RNA polymerase activity. The simple presence of histones had a strong inhibitory effect on transcription that could be gradually overcome by an increase in acetylation levels. The most likely explanation, according to the authors, was that the specificity of DNA-histone binding was altered by acetylation, and it was proposed that this might serve as a transcriptional switch.

The confirmation of the above hypothesis and the explanation of the role of histone acetylation *in vivo* came, however, only in 1990s. First, it was shown that histone tails are crucial for efficient transcription in yeast (Durrin et al., 1991). Shortly after it was found that the known yeast transcription regulator, Gcn5p, is homologous to the histone acetyltransferase (HAT) purified from *Tetrahymena* (Brownell et al., 1996), and in parallel work that the yeast repressor factor Rpdp3 is a histone deacetylase (HDAC) (Taunton et al., 1996). The identification of acetylated amino acids in histone tails followed, and Gcn5p was shown to specifically acetylate H3 at lysine 14 (H3K14Ac), and H4 at lysine 8 and 16 (H4K8Ac and H4K16Ac) (Kuo et al., 1996). Finally, it was demonstrated that histone lysine acetylation can be read by a specialized protein domain – the bromodomain – found in many of known transcriptional activators (Dhalluin et al., 1999).

Over time many other HAT and HDAC enzymes were identified (Yang and Seto, 2007), and today we know that all four core histones can be acetylated at specific lysine residues (Fig. 3). Eventually, the genome-wide distributions of various histone lysine acetylations were mapped and found to globally overlap with active transcription (Wang et al., 2008). The HATs, HDACs and bromodomain-containing proteins became common therapeutic targets as it was found that many human diseases, like cancer and metabolic disorders, correlate with an aberrant form of the above proteins (Keppler and Archer, 2008).

### 2.2.2. Histone phosphorylation

At the time when David Allis proposed the idea of the histone code, acetylation had been the most studied and appreciated modification, yet it had been known for a long time that histone tails carry many other PTMs, including phosphorylation and methylation (Strahl and Allis, 2000). What was often lacking at that time, however, was a functional link between the modifications and their effect on gene expression. Phosphorylation, another negatively charged modification, came first in the spotlight of researchers. Luis Mahadevan and his colleagues used growth factors to induce early-response genes, *c-jun* and *c-fos*, and observed immediate phosphorylation of chromatin-associated protein "pp15", which correlated with the activation of the above genes. Pp15 was identified as histone H3, with the main phosphorylation site at serine 10 (H3S10P) (Mahadevan et al., 1991). Not surprisingly, phosphorylation did not act alone in the activation of early-response genes, and acetylation on the same histones was observed, H3K9AcS10P (Clayton et al., 2000). In addition, the authors observed other unidentified PTMs present on the dual modified histones, suggesting an even more complex activation mechanism.

### 2.2.3. Histone methylation

Histone methylation puzzled researches for a bit longer. It is electrostatically neutral, and therefore cannot alter histone-DNA interactions, suggesting that it rather serves as a recruiter of other nuclear regulators. The first breakthrough came in 1999, when two different groups isolated histone methyltransferase enzymes (HMTs). Chen and co-workers isolated a transcriptional coactivator with HMTase activity and specificity for H3 arginine residues – coactivator associated arginine methyltransferase 1 (CAMR1) (Chen et al., 1999). In the second work, Strahl *et al.* purified an HMTase from the transcriptionally active macronuclei of *Tetrahymena*, with preferential specificity for H3 lysine 4 (H3K4me). The authors also confirmed the conservation of this modification in yeast and mammalian cells (Strahl et al., 1999). Both studies concluded that histone methylation acts in transcriptional activation.

Shortly after, a mammalian HMTase was isolated, suppressor of variegation 3-9 homolog (Suv39h), this time with specificity for H3 lysine 9 (H3K9me) (Rea et al., 2000). Unexpectedly, the known homologs of Suv39h, *S. pombe* Clr4 and *Drosophila* Su(var)3-9, were known repressors acting in heterochromatin silencing (Ivanova et al., 1998; Tschiersch

et al., 1994). Rea *et al.* confirmed that the Clr4 homolog is an active HMTase with H3K9 specificity. Next, they used differentially modified H3 peptides as a substrate for an *in vitro* HMTase assay, and found that H3K9Ac and H3S10P block the ability of Suv39h to methylate K9. This result pointed towards the possibility of an *in vivo* regulation mechanism protecting active chromatin from deposition of repressive marks. Finally, based on the obtained data, they proposed that H3K9me acts as a key determinant of the histone code defining heterochromatin. Indeed, H3K9me was later shown to act as a binding platform for heterochromatin protein 1 (HP1) homologs (Bannister et al., 2001; Lachner et al., 2001), known heterochromatic factors acting in gene silencing and chromatin compaction (Eissenberg and Elgin, 2000; Platero et al., 1995). Interestingly, binding of the HP1 proteins to the H3K9me mark was found to be regulated by H3S10P (Fischle et al., 2005). Fischle *et al.* showed that phosphorylation of H3SP10 by kinase Aurora B during M-phase ejects HP1 proteins from chromatin. At the end of the M-phase, phosphorylation is removed, and HP1 binding can be restored.

The next repressive HMTase that was discovered was the Extra-sex-combs and Enhancer-of-zeste (ESC-E(Z)) complex, which was purified from fly embryos with methylation activity directed towards H3 lysine 27 (H3K27me) (Müller et al., 2002). Mutations in the catalytic subunit of the complex, E(Z), disrupted methyltransferase activity *in vitro*, and *in vivo* led to the derepression of known ESC-E(Z) targets, Homeobox (*Hox*) genes. Interestingly, the ESC component was found to be a H3K27me reader, with the ability to allosterically stimulate the HMTase activity of E(Z) (Hansen et al., 2008; Margueron et al., 2009). Disruption of the ability of ESC to bind H3K27me in flies led to a global reduction of methylation and defects in development. This elegant combination of H3K27me reader and writer factors within one complex allows for the epigenetic transmission of repressed state across cell divisions. Similarly to H3K9me, methylation of the H3K27 residue is inhibited by active marks that are already present on the H3 tail, thus restricting ESC-E(Z) activity to already silent chromatin (Schmitges et al., 2011). Note: the ESC-E(Z) complex is also known as Polycomb repressive complex 2 (PRC2).

Unlike histone acetylation, where HATs deposit the mark and HDACs remove it, histone methylation was originally thought to be a permanent mark. This assumption was later disproved, and methylation was found to be removed by histone demethylases (HDMs). Two

different classes of HDMs, lysine specific demethylases (LSD1 and LSD2) and Jumonji-C (JmjC)-domain-containing histone demethylases, were found (Karytinos et al., 2009; Klose et al., 2006; Shi et al., 2004). The LSD1 enzyme specifically demethylates the H3K4me mark. The knockdown of LSD1 leads to an increase of H3K4me and, consequently, to upregulation of gene expression (Shi et al., 2004). Similarly, H3K9me is a reversible mark removed by JmjC domain-containing histone demethylation protein 2A (JHDM2A). The knockdown of JHDM2A results in the increase of H3K9me levels at a subset of genes, accompanied by a decrease in their expression (Yamane et al., 2006). More detailed analyses of HDMs revealed that they have a wider range of substrates than originally thought, for example LSD1 can demethylate H3K9me as well; eventually many other homologs of JmjC demethylases with specificity for other histone methylations were found (Højfeldt et al., 2013).

Today we know that the most common methylation sites of histone H3 are lysines at position 4, 9, 27, 36, 79 and arginines at position 2, 17, 26 (Fig. 3; Bannister and Kouzarides, 2011). Residues can be mono- (me1), di- (me2) or tri- (me3) methylated, allowing for yet another level of regulation (more in Chapter 5). In general, individual histone modifications show a discrete localization to active or silent chromatin regions, and in addition, they can have a specific distribution pattern within their chromatin region, e.g. different active marks in promoters and gene bodies. The distribution of the most important histone PTMs is further described in the next chapter (Fig. 5).

## 2.3. Distribution of distinct histone marks

The genome-wide mapping of histone modifications revealed that active genes associate with H3K4me1/2/3, H3K36me3 and hyperacetylated histones (Barski et al., 2007; Heintzman et al., 2007, 2009; Mikkelsen et al., 2007; Wang et al., 2008). The promoters of active genes show a high enrichment of RNA polymerase and H3K4me3, while active enhancers are marked by H3K4me1, p300 (HAT) and H3K27Ac (Creyghton et al., 2010; Heintzman et al., 2009). The H3K4me2 mark is present in enhancers, promoters and to some extent gene bodies (Heintzman et al., 2009). Finally, H3K36me3 marks the gene bodies of transcribed genes (Barski et al., 2007).

The repressed chromatin regions are characterized by H3K9, H3K27 and H4K20 methylation marks. H4K20me3 and H3K9me3 play a role in the silencing of constitutive heterochromatin, while H3K9me2 and H3K27me3 are associated with the silencing of facultative heterochromatin (Mikkelsen et al., 2007; Peters et al., 2003; Schotta et al., 2004).

Interestingly, H3K27me3 can co-occur with H3K4me3 at the promoters of developmental genes in embryonic stem cells (ESCs), and this chromatin type is referred as bivalent (Bernstein et al., 2006). The proposed function of bivalent domains is to poise promoters for rapid activation upon developmental stimulation. The nucleosomes were originally thought to be modified in an asymmetric way, with one H3 tail carrying H3K4me3 and the other H3K27me3 (Voigt et al., 2012). More recently, however, Shema *et al.* used single molecule technologies to show that this bivalent mark can even occur on the same H3 tail (Shema et al., 2016).

The different functional chromatin domains must be separated from each other to prevent spreading of epigenetic modifications or block unwanted interactions between regulatory elements, e.g. enhancers and promoters. The information about these "borders" is encoded in the DNA sequence, referred to as insulators or boundary elements. In mammals, the CCCTC-binding factor (CTCF) was shown to act as an insulator of neighbouring regions, forming so called topologically associating domains (TADs) (Dixon et al., 2012; Gaszner and Felsenfeld, 2006; Nora et al., 2017). CTCF binds to more than 20 000 sites in the genome, and its binding was shown to relatively correlate with all three methylation states of H3K4me, H3K9me1 and H2A.Z variant (Barski et al., 2007).

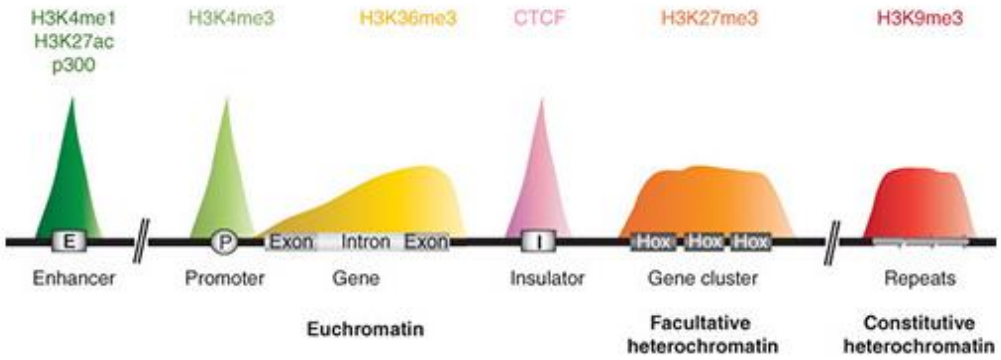



**Figure 5. Distribution of histone marks**

Enhancer, promoter and gene body are indicated in the scheme as an example of actively marked gene. H3K4me1, H3K27ac and HATase p300, marks active enhancers. H3K4me3 and hyperacetylation mark active promoters, and H3K36me3 marks gene bodies of actively transcribed genes. Repressed genes can be marked by H3K27me3, facultative heterochromatin, or alternatively by H3K9me3 in case of constitutive heterochromatin. Ctcf marks insulators creating boundaries between local chromatin domains. Altogether, creating a map of functional elements and chromatin domains. Reprinted from (Allis et al., 2015) with the permission from Cold Spring Harbor Laboratory Press ©.

## 3. Transcription factors

Unlike in prokaryotes, where DNA is freely accessible to the transcriptional machinery, the DNA of eukaryotes is by default in a transcriptionally restricted state (Struhl, 1999). This state can be further reinforced or elevated by chromatin modifications and chromatin associated factors. The above chapters summarized how different histone modifications affect gene expression, and act as transmitters of the functional status across cell divisions. However, the initial signal for establishment of an active or inactive chromatin environment must come from regulatory factors that have the ability to recognize and bind specific DNA sequences.

The power of DNA binding factors was first demonstrated by Davis *et al*., who showed that fibroblasts can be converted to myoblasts by overexpression of a single factor, Myoblast-determination (MyoD) protein (Davis et al., 1987). Other examples followed; in one study the targeted expression of the Paired box 6 (Pax6) protein in developing *Drosophila* was shown to induce the growth of ectopic fully developed eyes on wings or legs (Halder et al., 1995). A decade ago, Shin Yamanaka and his group succeeded in converting mouse fibroblast into pluripotent stem cells by overexpressing four factors, Oct3/4, Sox2, c-Myc and Klf4 (Octamer-binding transcription factor 3/4, Sex-determining-region-Y box 2, Myelocytomatosis-viral-oncogene homolog and Kruppel-like factor 4) (Takahashi and Yamanaka, 2006), a major discovery that started a whole new field.

All of the above factors are so called “transcription factors”, a class of proteins with the ability to recognize and bind specific DNA sequences to activate or repress gene activity. The ability of most transcription factors (TFs) to bind their cognate DNA motif is regulated by nucleosome distribution, epigenetic modifications and higher-order chromatin organization. Indeed, genome-wide mapping studies of TF binding have shown that most TFs bind only a minority of their consensus target sites in the genome, mostly in already accessible or active regions (Li et al., 2011; Slattery et al., 2014). Yet, there is a subgroup of TFs, pioneer factors, which are able to recognize and bind their specific DNA motif irrespective of nucleosomes (Cirillo et al., 1998). The forkhead box A (FoxA) possesses a DNA binding domain that is very similar to the nucleosome-binding domain of linker histone H1. This is a rather unusual resemblance, considering that H1 acts in compaction and thus repression, whereas FoxA acts as a transcriptional activator. Biochemical, and later *in vivo*, analyses revealed that FoxA can

bind its full DNA motif on the nucleosome, displace H1, and using its C-terminus to interact with the core histones, opening chromatin and making it accessible for other co-factors or downstream regulators (Cirillo et al., 1998; Iwafuchi-Doi et al., 2016). Three of the four proteins previously mentioned as having a capacity to reprogram (Oct3/4, Sox2 and Klf4 (OSK)) were also found to act as pioneer factors (Soufi et al., 2012, 2015). Interestingly, the fourth factor, c-Myc, prefers to bind to active chromatin on its own, but can bind inaccessible chromatin in cooperation with the OSK pioneer factors (Soufi et al., 2012, 2015). Analogously other transcription factors were found to cooperate with pioneer factors to bind inaccessible chromatin (Zaret and Carroll, 2011; Zaret and Mango, 2016). It is important to note that even pioneer factors cannot efficiently access epigenetically repressed chromatin like H3K9me-marked heterochromatic domains (Soufi et al., 2012).

Nevertheless, some transcription factors were found to access their sites on nucleosomes independently of pioneer factors. TFs that cannot bind their targets on nucleosomes individually were shown to be able to do so simultaneously, implying that cooperative binding might be a general mechanism (Adams and Workman, 1995). Other models propose that Adenosine triphosphate (ATP)-dependent chromatin remodelling complexes, remodellers, constantly scan the genome, loosen chromatin, and thus create access points for transcription factors (Erdel et al., 2010; de la Serna et al., 2006). Alternatively, TFs can couple with remodellers, accessing their sites together (Cosma et al., 1999).

In general, transcription factors that bind DNA in a sequence-specific manner contain one of the following DNA binding domains: helix-turn-helix (HTH), zinc-coordinating or zipper-type (Luscombe, Austin, Berman†, & Thornton, 2000). The zinc-coordinating type is the most frequently occurring class in the genome of *Homo sapiens.* Transcription factors are usually classified based on their DNA binding domains or, alternatively, according to their hierarchy in binding to chromatin (e.g. pioneer factors). The following chapter is going to introduce a transcription factor directly related to my PhD project.

## 3.1. Activity dependent-neuroprotective protein

The Activity-dependent neuroprotective protein (Adnp) is a putative transcription factor that is highly conserved in human, rat and mouse (Bassan et al., 1999; Sigalov et al., 2001; Zamostiano et al., 2001). Adnp was originally discovered as a protein associated with neuroprotection (Bassan et al., 1999), but later was found to be required during mouse development – its deletion is embryonically lethal (Pinhasov et al., 2003). This implicates that Adnp functions beyond the nervous system, yet its molecular function is not understood. The following paragraphs summarize current knowledge about the Adnp protein.

### 3.1.1. Domain organization

Adnp contains nine N-terminal zinc fingers, a C-terminal homeodomain, a C-terminal Proline-X-Valine-X-Leucine (PxVxL) pentapeptide motif, and a histone H3K9 mimic sequence (Fig. 6; Bassan et al., 1999; Mandel et al., 2007; Mosch et al., 2011; Zamostiano et al., 2001). All of the zinc fingers (ZF) in Adnp are the cysteine-cysteine-histidine-histidine (C2H2) type. The C2H2 motif is characterized by a tetrahedral coordination of one or two zinc ions (Zn) (Luscombe et al., 2000). ZFs are usually found in multiple copies within one protein, and are frequently accompanied by another DNA binding domain (Bateman et al., 2017). Adnp contains two conserved (6, 7), three atypical (5, 8, 9) and four degenerate (1-4) ZFs; atypical ZFs deviate from the consensus sequence but can coordinate Zn, while degenerate ZFs are too different from the original C2H2 motif and have thus lost the ability to bind Zn (Bateman et al., 2017). The homeodomain shows the closest homology to the homeobox domain of the Hox protein family. Homeodomains contain a HLH motif, which consists of two α-helices connected by a flexible linker. The HTH motif invariably binds in the DNA major groove; the first helix helps stabilize the structure, and the second helix, commonly known as the recognition helix, is inserted in the groove forming a number of hydrogen bonds and hydrophobic interactions with DNA (Luscombe et al., 2000). The C-terminal PxVxL pentapeptide, where x can be any amino acid, is a protein interaction motif that can be read by HP1 proteins (Thiru et al., 2004), and similarly the H3 mimic was shown to be recognized by HP1s or alternatively by HMTs (Sampath et al., 2007).


**Figure 6. Domain architecture of the mouse Activity-dependent neuroprotective protein (Adnp).**

The Adnp protein consists of nine N-terminal zinc fingers (ZFs) - two typical ZFs (6, 7), three atypical ZFs (5, 8, 9) and four degenerate ZFs (1, 2, 3, 4); followed by C-terminal DNA binding motif, homeobox domain. Two additional motifs are adjacent to homeobox domain, histone H3 mimic and PxVxL; both of these motifs can facilitate interaction with HP1 homologs.

### 3.1.2. Interacting partners

The Adnp protein was found to co-immunoprecipitate one of the three HP1 homologs, HP1α, from the embryonic carcinoma cell line P19 as well as from differentiated P19 cells (Mandel et al., 2007). This interaction was further supported by another work that showed HP1α- or HP1β-dependent Adnp association with heterochromatin (Mosch et al., 2011). The presence of the PxVxL motif in Adnp suggests that this is a direct interaction, but this has not been confirmed biochemically. In another study, immunoprecipitation of ADNP from human embryonic kidney 293 (HEK293) cells followed by mass spectrometry identified Brahma-related gene 1 (BRG1), BRG1-associated factor 250a and 170 (BAF250a, BAF170), as co-immunoprecipitating proteins, all components of the Switch/sucrose non-fermentable (SWI/SNF) chromatin-remodelling complex (Mandel and Gozes, 2007). To date, there are no other publications describing Adnp interactors, although Adnp was found in several works as a "prey". For example, Adnp was identified as an interactor of HP1 homologs (Nozawa et al., 2010; Rosnoblet et al., 2011; Vermeulen et al., 2010), and the mutation of a residue crucial for PxVxL motif recognition in HP1 abolished its interaction with ADNP in HEK293 cells (Nozawa et al., 2010). Large-scale interactome profiling of various transcription factors found Adnp in Forkhead box (Fox) homologs, Nuclear factor of activated T-cells 1 (Nfatc1) or TEF1-abaA (TEA)-domain transcription factor 2 (Tead2) purifications (Li et al., 2015), although the functional relevance of these interactions is unclear.

### 3.1.3. Activity

The association of Adnp with chromatin was confirmed by chromatin immunoprecipitation (ChIP) and immunofluorescence assays (Mandel et al., 2007; Mosch et al., 2011). ADNP was proposed to act as a regulator of gene expression in association with HP1 and/or the SWI/SNF complex, and the knockdown of ADNP in HEK293 cells induced distinct morphological changes and reduced viability (Mandel and Gozes, 2007). Moreover, complete deficiency in Adnp during mouse development results in general organogenesis defects, and failure in neural tube closure resulting in death at days 8.5-9.5 of gestation (E8.5-9.5) (Pinhasov et al., 2003). The *in situ* hybridization of the pluripotency marker Oct4 and the neuronal marker Pax6 in control and mutant E8.2 embryos revealed that mutants failed to downregulate Oct4 or induce Pax6 expression. In the follow-up work from the same laboratory, Mandel *et al.* analysed gene expression by microarrays in Adnp-deficient mouse embryos (Mandel et al., 2007). The embryos exhibited upregulation of transcripts associated with lipid metabolism and extra-embryonic tissues, and downregulation of organogenesis, and in particular neurogenesis, related transcripts.

ADNP is abundantly expressed in different human adult tissues with highest expression levels in cerebellum, cortex, heart, skeletal muscles, kidney, placenta and testis (Zamostiano et al., 2001). The levels of ADNP are often elevated in malignant cells, and the knockdown of Adnp in intestinal cancer cells reduced their viability by 90%, which was accompanied by upregulation of the tumour suppressor protein p53 (Zamostiano et al., 2001). More recently, *de novo* mutations in the *ADNP* gene were found to underlie ADNP syndrome (also known as Helsmoortel-Van der Aa syndrome/HVDAS), an extremely rare complex neurodevelopment disorder (Helsmoortel et al., 2014). This syndrome is commonly associated with neurological, cardiovascular, musculoskeletal, gastrointestinal, endocrine and immune defects, as well as vision, hearing, growth, feeding and sleep impairment. Additionally, patients exhibit delays in speech and intelligence, and many develop behavioural conditions such as autism spectrum disorders (ASD). It is estimated that at least 0.17% of ASD patients have an *ADNP* mutation, making it one of the most frequent ASD-associated genes known to date (Helsmoortel et al., 2014).

## 4. Chromatin remodellers

Although chromatin generally restricts access of proteins to DNA, the degree to which individual nucleosomes inhibit transcription factors from interacting with their specific sites is quite variable, and can be further modulated by altering the nucleosome positioning via chromatin remodellers (Workman and Kingston, 1998). Genome-wide mapping of nucleosome positioning in *S. cerevisiae* has revealed a stereotyped Pol II promoter organization in which a nucleosome-free region (NRF) is present within approx. 200 bps upstream of the start codon flanked on both sides by well positioned nucleosomes (Yuan et al., 2005). This suggests that transcriptional activation requires exposure of DNA, which in turn means that access of transcription factors to regulatory elements, and therefore gene expression, can be regulated by remodellers and nucleosome positioning.

Remodellers use ATP hydrolysis as an energy source to slide, space, eject or assemble nucleosomes; or to replace individual histones within the nucleosome. In mammals, approx. 30 genes encoding for chromatin-remodelling ATPase subunits have been identified (Ho and Crabtree, 2010). The vast majority of remodelling ATPases appears to be genetically nonredundant as their deletions result in early embryogenesis phenotypes, suggesting specific biological roles. Phylogenetic and functional analyses have grouped all remodeling ATPases into the RNA/DNA helicase superfamily 2 (Clapier et al., 2017). Based on the sequence and the structural conservation of individual ATPases, the chromatin remodellers can be further divided into four subfamilies: Imitation switch (ISWI), INO80, SWI/SNF and Chromodomain helicase DNA-binding (CHD) (Clapier and Cairns, 2014). These subfamilies are conserved from yeast to human and all contain their subtype ATPase domain, but differ in additional domains and features. Furthermore, higher eukaryotes evolved multiple remodeller subtypes within each subfamily, often with cell-type or developmentally specific functions (Ho and Crabtree, 2010; Kadoch and Crabtree, 2015).

ISWI remodellers act mostly as repressors, assembling and regularly spacing nucleosomes in a way that limits DNA accessibility (Clapier et al., 2017). An exception in this subfamily is the Nucleosome-remodelling-factor (NURF) complex, which, with the help of additional subunits, randomizes spacing and facilitates transcriptional initiation (Xiao et al., 2001). Apart of the ATPase domain, ISWI members contain a typical C-terminal HAND-SANT-SLIDE

(HSS) domain that can bind the unmodified H3 tail and the linker DNA (Boyer et al., 2004; Dang and Bartholomew, 2007; Grüne et al., 2003).

INO80 members mostly function in nucleosome editing that is associated with transcriptional activation and DNA repair (Clapier et al., 2017). They can remove individual histones within an octamer and replace them with canonical or variant histones. For instance, the SWR1 complex replaces canonical H2A for its H2A.Z variant in gene promoters (Mizuguchi et al., 2004), and so does E1A binding protein p400 (EP400), which can in addition exchange H3.1 with its H3.3 variant in promoters and enhancers (Pradhan et al., 2016). Interestingly, another family member, INO80c, prevents H2A.Z depositioning outside of gene promoters (Papamichos-Chronakis et al., 2011).

SWI/SNF type remodellers are the best described subfamily. They typically provide chromatin access by repositioning and/or ejecting nucleosomes, which can facilitate either activating or repressive functions and DNA repair (Clapier et al., 2017). SWI/SNF remodellers contain an N-terminal helicase SANT-associated (HAS) domain, which can bind actin and actin-related proteins (Schubert et al., 2013), and a C-terminal bromodomain accompanied by a pair of DNA binding motifs, AT-hooks (Clapier et al., 2017).

CHD family members are versatile remodellers providing chromatin access by nucleosome sliding, but also performing nucleosome assembly and editing (Clapier et al., 2017). They are characterised by two N-terminal tandemly arranged chromatin organization modifier (chromo) domains (Delmas et al., 1993; Woodage et al., 1997). The family is further divided into three subtypes based on the presence of additional domains (Marfella and Imbalzano, 2007). Subtype 1 (CHD1/CHD2) contains a C-terminal DNA binding domain consisting of a SANT-SLIDE motif similar to ISWI (Ryan et al., 2011), subtype 2 (CHD3/CHD4/CHD5) contains a paired plant homeodomain (PHD) zinc-finger-like domain at the N-terminus (Woodage et al., 1997), and subtype 3 (CHD6-CHD9) contains two C-terminal domains, SANT and BRK (Brahma and Kismet) (Shur and Benayahu, 2005).

CHD3/CHD4 ATPases are fundamental subunits of the Nucleosome remodelling and histone deacetylase (NuRD) complex. This complex functions in transcriptional activation and repression, in replication-coupled chromatin assembly, and in DNA repair (Basta and Rauchman, 2017; Lai and Wade, 2011). NuRD contains another enzymatic subunit, histone deacetylase 1 (HDAC1) or HDAC2 that catalyse histone deacetylation, and other non-

enzymatic subunits: metastasis-associated gene 1 (MTA1), MTA2 and MTA3, methyl-CpG-binding domain 2 (MBD2) and MDB3, retinoblastoma-binding protein 4 (RBBP4; also known as RBAP48) and RBBP7 (RBAP46), and GATAD2A (p66α) and GATAD2B (p66β) (Denslow and Wade, 2007; Lai and Wade, 2011). The complex is possibly targeted across the genome by MTA and MBD subunits; MTA by associating with transcription factors (Fujita et al., 2004), MBD2 by binding to methylated DNA (Hendrich and Bird, 1998), and MBD3 by binding to a product of active DNA demethylation, 5-hydroxymethylcytosine (Yildirim et al., 2011). In addition, the CHD4 subunit can bind DNA by its chromodomain (Bouazoune et al., 2002), and differentially modified H3 tails via its PHD fingers (Musselman et al., 2009). CHD4 seems to have NuRD independent functions, and it is possible that some of the assigned NuRD functions are in fact CHD4 specific (Basta and Rauchman, 2017). Biochemically, CHD4 does not require other NuRD components for its remodelling activities (Wang and Zhang, 2001), and *Chd4* mouse knockout embryos die preimplantation, while other NuRD components show phenotypes at later stages (O'Shaughnessy-Kirwan et al., 2015).

# 5. Chromatin readers - HP1 proteins

HP1 was originally discovered in *Drosophila*, by Sarah Elgin's laboratory, as a non-histone chromosomal protein associating with heterochromatin (James and Elgin, 1986). Shortly after, the same group identified HP1 as a dominant suppressor of position-effect variegation (PEV), a phenomenon of mosaic silencing observed when a euchromatic gene is placed near or within heterochromatin (Eissenberg et al., 1990). The family of HP1 proteins quickly expanded as other homologs were identified in *Drosophila* and other organisms (Singh et al., 1991). Today, we know that the HP1 family is evolutionary conserved with members across nearly all eukaryotes, ranging from fission yeast to plants to mammals. For instance, the fission yeast homolog shows overall 20% amino-acid sequence identity with mammalian homologs, and its chromo domain can be functionally replaced by the mouse HP1β chromo domain (Canzio et al., 2014; Wang et al., 2000). In mammals, the HP1 proteins are encoded by the *Chromobox* (*Cbx*) genes. The Cbx superfamily is defined by a conserved N-terminal domain, the chromatin organization modifier (chromo) domain (CD) (Paro and Hognesst, 1991). Both mice and humans possess three HP1 homologs, HP1α, HP1β and HP1γ, which are encoded by the *Cbx5*, *Cbx1* and *Cbx3* genes in mice, and the *CBX5*, *CBX1* and *CBX3* genes in humans, respectively. The murine and human genes are located on syntenic chromosomal regions, and the proteins show an extremely high level of sequence identity – 98% for HP1α, 100% for HP1β, and 99.5% for HP1γ (Bateman et al., 2017; Lomberk et al., 2006a).

## 5.1. Domain organization

Apart from the N-terminal chromo domain, HP1 proteins contain another structured module, the C-terminal chromo-shadow domain (CSD). Both domains have flexible extensions, N-terminal (NTE) and C-terminal (CTE), and are connected to each other by a flexible linker called the hinge (Fig. 7). The HP1 chromo domain specifically recognizes methylated lysine in the ARKS motif (ARKmeS), which is present in the H3 histone (H3K9) (Bannister et al., 2001; Jacobs et al., 2001; Lachner et al., 2001), but also in other proteins, such as the Eukaryotic HMTase 2 (Ehmt2) (Sampath et al., 2007). The chromo domain was shown to have the highest affinity to H3K9me3 and H3K9me2 peptides, followed by

H3K9me1, with the unmodified peptide showing the weakest affinity (Fischle et al., 2005; Yamada et al., 2005). The binding constants are rather weak, in the μmol.L$^{-1}$ (μM) range; however, the affinity seems to be further fortified *in vivo* by additional mechanisms like the unspecific interaction of the hinge with nucleosomal DNA (Mishima et al., 2013) or by PTMs – the phosphorylation of tandem serines in the NTE of HP1α increases its affinity approximately fivefold (Hiragami-Hamada *et al.*, 2011). Interestingly, this array of serines is not present in the other two homologs. The structure of the HP1β chromo domain resolved by nuclear magnetic resonance (NMR) revealed an N-terminal three-stranded antiparallel β-sheet folded against a C-terminal α-helix (Ball et al., 1997), but it did not provide any explanation for the specificity towards the ARKmeS motif. The answer came from the structure of the chromo domain bound to the H3K9me2/3 peptide that revealed an aromatic cage, serving as a hydrophobic pocket for K9me2/3 (Jacobs and Khorasanizadeh, 2002; Nielsen et al., 2002). The other residues of the peptide form bonds with the β-sheet of the chromo domain, and are equally important in establishing a stable specific interaction. A point mutation of alanine from the ARKS motif to methionine reduces the affinity of the CD for H3K9me2 25-fold (Jacobs and Khorasanizadeh, 2002). Similarly, substitution of the CD residue creating contact with the ARKS alanine, valine (V26 in *Drosophila),* leads to the mislocalisation of HP1 and impaired silencing *in vivo* (Platero et al., 1995).

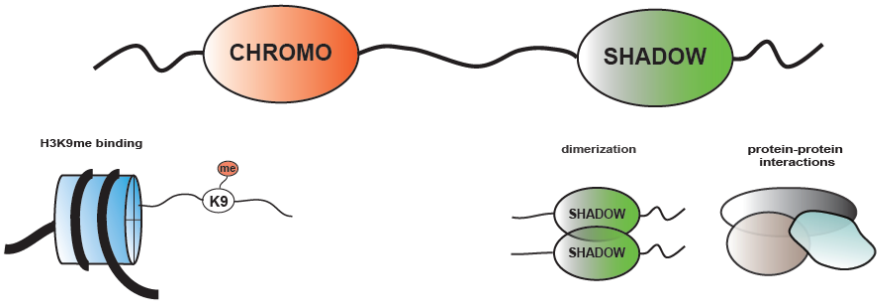



**Figure 7. A schematic representation of HP1 proteins.**

HP1 proteins consist of two conserved domains, chromo and chromo-shadow, connected by a less conserved linker, hinge. In addition, there are N-terminal and C-terminal flexible extensions (NTE and CTE). The chromo domain serves as a recognition and binding module for the H3K9 methylation mark. The chromo-shadow acts as an interaction platform – mediating formation of HP1 homo- or heterodimers, and/or interaction with other proteins.

The chromo-shadow domain adopts a very similar structure to the CD, consisting of a three-stranded antiparallel β-sheet followed by two α-helices (Brasher et al., 2000; Cowieson et al., 2000). The second a-helix, which is missing in the CD, mediates the CSD dimerisation. The fold of the CSD dimer creates a nonpolar binding pocket, which serves as a protein interaction platform that can be bound by proteins containing a PxVxL motif (Thiru et al., 2004). The protein interactions can, in addition, be regulated by the CTE residues (Mendez et al., 2013). Interestingly, the CSD can form homo- and heterodimers, both *in vitro* and *in vivo* (Nielsen et al., 2001).

The hinge linker is the least conserved region in the HP1 family, both within and across species. It contains a nuclear localisation signal, and carries multiple PTMs, with phosphorylation and SUMOylation being the best functionally characterised (LeRoy et al., 2009; Maison et al., 2011; Shimada et al., 2009; Zhao et al., 2001). The hinge contains a high number of KR residues that give it an overall positive charge, and has been implicated in nucleic acid binding (Keller et al., 2012; Maison et al., 2011; Mishima et al., 2013; Muchardt et al., 2002).

## 5.2. Interacting partners

Mammalian HP1s interact with a wide range of proteins, with BioGRID, a protein interaction database, listing nearly 200 HP1-specific interactions. Many of these, however, were not verified biochemically, and therefore are not necessarily direct. For the purpose of this thesis, I am going to list only the best-characterised and/or most relevant ones. In general, the HP1 interactors can be divided into three groups: (1) transcriptional regulators and chromatin-related proteins, (2) DNA replication and repair factors, (3) nuclear structure proteins (Table 1) (Lomberk et al., 2006a). Not surprisingly, group (1) contains histones and histone modifying enzymes. HP1s interact with two HKMTases that create their binding substrate, H3K9me, the Ehmt1/2 and Suv39h1/2 proteins (Aagaard et al., 1999; Melcher et al., 2000; Nozawa et al., 2010; Rosnoblet et al., 2011; Sampath et al., 2007). Ehmt1 and Ehmt2 form a heterodimeric complex that acts as the primary enzyme for H3K9me1 and H3K9me2 in silent euchromatin and in retrotransposons, while Suv39h1 and Suv39h2 act as two

redundant paralogs depositing H3K9me3 in pericentric heterochromatin[1] (Maksakova et al., 2013; Peters et al., 2003; Rea et al., 2000; Rice et al., 2003; Tachibana et al., 2002, 2005). HP1s interact with another factor involved in retrotransposon silencing, tripartite motif containing 28 (Trim28). Trim28 contains the PxVxL motif that mediates direct interaction with HP1s, and acts as a corepressor mediating transcriptional silencing via interaction with Krüppel-associated box (KRAB) domain-containing zinc-finger proteins (Friedman et al., 1996; Lechner et al., 2000; Rowe et al., 2010; Sripathy et al., 2006). The HP1 proteins were shown to associate with the nuclear envelope by interacting with its structural components Lamin B and Lamin B receptor (LBR) (Kourmouli et al., 2000; Lechner et al., 2005; Ye and Worman, 1996). This coincides with the localisation of heterochromatin to the nuclear periphery (Fig. 1; (Akhtar and Gasser, 2007)).

**Table 1. Summary of the mammalian HP1-interacting proteins**

| Interactor | HP1 homolog | Reference |
|---|---|---|
| *(1) Transcriptional regulators and chromatin-related proteins* | | |
| H1 | α | (Hale et al., 2006; Nielsen et al., 2001) |
| Core histones | α, β, γ | (Nielsen et al., 2001; Rosnoblet et al., 2011) |
| Dnmt1/3 | α, β, γ | (Fuks et al., 2003; Lehnertz et al., 2003; Smallwood et al., 2007) |
| Suv39h1/2 | α, β | (Aagaard et al., 1999; Melcher et al., 2000; Nozawa et al., 2010) |
| Ehmt1/2 | α, γ | (Nozawa et al., 2010; Sampath et al., 2007) |
| Brg1 | α | (Lechner et al., 2005; Nozawa et al., 2010) |
| Chd4 | β, γ | (Rosnoblet et al., 2011) |
| Atrx1 | α, β | (Lechner et al., 2005) |
| Trim28 | α, β, γ | (Lechner et al., 2000; Nozawa et al., 2010; Rosnoblet et al., 2011) |
| Adnp | α, β, γ | (Mosch et al., 2011; Nozawa et al., 2010; Rosnoblet et al., 2011) |
| *(2) DNA replication and repair factors* | | |
| Caf1* | α, β, γ | (Lechner et al., 2005; Nozawa et al., 2010; Rosnoblet et al., 2011) |
| Rbb4* | α, β, γ | (Nozawa et al., 2010; Rosnoblet et al., 2011) |
| Ku70 | β, γ | (Lomberk et al., 2006b; Rosnoblet et al., 2011) |
| ORC | α | (Prasanth et al., 2010) |

[1] Pericentric heterochromatin is a region of chromatin that is found juxtaposed to the centromere. It consists of tandemly arranged repeats, which in mouse are called the major satellite repeats.

| | | |
|---|---|---|
| *(3) Nuclear structure proteins* | | |
| Lamin B | β | (Kourmouli et al., 2000) |
| Lbr | α, γ | (Lechner et al., 2005; Ye and Worman, 1996) |

(Trim28; Tripartite Motif Containing 28), (LBR; Lamin B receptor), (ORC, Origin recognition complex), (Rbb4 Retinoblastoma-binding protein 4), (Dnmt1/3 DNA methyltransferase 1 and 3 ), (*Caf1 and Rbb4 form a Caf complex)

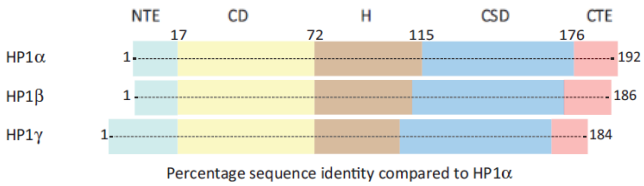
### 5.3. Role of mammalian HP1s

The most common role of HP1is the silencing of chromatin. For this, HP1s use their two-domain structure to physically link their substrate histone mark, H3K9me, with the repressive non-histone proteins that they interact with. In this model, silencing involves an iterative self-sustaining loop, in which HP1 is recruited by H3K9me and/or by the HKMTase depositing the mark. In turn, HP1 recruits more HKMTases, which leads to propagation of the H3K9me mark along the locus, and consequently, to spreading of HP1 binding. The HP1 association with constitutive heterochromatin is greatly reduced in *Suv39h* double-null (dn) cells (García-Cao et al., 2004; Lachner et al., 2001), and similarly, deletion of the *Ehmt2* gene leads to depletion of HP1 binding within silent euchromatin (Tachibana et al., 2005). Alternatively, HP1s can be recruited by Trim28, which also interacts with an H3K9me3-depositing HKMTase, SET (Suv39, E(z), Trithorax)-domain bifurcated 1 (SETDB1) (Schultz et al., 2002).

Artificial targeting of Trim28 to a reporter gene is sufficient to induce silencing, but it requires an intact interaction with HP1 and Setdb1 (Sripathy et al., 2006). Equivalently, tethering of HP1α or HP1β is sufficient to recruit Trim28 and Setdb1, and to repress the target reporter gene (Ayyanathan et al., 2003; Verschure et al., 2005). A later study by Hathaway et al., studying the dynamics of heterochromatin formation and maintenance, found that transient targeting of HP1α upstream of the promoter of the endogenous *Oct4* allele in mouse ES cells, leads to the erasure of the active marks, H3K4me3 and H3K27ac, and to the deposition of H3K9me3. The H3K9me3 mark is first deposited at the tethering site but continues to spread, eventually forming a domain spanning across approx. 10 kilo base pairs (kbp) (Hathaway et al., 2012). The acquired repressed state can be inherited over multiple mitotic divisions in the absence of the originally tethered HP1 (Hathaway et al., 2012;

Verschure et al., 2005). It is important to say, however, that the propagation of the silent state and the presence of H3K9me3 was accompanied by the deposition of the DNA methylation at the target locus (Hathaway et al., 2012). This is consistent with the silencing mechanism observed at naturally repressed chromatin sites. Both HP1s and HKMTases were shown to interact with the *de novo* DNA methyltransferase 3 homologs (Dnmt3a and/or Dnmt3b) (Table 1; Fuks et al., 2003; Lehnertz et al., 2003; Li et al., 2006). The study by Lehnertz *et al.* revealed that the Suv39h-mediated H3K9 methylation, and the HP1 binding, precede the establishment of the DNA methylation at the pericentric heterochromatin. In addition, the HP1 proteins co-immunoprecipitate the maintenance DNA methyltransferase 1 (Dnmt1) (Table 1; (Fuks et al., 2003; Smallwood et al., 2007)). All of the above-mentioned factors – Dnmt1, HKMTases and HP1s, were shown to, directly or indirectly, associate with the replication fork, contributing to the propagation of the silent state across the cell divisions (Table 1; Chuang et al., 1997; Loyola et al., 2009; Sarraf and Stancheva, 2004; Talbert and Henikoff, 2016).

Tethering of one of the HP1 homologs, HP1α, results in the recruitment of the endogenous homologs, HP1β and/or HP1γ, and analogously, tethering of HP1β results in the recruitment of HP1α (Hathaway et al., 2012; Verschure et al., 2005). The ability of individual HP1s to recruit each other and to form heterodimers suggests their cooperative, and perhaps redundant, action in gene silencing. The three mammalian HP1 homologs share a high level of sequence identity (Fig. 8). However, homolog-specific localisation patterns were also observed (Minc et al., 2000), and domain-swapping experiments between different HP1 homologs pointed towards non-redundant functions of the corresponding domains (Kato et al., 2007).




Percentage sequence identity compared to HP1α

| | NTE | CD | H | CSD | CTE |
|---|---|---|---|---|---|
| HP1β | 35% | 82% | 33% | 82% | 38% |
| HP1γ | 15% | 71% | 36% | 87% | 19% |

**Figure 8. Sequence identity of the mammalian HP1 homologs.**

The domain map of the mammalian HP1α, HP1β and HP1γ, illustrating the level of conservation between the individual homologs. The percentage represents sequence identity of HP1β and HP1γ compared to HP1α. (Adopted from Canzio, Larson and Narlikar, 2014)

*Cbx5* deficient mice are viable and fertile, showing no distinguishable phenotype compared to wild-type littermates (Aucott et al., 2008; Brown et al., 2010). The *Cbx1* null mutation, however, leads to perinatal lethality in mice, which is most likely caused by defects in brain development (Aucott et al., 2008). The *in vitro* culture of *Cbx1* null neurospheres uncovered a severe genomic instability, indicating defects in centromere functions. Finally, *Cbx3* mutants exhibit high postnatal lethality (99%), with the surviving animals being infertile (Brown et al., 2010; Naruse et al., 2007). The analysis of the mutant embryos and the surviving adult animals revealed an early loss of germ cells, impaired cell cycle progression, an increase in the transcription levels of retrotransposons, and defects in energy homeostasis (Abe et al., 2011; Aydin et al., 2015; Brown et al., 2010). Altogether, the above-presented data suggest that HP1s have redundant but also homolog-specific non-redundant functions. However, detailed mechanistic analysis of their redundancy and homolog-specialised roles have remained largely unexplored.

# *Results*

# 1. H3K9me-independent targeting of mammalian HP1 homologs

HP1 proteins constitute a fundamental component of heterochromatin, recognizing and binding the hallmark of heterochromatin H3K9me via their chromo domain and consequently mediating silencing of the bound regions. Interestingly, the chromo domain is required for H3K9me-mediated silencing (Platero et al., 1995), but is not sufficient for the initial HP1 targeting and heterochromatin formation (Smothers and Henikoff, 2001), suggesting that additional mechanisms and/or factors are involved in the *de novo* targeting. Furthermore, HP1β and HP1γ were shown to localise not only to heterochromatin but also to euchromatin, possibly in a H3K9me-independent manner (Mattout et al., 2015; Minc et al., 2000; Smallwood et al., 2012). Several studies even proposed the role of HP1γ in the activation of gene expression, transcription or pre-mRNA splicing (Huang et al., 2017; Lomberk et al., 2006b; Skourti-Stathaki et al., 2014; Smallwood et al., 2012). This is quite striking considering that mammalian HP1 homologs are highly conserved and share nearly identical domains. The flexible regions, however, are not as conserved, and their diversification between the individual homologs could explain the different localisation patterns and functions. Indeed, the flexible regions were described to carry homolog-specific posttranslational modifications (LeRoy et al., 2009; Lomberk et al., 2006b; Maison et al., 2011), to modulate the affinity for H3K9me (Hiragami-Hamada et al., 2011), to affect protein interactions (Lomberk et al., 2006b), and to mediate RNA binding (Maison et al., 2011; Muchardt et al., 2002). Most importantly, these modifications and/or interactions were shown to influence the chromatin localisation of individual HP1s.

The aim of my PhD was to investigate the mechanisms that target HP1 proteins to specific sites in the genome. On the one hand, I focused on the role of RNA binding in this process, and my most significant findings are highlighted in chapter 1.1. In Chapter 1.2, I briefly summarize my second line of research, which resulted in the discovery of a novel HP1-containing protein complex that is targeted to euchromatin by a transcription factor. A manuscript describing my findings in detail can be found in the appendix of this thesis.

## 1.1. RNA binding properties of mammalian HP1 homologs

*(Unpublished observations)*

Veronika Ostapcuk[1,2], Claudia Keller[1,2], Ricardo Aidaxo[2,3], Morgan Callon[2,3], Antoine Clery[4], Mauro Zurini[5], Sebastian Hiller[2,3], Beat Fierz[6], Marc Bühler[1,2]

[1] FMI, Maulbeerstrasse 66, 4058 Basel, Switzerland

[2] University of Basel, Petersplatz 10, 4003 Basel, Switzerland

[3] Biozentrum, Klingelbergstrasse 50/70, 4056 Basel, Switzerland

[4] Novartis Pharma AG, 4002, Basel, Switzerland

[4] ETH, IMBB, Otto-Stern-Weg 5, 8093 Zürich, Switzerland

[6] EPFL, ISIC, CH B3 485, 1015 Lausanne, Switzerland

### 1.1.1. Introduction

The conventional view of heterochromatin as a silent rigid structure inaccessible to the transcriptional machinery has been gradually challenged over the two last decades. Firstly, the rapid development of genome-wide sequencing technologies revealed that transcription of the genome is more widespread than originally expected (Birney et al., 2007; Djebali et al., 2012; Kapranov et al., 2007), and secondly, findings from different organisms revealed that heterochromatic transcripts can, in fact, be required for efficient silencing (Aravin et al., 2001; Deng et al., 2009; Matzke et al., 2009; Volpe et al., 2002). The process of RNA-mediated heterochromatin silencing has been most extensively studied in *S. pombe*, where it was found that deletion of RNA interference (RNAi) components results in the loss of H3K9 methylation and HP1 binding, and consequently, upregulation of centromeric transcripts (Volpe et al., 2002). In addition, it was discovered that silencing need not occur at the transcriptional level by restricting access of RNA polymerase, or post-transcriptionally by degradation of transcripts in the nucleoplasm, but can also occur co-transcriptionally. In co-transcriptional gene silencing RNA degradation is induced by the chromatin context from which the transcript is produced, and the transcript itself is required for silencing (Bühler et al., 2006).

Most recently, a heterochromatin-dependent transcriptional machinery was described in *Drosophila*, another well-studied model organism of RNA-mediated heterochromatin

silencing (Andersen et al., 2017). The study revealed that the core transcription machinery can be recruited to heterochromatin by the HP1 paralog, Rhino/HP1d, which results in the initiation of transcription at Rhino bound regions instead of at TSSs that are transcriptionally inaccessible. Moreover, the HP1 proteins, which were assumed to act as a static "glue" of heterochromatin, were found to be highly mobile (Cheutin et al., 2003, 2004; Festenstein et al., 2003). The dynamic nature of HP1s was further reinforced by the finding that HP1s bind centromeric and telomeric transcripts, and that abolishing these interactions impairs HP1 localisation patterns (Deng et al., 2009; Keller et al., 2012; Maison et al., 2011; Muchardt et al., 2002).

Mammalian HP1 homologs were shown to be recruited by RNA transcripts to specific heterochromatic loci. All three homologs, HP1α, HP1β and HP1γ, were shown to bind RNA by an RNA-immunoprecipitation assay (RIP), an electrophoretic mobility shift assay (EMSA) and/or North-western assay, but only the HP1α-RNA interaction, which appears to be the strongest, has been characterised further (Maison et al., 2011; Muchardt et al., 2002). The study by Muchardt *et al*. found that the hinge region and its positive residues are required for RNA binding *in vitro*, and that this interaction is necessary for HP1α localization to pericentric heterochromatin in mouse fibroblast cells. The more recent work by Maison *et al*. made very similar observations, but in addition found that the SUMO-modified HP1α protein can specifically bind to the major satellite transcripts, which then guide *de novo* targeting of HP1α to the pericentric domains.

A previous study from our laboratory focusing on the *S. pombe* HP1 homolog, Swi6, revealed that Swi6 binds heterochromatic RNAs through its positively charged hinge region, and hands them over to the RNA degradation machinery (Fig. 9; Keller et al., 2012). Interestingly, RNA binding changes the conformation of the Swi6 chromodomain, affecting its ability to bind to the H3K9me mark, and eventually leads to the eviction of Swi6 from the heterochromatin (Fig. 9). The follow up study by Keller *et al.* found that abolishing the RNA binding of Swi6 *in vivo* results in the spreading of Swi6 and H3K9me into neighbouring euchromatic regions (Keller et al., 2013).

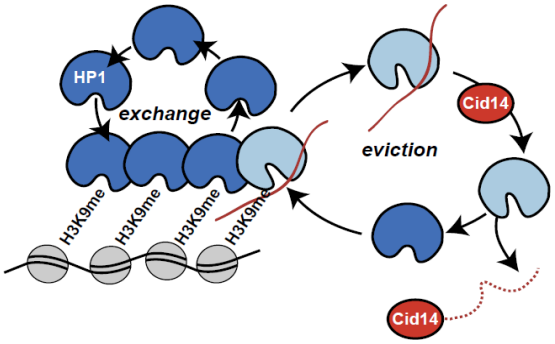



**Figure 9. Model for HP1$^{Swi6}$-mediated degradation of heterochromatic RNAs.**

HP1$^{Swi6}$ (dark blue) dynamically exchanges within H3K9me nucleosomes (grey). Transcribed RNAs (red) associate with HP1$^{Swi6}$ and cause a conformational change (light blue), which evicts HP1$^{Swi6}$ from heterochromatin due to the competition between RNA and H3K9me. The RNA is passed on to the RNA decay pathway (Cid14 (red)), and the HP1 conformation is restored. This mechanism serves as a checkpoint at the borders between heterochromatin and euchromatin, preventing heterochromatin spreading into euchromatic regions. (Keller et al., 2012, 2013)

### 1.1.2. Objective

Because the above findings depict different roles of RNA in modulating HP1 functions between mammals and fission yeast (recruitment to and eviction from chromatin, respectively) we decided to revisit the RNA-mediated regulation of the mammalian HP1 homologs. The objective of this project was to dissect the RNA binding properties of the mammalian HP1 proteins *in vitro* with the help of the methods established in our laboratory for Swi6, and in the following step, apply the results we obtained to study mouse embryonic stem cells, investigating the roles of HP1-RNA interactions genome wide. The following chapter summarises our most important *in vitro* findings.

### 1.1.3. Results

First, I set out to recapitulate the previously published RNA binding by HP1 proteins *in vitro*. To do so, I overexpressed N-terminal Glutathione-S-transferase (GST)-fusions of mouse HP1α, HP1β and HP1γ, respectively, in *Escherichia coli*, and captured the fusion proteins on Glutathione sepharose. I cleaved off the GST tag, and purified the recombinant HP1 proteins by anion exchange chromatography (AEC) and size exclusion chromatography (SEC). Finally, I used the recombinant HP1 proteins to perform EMSA assays with fluorescently labelled RNA probes. In these assays, the recombinant HP1 proteins did not bind RNA at the expected protein:RNA ratios (data not shown). RNA probes of a different sequence and length were assayed as well, but with the same result. Therefore, I decided to use NMR spectroscopy, which is a more sensitive method to test RNA binding. We purified $^{15}$N labelled HP1s and measured $^{1}$H-$^{15}$N-HSQC (2D) NMR spectra, followed by the titration of unlabelled RNA. However, this method was also unable to detect any significant RNA interaction (data now shown). To tackle the discrepancy between our observations and the previously published data, I thoroughly investigated the methodology of the prior publications, and found that their results were obtained using recombinant GST-HP1 fusion proteins (Maison et al., 2011; Muchardt et al., 2002). The GST tag has a similar size as HP1 proteins and on its own forms a dimer (Parker et al., 1990); it is, therefore, foreseeable that it can significantly influence the biophysical properties of HP1s.

In the next experiment, I decided to test in parallel the RNA binding of the untagged and GST-tagged HP1 proteins, and compare the observed binding affinity, if any, with the Swi6 ortholog. For this, I purified recombinant HP1α, HP1β, HP1γ and Swi6 (Fig. 10A), their GST fusions (Fig. 10B), and the GST tag as a control (Fig. 10C). For Swi6, the mutation of the positively charged hinge residues, R and K, to alanine (A) drastically reduced the RNA binding as compared to the wild-type protein (Keller et al., 2012). I designed the same mutants for the HP1 proteins, HP1α-KR13A, HP1β-KR12A, HP1γ-KR11A, and purified their GST fusions as well (Fig. 10C).

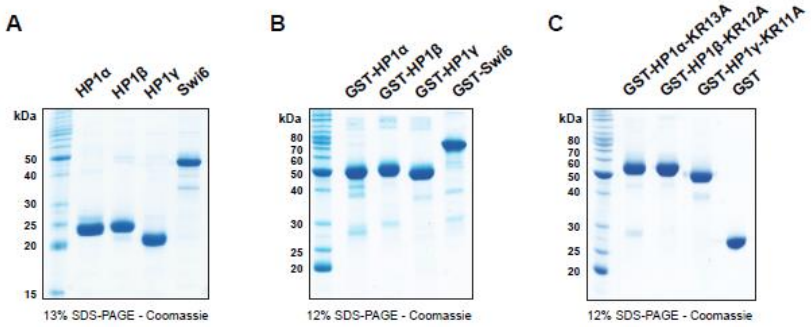



**Figure 10. Recombinant mammalian HP1 proteins.**

(A) Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) of the recombinant mouse HP1α, HP1β, HP1γ, and fission yeast Swi6. (B) SDS-PAGE of the recombinant N-terminal Glutathione-S-transferase (GST) fusions of HP1α, HP1β, HP1γ and Swi6. (C) SDS-PAGE of the recombinant GST-HP1α-KR13A, GST-HP1β-KR12A, GST-HP1γ-KR11A mutants (all positively charged residues in hinge, K and R, mutated to alanine, A), and GST tag.

(A-C) All proteins were overexpressed in *Escherichia coli*, and captured on Glutathione sepharose. The GST tag was cleaved off (A) or left intact (B, C), proteins were subsequently purified by anion exchange chromatography (AEX) and size exclusion chromatography (SEC).

The above-purified recombinant proteins were assayed for RNA binding by EMSA using a fluorescently labelled AU-rich RNA probe. In accordance with our previous observations, the untagged HP1 proteins did not bind the RNA probe (Fig. 11A-C), although RNA mobility was slightly affected by HP1α at the highest protein:RNA ratios (200:1 to 400:1) (Fig. 11A). For comparison, Swi6, which has a rather weak RNA binding constant of approx. 40 μM (Keller et al., 2012), interacted with RNA already at a 5:1 ratio, and fully bound at a 50:1 ratio (Fig. 11D). The second step was to assay the GST-tagged HP1 proteins. The GST-tag had no effect on HP1γ but slightly enhanced the HP1β-RNA interaction, and most importantly, it induced RNA binding by HP1α (Fig. 11A-C, second lane). Interestingly, the GST tag reduced the affinity of Swi6 to RNA, and on its own had no influence on RNA mobility (Fig. 3D, second and third lane). Next, I wanted to test whether this is a GST-induced artefact or an HP1-specific interaction. To do so, I assayed the RNA binding of the GST-HP1-KRxA mutants. The previously observed RNA binding by GST-HP1α was greatly reduced in the GST-HP1α-KR13A mutant (Fig. 3A, third lane). Not surprisingly, the GST-HP1β-KR12A and GST-HP1γ-KR11A mutants showed similar behaviour as their wild-type proteins (Fig. 11B, 11C; third lane). In summary, these results demonstrated that HP1α has the potential to interact with RNA via its positively charged hinge region.

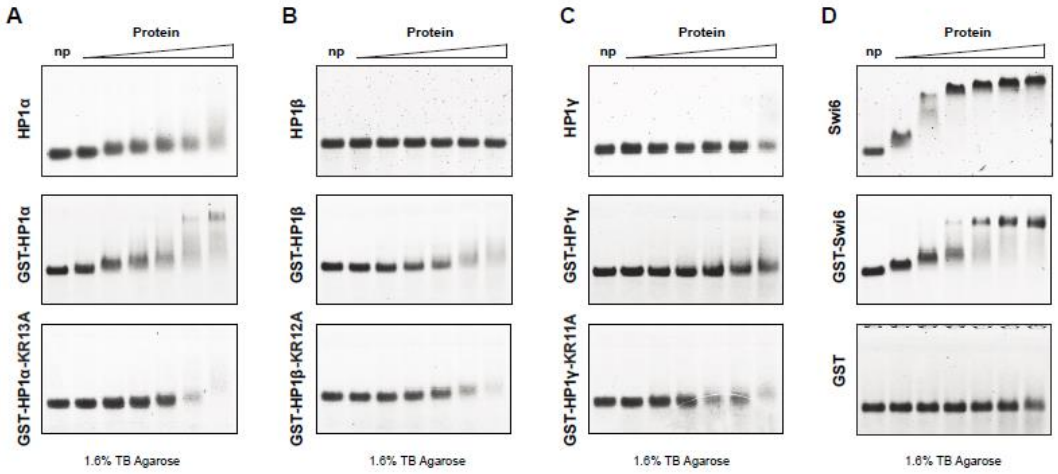



**Figure 11. Recombinant GST-HP1α fusion protein binds AU-rich RNA.**

(A) Electrophoretic mobility shift assay (EMSA) of the fluorescently labelled AU-rich RNA probe and recombinant HP1α, GST-HP1α or GST-HP1α-KR13A.

(B) EMSA of the fluorescently labelled AU-rich RNA probe and recombinant HP1β, GST-HP1β or GST-HP1β-KR12A.

(C) EMSA of the fluorescently labelled AU-rich RNA probe and recombinant HP1γ, GST-HP1γ or GST-HP1γ-KR11A.

(A-C) RNA probe was labelled with fluorescein-UTP by *in vitro* transcription. Protein-RNA mixtures were separated on 1.6% TB agarose gels, and the fluorescent signal was detected using a typhoon scanner. All wells contained constant amounts of the RNA probe. First wells, np, had no recombinant protein. Second to seventh wells contained increasing concentrations of the recombinant proteins; protein:RNA ratio of 5:1, 25:1, 50:1, 100:1, 200:1, 400:1.

Since both, the GST tag and the HP1 chromoshadow domain, can independently dimerise (Cowieson et al., 2000; Parker et al., 1990), I hypothesized that the GST-HP1α fusions can form alternative dimers or oligomers, which change the conformation of HP1α allowing the RNA interaction. In this theoretical conformation, the GST dimerisation connects two chromodomains, while their chromoshadow domains interact together, forming a closed dimer. Alternatively, the chromoshadow domains from one GST-HP1α dimer can interact with the chromoshadow domains from the neighbouring GST-HP1α dimers, forming an oligomer. Interestingly, this is the exact description of the proposed HP1 binding to the nucleosomes (Canzio et al., 2014). In this analogy, the GST dimer represents the two H3K9me-tails of a single nucleosome. I had, in fact, considered the effect of the H3K9me3 binding previously, and measured the RNA affinity of the untagged HP1α bound to the H3K9me3 peptide by surface plasmon resonance (SPR). This experiment did not reveal any

significant RNA binding (data not shown), possibly because a nucleosome with two HP1 binding sites is required to stimulate the RNA interaction. Therefore, I aimed to reconstitute the H3K9me nucleosomes *in vitro* in collaboration with the laboratory of Dr. Beat Fierz from EPFL.

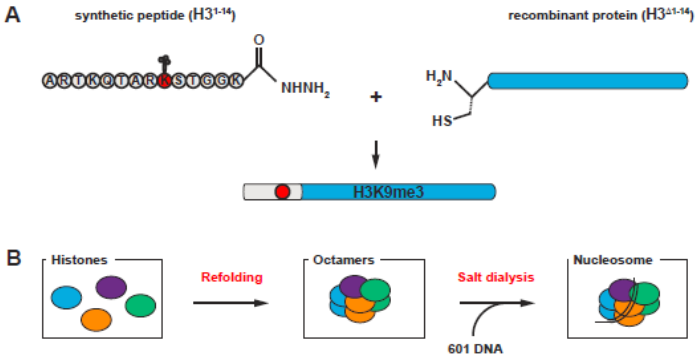



**Figure 12. Schematic representation of reconstitution of semisynthetic H3K9me3 nucleosomes.**

(A) H3 histone lacking its N-terminus ($H3^{d1-14}$) is produced recombinantly, by overexpression and purification from *E. coli.* The H3 N-terminal tail ($H3^{1-14}$) is chemically synthesised, including the trimethylation at the lysine 9 (red dot). The two components are then chemically ligated together, forming H3K9me3 histone.
(B) Unmodified histones, H2A, H2B and H4, are produced recombinantly from *E. coli*, and in combination with H3K9me3, used to form a histone octamer. In the last step, the 601 DNA is introduced to the octamer, ultimately forming H3K9me3 nucleosome.

We followed the general reconstitution protocol established by Luger *et al.* for the production of unmodified nucleosomes (Fig. 12B; Luger et al., 1997), but used a semisynthetic approach for the reconstitution of H3K9me3 nucleosomes established by the Fierz laboratory (Fig. 12A-B). In the first step, we produced recombinant core histones and synthetic $H3K9me3^{1-14}$ peptide (Fig. 13A-B). The K9me3-modified H3 histone was synthesised by the chemical ligation of $H3K9me3^{1-14}$ and the N-terminally truncated recombinant $H3^{d1-14}$ histone. Next, we combined the individual core histones at equimolar ratios, and purified the resulting histone octamers by SEC (Fig. 13C). In the last step, 601 DNA was assembled on octamers with the help of salt gradient dialysis, resulting in the formation of nucleosomes (Fig. 13D). The purity and size of the H3K9me3 nucleosomes was confirmed by the SEC-MALLS method (multiangle light laser scattering), which detected a uniform protein mixture of the approx. size of 205 kDa (Fig. 13E). This is in agreement with the previously determined size of a single unmodified nucleosome, which is 206 kDa (Luger

et al., 1997). In summary, we produced highly purified H3K9me0 and H3K9me3 nucleosomes.

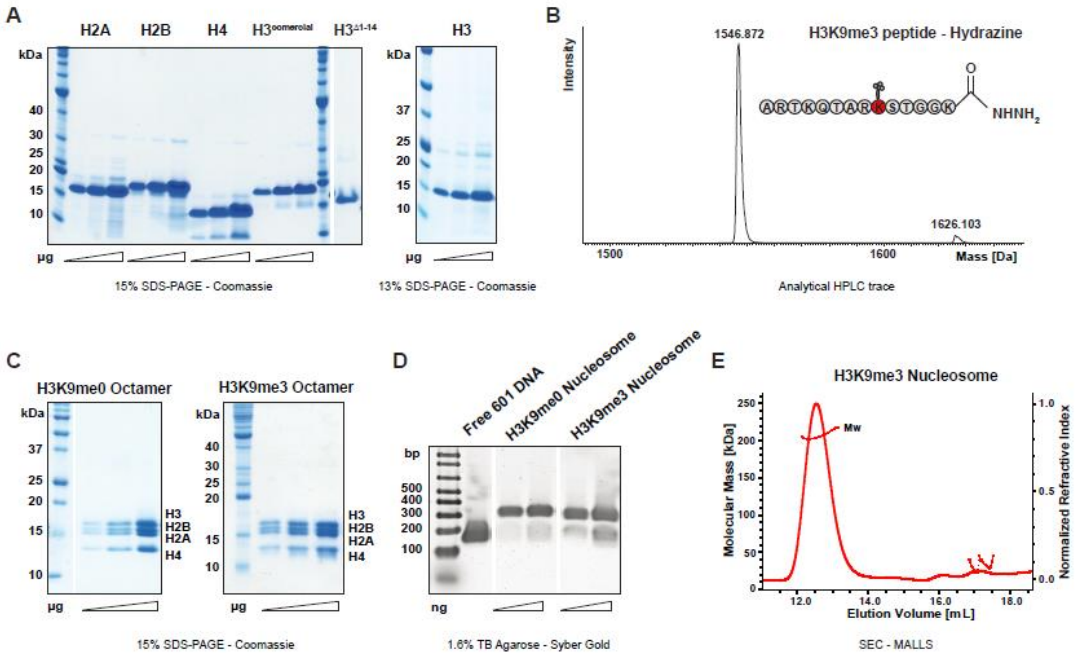



**Figure 13. Reconstitution of the H3K9me0 and H3K9me3 nucleosomes.**

(A) SDS-PAGE of the recombinant H2A, H2B, $H3^{d1-14}$ and H3 proteins. Commercial H3 histone was used as a loading control. B. Analytical HPLC trace of the synthetically produced H3K9me3 peptide used for synthesis of the H3K9me3 histone.

(C) SDS-PAGE of the SEC-purified histone octamers containing unmodified H3 (H3K9me0) or K9me3-modified H3.

(D) DNA agarose gel of the final H3K9me0 and H3K9me3 nucleosomes. Free 601 DNA was loaded as a control.

(E) SEC-MALLS (Multiangle light laser scattering) profile of the reconstituted H3K9me3 nucleosome.

Next, we tested the RNA binding by HP1α in the context of nucleosomes by the EMSA assays. The recombinant HP1α protein bound to nucleosomes induced an RNA shift at similar protein:RNA ratios as the GST-HP1α fusion (Fig. 14, Fig. 11A); we did not observe the same RNA shift with HP1α alone or with HP1α bound to the H3K9me3 peptide (Fig. 14). Interestingly, HP1α RNA binding was enhanced by nucleosomes irrespectively of the K9 methylation status (Fig. 14). The RNA interaction was abolished in the HP1α hinge mutant, HP1α-KR13A (Fig. 14). Because we did not observe the same RNA shift by HP1β, we concluded that enhancement of the RNA binding by nucleosome is HP1α-specific (Fig. 14).

In agreement with the previously published data, the affinity of Swi6 to RNA was reduced in the context of nucleosomes due to the competition between the RNA and nucleosome binding (Fig. 14; Keller et al., 2012).

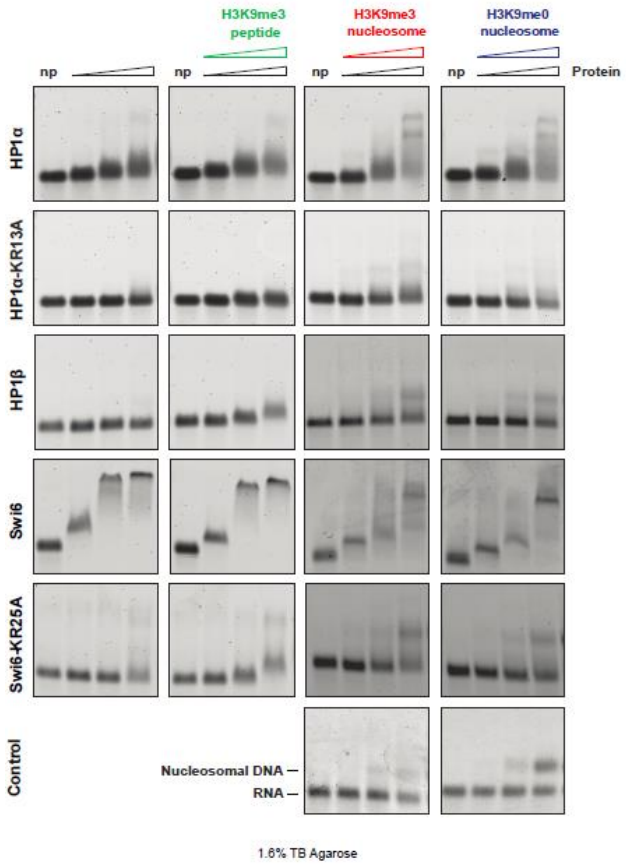



**Figure 14. Enhancement of the HP1α RNA binding in the context of the nucleosome.**

EMSAs assaying the binding of the fluorescently labelled AU-rich RNA probe by the recombinant HP1α, HP1α-KR13A, HP1β, Swi6 or Swi6-KR25A proteins in the context of H3K9me3 peptide, H3K9me0 nucleosome or H3K9me3 nucleosome. RNA probe was labelled with fluorescein-UTP by *in vitro* transcription. Protein-RNA complexes were separated on 1.6% TB agarose gels, and the fluorescent signal was detected using a typhoon scanner; the typhoon scanner also detected unlabelled nucleosomal DNA signal. All wells contained constant amounts of the RNA probe. First wells, np, had no recombinant proteins; all the second to forth wells contained increasing concentrations of the recombinant proteins, peptides or nucleosomes. First column, protein:RNA ratio of 5:1, 50:1 and 250:1. Second column, protein:peptide:RNA ratio of 5:5:1, 50:50:1 and 250:250:1. Third and fourth column, protein:nucleosome:RNA ratio of 5:2.5:1, 50:25:1, 250:125:1. Control EMSA contains only nucleosomes and RNA at the ratio of 2.5:1, 25:1 and 125:1.

Our next goal was to further characterise the nucleosome-HP1α-RNA interaction by NMR. We purified $^{15}$N labelled HP1α and unlabelled H3K9me3 nucleosomes, and attempted to measure the 2D NMR spectra of the complex, followed by RNA titration. When we

combined HP1α with H3K9me3 nucleosomes, we observed massive protein precipitation, which made it impossible to measure NMR spectra. Instead, we inspected our sample by EM. This revealed uniformly shaped structures, which are suggestive of a phase separation of the nucleosome-HP1α complex (data not shown). Indeed, DNA was recently shown to promote the phase separation of HP1α (Larson et al., 2017). However at that point, we were not able to continue with the biochemical characterisation of the nucleosome-HP1α-RNA interaction. Therefore, I considered creating the endogenous HP1α-KR13A mutation in mouse ES cells, comparing its behaviour with the wild-type protein. However, before initiating this experiment, I wanted to perform one last control *in vitro* experiment, to measure the affinity of the HP1α-KR13A mutant to the H3K9me3 peptide by SPR. This revealed that the affinity of the HP1α-KR13A mutant is about 60-fold weaker compared to the wild type protein (Fig. 15A, 15B). The difference in the binding constants is most likely even more severe in the context of the nucleosomes, as it was previously shown that the hinge linker contributes to the nucleosome binding *in vitro* (Mishima et al., 2013).

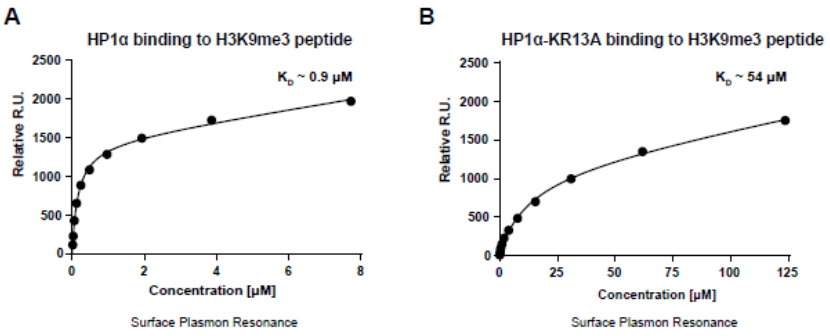



**Figure 15. Intact HP1α hinge is required for efficient H3K9me3 binding.**

(A) Surface plasmon resonance (SPR) measurement of recombinant HP1α affinity to the immobilised H3K9me3 peptide. (B) SPR measurement of recombinant HP1α-KR13A mutant affinity to the immobilised H3K9me3 peptide.

(A, B) Equilibrium dissociation constants ($K_D$) were determined by the SPR software (GE Healthcare). Relative R.U. – relative response units.

The above result discouraged us from creating the endogenous HP1α-KR13A mutation because we would not be able to separate the phenotypes caused by the loss of RNA binding and/or impaired H3K9me3 binding. To overcome this issue, we attempted to assign HP1α

residues by NMR to identify the ones affected by RNA, and eventually use this information for designing an RNA binding-specific mutant. Unfortunately, the instability of HP1α at higher concentrations, which is a requirement for such an experiment, prevented us from determining the exact residues affected by RNA. Finally, we decided to postpone this project until new evidence helping to tackle the HP1-RNA interaction arises.

In summary, our data confirm that HP1α does interact with RNA, while the other two homologs, HP1β and HP1γ do not. The affinity of the interaction is extremely weak *in vitro,* but can be further reinforced in the context of the nucleosome. The HP1α-RNA affinities are similar for GC- or AU-rich substrates (data not shown), which suggests that this is a rather unspecific interaction. RNA binding is mediated by the positively charged hinge linker, and mutation of the KR residues abolishes the interaction. The above results, their implications and possible follow-up experiments, are further discussed in the last chapter (Discussion and Outlook).

### 1.1.4. Collaborations

The NMR measurements were carried out in collaboration with the group of Dr. Sebastian Hiller, namely by Ricardo Adaixo and Dr. Morgan Callon, and in later stages with Dr. Antoine Clery. Dr. Morgan Callon did the EM analysis. The nucleosome reconstitution and the following experiments with nucleosomes were performed in collaboration with Dr. Claudia Keller, Dr. Beat Fierz, and Dr. Mauro Zurini.

### 1.1.5. Methods

Expression and purification of HP1 proteins, EMSA assays, NMR experiments, and SPR measurements were carried out according to (Keller et al., 2012). Nucleosome reconstitution was performed according to (Kilic et al., 2015; Luger et al., 1997).

## 1.2. Transcription factor-mediated recruitment of the HP1 proteins to euchromatin

*(Ostapcuk et al., accepted for publication at Nature; see Appendix for full manuscript)*

Veronika Ostapcuk[1,2], Fabio Mohn[1,5], Sarah H. Carl[1,3,5], Anja Basters[1,5], Daniel Hess[1], Vytautas Iesmantavicius[1], Lisa Lampersberger[4], Matyas Flemr[1], Yukiko Shimada[1], Nicolas H. Thomä[1], Jörg Betschinger[1], Marc Bühler[1,2]

[1]FMI, Maulbeerstrasse 66, 4058 Basel, Switzerland

[2] University of Basel, Petersplatz 10, 4003 Basel, Switzerland

[3] Swiss Institute of Bioinformatics, 4058 Basel, Switzerland

[4] University of Vienna, Dr. Bohrgasse 9, 1030 Vienna, Austria

Prevailing models suggest that HP1 proteins are targeted to chromatin via their chromo domain, which recognizes and binds the H3K9me mark. However, the affinity of the chromo domain to H3K9me is rather low (Canzio et al., 2014), and HP1s show distinct localization patterns in the nucleus (Minc, Courvalin and Buendia, 2000). In addition, previous works have demonstrated that HP1β and HP1γ can also localise to euchromatic regions (Mattout et al., 2015; Smallwood et al., 2012; Vakoc et al., 2005). It is not clear, however, how the homolog-specific targeting is achieved, or what factors and/or mechanisms act as a recruiter of HP1s to euchromatic regions. In this project, I set out to investigate whether the HP1 proteins can be targeted to euchromatic regions by homolog-specific protein interactors. To test this, I exploited the recent advances in genome editing to endogenously tag individual HP1 homologs, and determined their interactomes (Fig. 16, 17). Interestingly, the majority of highly enriched HP1 interactors we observed were DNA-binding proteins (Fig. 17, green).

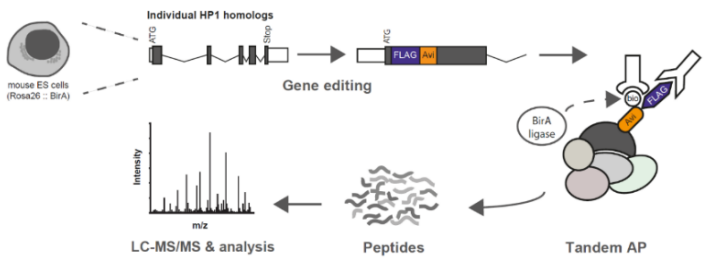



**Figure 16.**

Scheme depicting the endogenous tagging of the individual HP1 genes in mouse ES cells with the FLAG-Avi tag (FLAvi), followed by tandem affinity purification of HP1-containing protein complexes for LC-MS/MS (Liquid chromatography followed by tandem mass spectrometry).

Next, I inspected the individual HP1 interactors to find the one that is (1) a DNA-binder, (2) HP1β- and/or HP1γ-specific, (3) highly enriched, and (4) has a PxVxL HP1-interaction motif. The best candidate, fulfilling all of the above criteria, was a transcription factor, Adnp. I endogenously tagged Adnp, and confirmed its interaction with HP1β and HP1γ. More importantly, the tandem affinity purification (TAP) of tagged Adnp in 500mM NaCl did not disturb its interaction with HP1s or with another highly enriched protein, chromatin remodeller Chd4 (Fig. 18). This indicates that Adnp, HP1β and/or HP1γ, and Chd4 form a stable complex in mouse ES cells – which we named the ChAHP complex.

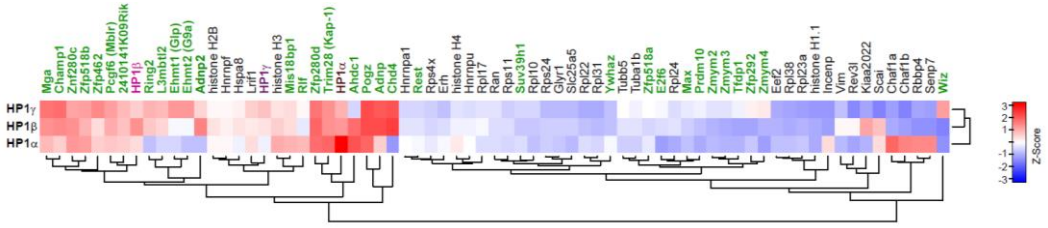



**Figure 17. The HP1 interactome of mouse embryonic stem cells.**

Heat map showing the variation in co-purifying (Z-score) proteins across HP1 isoform-specific TAP-LC-MS/MS experiments. Proteins that were significantly enriched in at least one of the individual HP1 purifications were included in the analysis.

We reconstituted the ChAHP complex *in vitro* from insect cells and dissected the individual interactions. Together with proteome studies from mouse ESCs, this revealed Adnp as a core bridging module interacting with Chd4 and HP1s. Using ChIP-sequencing, we identified over 15 000 ChAHP-bound genomic sites. Unexpectedly, these sites are devoid of H3K9me2 or H3K9me3. Instead, the complex appears to be targeted via a highly conserved DNA motif recognized by Adnp, and deletion of Adnp leads to the depletion of HP1γ binding at the ChAHP sites (Fig. 19A). Similarly, deletion of the Adnp binding motif at the endogenous site delocalizes all of the ChAHP components (Fig. 19B-E).

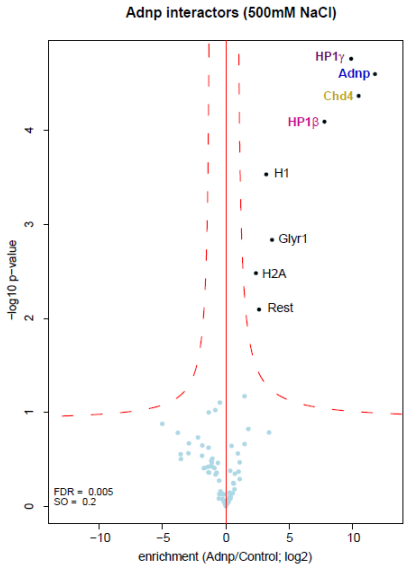



**Figure 18. Adnp, HP1β and/or HP1γ, and Chd4 form a novel complex – ChAHP.**

TAP-LC-MS/MS of endogenously FLAG/Avi-tagged Adnp. Protein purification was performed in the presence of 500mM NaCl. Parental mES cell line serves as background control. n=3 biological replicates (i.e. 3 independent *Adnp*$^{FlagAvi/FlagAvi}$ mES cell lines).

Additionally, deletion of Adnp or HP1s (triple knock-out) leads to the derepression of lineage specification factors bound by ChAHP, suggesting that ChAHP acts as a transcriptional repressor in mES cells. Indeed, we have found that ChAHP binding renders chromatin inaccessible. Ultimately, we investigated the role of ChAHP in Helsmoortel-Van der Aa syndrome, a disease caused by mutations in the *ADNP* gene. We discovered that patient Adnp protein (Adnp$^{Mut}$) fails to interact with HP1s and, consequently, cannot assemble functional ChAHP complexes. The DNA binding of Adnp$^{Mut}$ to the ChAHP genomic sites was not affected, suggesting that the most likely primary cause of Helsmoortel-Van der Aa syndrome is the inability of Adnp$^{Mut}$ to form a functional complex. Finally, we were able to rescue the Adnp$^{Mut}$ defect pharmacologically, restoring complex formation.

In summary, my work revealed that HP1 proteins can be recruited to genomic loci in a DNA sequence-specific, H3K9 methylation-independent, manner via an interaction with the transcription factor Adnp. In addition, I demonstrated that H3K9 methylation is not required for repression of ChAHP target genes, and finally yet importantly, I uncovered the most likely primary cause of the Helsmoortel-Van der Aa syndrome, and based on my results, propose a potential pharmacological treatment.

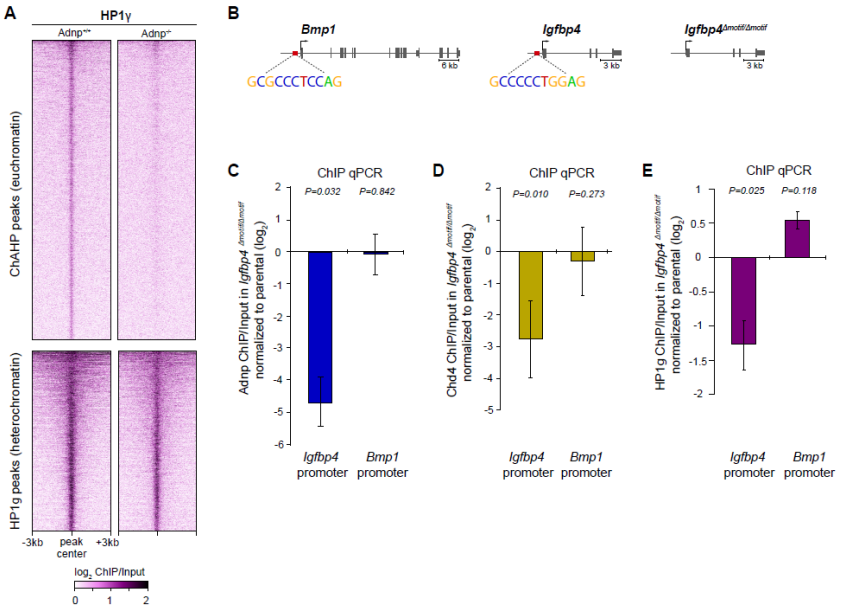



**Figure 19. DNA sequence specifies ChAHP association with euchromatin.**

(A) Heatmap of HP1γ ChIP-seq enrichment in ChAHP peaks (top) or heterochromatic HP1-bound sites (bottom) in *Adnp*$^{+/+}$ and *Adnp*$^{-/-}$ mES cells. Each row represents a 6-kb window centered on the respective peak signal. Rows are sorted by mean ChIP enrichment. Average of n=3 biological replicates (i.e. 3 independent mES cell lines).

(B) Schemes depicting the location of Adnp binding motifs in the *Bmp1* and *Igfbp4* genes (left, middle) and the *Igfbp4* locus with the motif deletion *Igfbp4*$^{\Delta motif/\Delta motif}$ (right).

(C) ChIP-qPCR measuring Adnp enrichments over *Igfbp4* and *Bmp1* promoters in *Igfbp4*$^{\Delta motif/\Delta motif}$ cells compared to parental line. n=3 independent biological replicates.

(D) ChIP-qPCR measuring Chd4 enrichments over *Igfbp4* and *Bmp1* promoters in *Igfbp4*$^{\Delta motif/\Delta motif}$ cells compared to parental line. n=6 independent biological replicates used. (E) HP1γ ChIP-qPCR enrichments over *Igfbp4* and *Bmp1* promoters in *Igfbp4*$^{\Delta motif/\Delta motif}$ cells compared to parental line. n=3 independent biological replicates.

### 1.2.1. Collaborations

The proteomic experiments were carried out in collaboration with the FMI Protein Analysis Facility, namely with Dr. Daniel Hess and Dr. Vytautas Iesmantavicius. The ChAHP complex reconstitution experiments were performed in collaboration with the group of Dr. Nicolas Thomä, namely by Dr. Anja Basters. The bioinformatic analysis of the genome-wide data was performed by Dr. Sarah H. Carl and Dr. Fabio Mohn. Dr. Fabio Mohn performed the chromatin accessibility experiments. All experiments related to Adnp$^{Mut}$ were performed with the help of Lisa Lampersberger, Dr. Yukiko Shimada, Dr. Matyas Flemr and Dr. Fabio Mohn. Finally, Dr. Jörg Betschinger advised on the stem cell related experiments.

# *Discussion and Outlook*

Traditionally, HP1 homologs, as their name suggests, are perceived as fundamental constituents of heterochromatin, to which they are recruited via the H3K9me mark. Their role, based on the widely accepted model, is to facilitate spreading of H3K9me along the locus, and to compact heterochromatin, rendering it physically inaccessible. This static perception of HP1s holds true although the general view of heterochromatin is being reshaped by emerging evidence about its highly dynamic nature. HP1 proteins themselves were found to be highly mobile molecules with the ability to interact with RNA, and to bind to euchromatin in a homolog-specific manner. HP1 binding to euchromatin is usually explained by the presence of the H3K9me mark, but the mark itself, as it lacks specificity, cannot act as a primary targeting mechanism; if it would, all three homologs would be recruited to the same sites across the genome, which is not the case. In addition, HP1γ was detected at sites lacking H3K9me (Smallwood et al., 2012). These findings raise several questions about HP1 recruitment and function. What determines HP1 targeting to chromatin? What gives specificity to individual homologs? What is the extent of their redundancy? What is their role in euchromatin that is not marked by H3K9me?

During my PhD, I focused on H3K9me-independent pathways of HP1 recruitment to chromatin, which appears to use the hinge and chromo-shadow domains. Notably, the chromo domain, which is required for H3K9me binding, is on its own insufficient for targeting to constitutive heterochromatin (Muchardt et al., 2002; Smothers and Henikoff, 2001). This indicates that even H3K9me-dependent recruitment requires other HP1 domains and/or additional factors.

All three HP1 domains (the N-terminal extension with the chromo domain, the hinge, and the chromo-shadow with the C-terminal extension) were shown to play a functional role in determining the chromatin-binding patterns of HP1s. In the following chapters, I describe and discuss the roles of the individual domains and their targeting mechanisms based on the current literature, using this as a foundation for the discussion of the main findings of my PhD.

In the first chapter, I concentrate on H3K9me-dependent recruitment, which was not my primary focus, yet I believe that my work contributes to our understanding of this pathway. In the second chapter, I discuss the role of the hinge domain and RNA in regulating HP1 function and chromatin binding, which I investigated using biochemical approaches. In the

third chapter, I focus on the main finding of my PhD – transcription factor-mediated recruitment of HP1s to euchromatin, a novel recruitment pathway that is independent and not associated with the H3K9me mark. In the last chapter, concluding remarks, I summarise this discussion, and present my perspective on HP1 recruitment to chromatin.

# 1. H3K9-methylation dependent recruitment of HP1s

In the following chapters, I am going to review H3K9-methylation dependent targeting of HP1s, and provide literature-based evidence, indicating that CD-mediated HP1 binding to H3K9me-marked chromatin depends also on other HP1 domains and on additional factors. In later chapters, I focus on mammalian HP1 homologs and their redundancy, and discuss my results that are relevant to this pathway.

## 1.1. Contributions of HP1 domains in mediating the H3K9me interaction

The targeting of HP1s to H3K9-methylation-marked chromatin requires the cooperative action of all three HP1 domains. The chromo domain binds the N-terminal H3 tail carrying K9 methylation and the C-terminal histone-fold domain of H3 (Bannister et al., 2001; Jacobs et al., 2001; Lachner et al., 2001; Nielsen et al., 2001). The hinge region contacts nucleosomal DNA (Canzio et al., 2013; Mishima et al., 2013), and the chromo-shadow domain interacts with HKMTases (Nozawa et al., 2010; Schotta et al., 2002; Yamamoto and Sonoda, 2003). Mutation of V22 in the HP1α chromo domain, which reduces its affinity for the ARKS motif, abolishes interactions with the nucleosome *in vitro* (Nielsen et al., 2001). This indirectly indicates that the role of the hinge is to stabilise the binding mediated by the CD, and on its own, it cannot form a stable interaction with the nucleosome. Similarly, the CSD creates stabilising contacts with the nucleosome, but on its own is not sufficient for the interaction (Canzio et al., 2013; Mishima et al., 2013; Nielsen et al., 2001).

The cumulative power of the individual domains in stabilising the interaction with H3K9 methylation is especially apparent when we compare the affinity of the isolated CD for the H3K9me peptide with the affinity of the full-length (FL) HP1 for the nucleosome. *In vitro* the stability of the FL-HP1-nucleosome complex is 100-fold stronger than of the CD-H3K9me3-

tail complex, with the affinity of FL-HP1 for the nucleosome being 100 nM (Canzio et al., 2013).

In addition, we should not omit the interaction with HKMTases that can potentially increase the affinity for H3K9 methylation-marked chromatin. Indeed, *in vitro* binding of the fly HP1α ortholog, HP1a, to nucleosomes is stimulated by the presence the of HKMTase, Su(var)39 (Eskeland et al., 2007). This effect is independent of the enzymatic activity of Su(var)39, but depends solely on the interaction between HP1 and Su(var)39.

### 1.2. Interdependency of HP1 and HKMTases

Based on the above presented data, it is not surprising that HP1 recruitment to heterochromatin requires the presence of the corresponding HKMTase, and this dependency is highly conserved across different systems, from fission yeast (Bannister et al., 2001; Nakayama et al., 2001), to flies (Schotta et al., 2002), to mammals (Lachner *et al.*, 2001; García-Cao *et al.*, 2004; Tachibana *et al.*, 2005). What is surprising, however, is that the proper localisation of HKMTases also depends on HP1 proteins. Deletion of HP1a results in mislocalisation of Su(var)39 outside of the pericentric heterochromatin into the euchromatic regions, where it ectopically methylates H3K9 (Schotta et al., 2002). Similarly, in *S. pombe* deletion of its HP1 ortholog, Swi6, results in the spreading of H3K9 methylation into euchromatin (Stunnenberg et al., 2015).

These results imply that the interaction of HKMTases with the HP1 chromo-shadow domain restricts their binding to heterochromatin. Such an interdependency model of the HP1-HKMTase association is further supported by domain swapping experiments, in which the HP1a chromo domain was replaced by the chromo domain of Polycomb (Pc) (HP1a-Pc) (Schotta et al., 2002). The Pc chromo domain has specificity for H3K27me3, and consequently, binds different chromatin regions than HP1s (Fischle et al., 2003). The chimeric HP1a-Pc protein binds to both, HP1 and Pc chromatin sites (Fischle et al., 2003; Platero et al., 1995), and more importantly, it has the ability to recruit Su(var)39 and endogenous HP1s to Pc sites (Platero et al., 1995; Schotta et al., 2002). This indicates that CSD-mediated interactions are strong enough to recruit HKMTases to ectopic sites.

### 1.3. Mammalian HP1 homologs

All three mammalian HP1s, HP1α, HP1β and HP1γ, bind H3K9me *in vitro*, and in cells localise to H3K9me-marked chromatin (García-Cao et al., 2004; Lachner et al., 2001; Tachibana et al., 2005). Biochemical experiments and microscopic observations, however, point towards certain differences between the homologs. Unlike HP1α, the HP1β CSD does not form any contacts with the methylated nucleosome, which results in similar affinities of HP1β for the H3K9me peptide and the nucleosome (both in μM range); suggesting it may have a weaker affinity compared to HP1α *in vivo* (Mishima et al., 2013; Munari et al., 2012). The NTE of HP1α contains a tandem array of serines, which upon phosphorylation increase the affinity of HP1α to H3K9me (Hiragami-Hamada et al., 2011), while phosphorylation of HP1γ reduces its affinity to H3K9me-marked chromatin (Lomberk et al., 2006b). This clearly demonstrates that the affinity of the CD to H3K9me can be modulated, which allows for differential regulation of individual homologs *in vivo.*

More recently, the phosphorylation (P) of the NTE of HP1α was shown to promote formation of phase-separated droplets (Larson et al., 2017). The P-NTE-HP1α protein has the ability to sequester H3K9-methylated nucleosomes in these droplets without an effect on solvation, which means that reactions and interactions can take place in the HP1α phase. Larson *et al.* also tested the oligomerisation potential of P-NTE-HP1α compared to HP1α or HP1β, finding that only the P-NTE-HP1α can form higher order oligomers, a feature required for heterochromatin compaction. It would be intriguing to test if the other two HP1 homologs can be phase-separated by P-NTE-HP1α. Finally, the authors proposed that *in vivo* this could serve as a repressive mechanism in which heterochromatin is physically sequestrated in an environment permissive only for repressive machinery or alternatively, for factors responsible for dissolving the phase droplets.

Microscopic analyses of HP1 localisation patterns in mouse and human cells revealed homolog-specific differences as well. HP1α and HP1β predominantly bind to the pericentric heterochromatin, yet some staining in euchromatin was also observed (Lachner et al., 2001; Minc et al., 1999; Tachibana et al., 2005); while HP1γ appears to be mostly dispersed in euchromatin (Minc et al., 2000; Tachibana et al., 2005). Interestingly, deletion of *Ehmt2*, which results in a loss of H3K9me2 in silent euchromatin, leads to delocalisation of all HP1

homologs to pericentric heterochromatin (Tachibana et al., 2005). In addition, all three homologs were detected by ChIP at other H3K9me-marked regions, telomeres and retrotransposons (García-Cao et al., 2004; Matsui et al., 2010).

Altogether, this suggests that despite differences in the HP1 localisations, they all have the potential to bind to H3K9 di- and tri-methylated chromatin, with specificity arising from the interactions with HKMTases or other associated factors, and the affinity being modulated by the small differences in HP1 sequence and PTMs. Support of this model, in addition to the already discussed findings, also comes from the fact that in *Suv39h* dn cells the H3K9me3 mark is depleted from pericentric regions while H3K9me2 persists, yet HP1 binding is lost. The requirement of additional factors, however, does not diminish the importance of the mark; it rather demonstrates the versatility and complexity of the underlying regulation.

The above findings suggest that HP1α has the strongest affinity to H3K9me and has a non-redundant function in heterochromatin compaction, but surprisingly *Cbx5*$^{-/-}$ mice, unlike *Cbx3* or *Cbx1* mutants, have no observable phenotype (Aucott et al., 2008; Brown et al., 2010). This can indicate that (a) HP1β and HP1γ may have H3K9me-independent roles that are essential for mouse development; (b) HP1β and/or HP1γ use alternative mechanisms in heterochromatin compaction *in vivo*, and this is sufficient to compensate for the lack of HP1α; (c) the unique HP1α properties in H3K9me binding and oligomerisation are not essential; or (d) combination of the previous options. In this regard, it would be interesting to test if double knock-out mice, e.g. *Cbx5*$^{-/-}$ *Cbx1*$^{-/-}$ have more severe, or earlier, phenotypes than *Cbx1* single KO mice.

### 1.3.1. Redundancy of mammalian HP1 homologs in mouse ES cells

As apparent from the previous paragraphs, our understanding of HP1 function comes mostly from biochemical experiments and genetic studies combined with immunofluorescence or single locus studies (e.g. ChIP). Genome-wide experiments are therefore required to further expand our knowledge about the roles of HP1s, and especially to study their redundancy. During my PhD, I took advantage of the recent developments in genome editing to endogenously tag individual HP1 homologs in mouse ES cells, and determined their genome-wide occupancy. Our analysis focused on the ChAHP-bound regions, which is discussed separately in chapter 3 of this section, and we have not yet fully

exploited the potential of our datasets. Therefore, I propose to compare the profiles of all three HP1s to determine homolog-specific and redundant binding sites across the whole genome, and in the second step, determine the binding of individual homologs in the absence of the other two to study their redundancy. Up to now, I created FLAvi-HP1β in *Cbx3*$^{-/-}$, and FLAvi-HP1γ in *Cbx1*$^{-/-}$ and performed ChIP-seq in these cell lines. My results are very encouraging, as we see complementation of binding in the corresponding cell lines.

Next, I propose to perform ChIP-seq of H3K9me2 and H3K9me3 in wild-type and HP1 conditional/straight KO cells that I generated during my PhD (Table 2); testing if methylation spreads outside of the normal regions or alternatively, if it is lost, and what is the extent of HP1 redundancy in this process.

**Table 2. List of HP1 knock-out mouse ES cells lines generated during my PhD**

| Single KOs | Double KOs | Triple KO |
|---|---|---|
| *Cbx5*$^{fl/fl}$ | *Cbx5*$^{fl/fl}$ *Cbx1*$^{-/-}$ | *Cbx5*$^{fl/fl}$ *Cbx3*$^{fl/fl}$ *Cbx1*$^{-/-}$ |
| *Cbx1*$^{-/-}$ | *Cbx5*$^{fl/fl}$ *Cbx3*$^{fl/fl}$ | |
| *Cbx3*$^{fl/fl}$ | *Cbx3*$^{fl/fl}$ *Cbx1*$^{-/-}$ | |

Thus far, we analysed only expression profiles of these cell lines by RNA-seq, and the results point towards significant redundancy of the HP1 homologs in mouse ES cells. This is especially evident when looking at transcript levels of endogenous retrovirus (ERVs), a subclass of long terminal repeat (LTR) retrotransposons occupying approximately 10% of the mouse genome (Chinwalla et al., 2002). Based on their similarity, ERVs are divided into three classes: I, II and III, which are all repressed by the known HP1 interactors. Classes I and II are silenced by the H3K9me3-depositing Trim28-Setdb1 pathway, and class III is silenced by the H3K9me2-depositing Ehmt1-Ehmt2 pathway (Maksakova et al., 2013).

Unexpectedly, the HP1 proteins were previously shown to be dispensable for the silencing of class I and II (Maksakova et al., 2011), and cause only mild derepression of the class III elements (Maksakova et al., 2013). Our datasets, however, reveal that HP1 proteins are essential for the silencing of ERVs. The single HP1 KOs and double KOs indeed have no effect, but loss of all three HP1s results in the upregulation of class II and III ERVs (Fig. 20). The functional redundancy of the HP1 homologs explains the discrepancy with the above-mentioned reports, in which single HP1 KOs or triple HP1 short interfering (si) RNA-induced

knock-downs (KD) were tested. For the triple HP1 KD, we have to take into account that the residual levels of the individual homologs can have a cumulative effect, and that this can be potentially sufficient for silencing, as is the case for the double KOs. In addition, the redundancy of the HP1 homologs can be seen from the effects of their depletion on the proliferation of mES cells. Single and double KOs do not show significant changes in growth while triple KO cells stop proliferating 6-8 days after treatment with 4-hydroxytamoxifen (data not shown).

Lastly, I propose to differentiate ES cells, and map the changes in HP1 behaviour during this process. Apart from the experiments focusing on the HP1 redundancy, it would be interesting to differentiate the single KOs into lineages corresponding to their mouse KO phenotypes: germ cell like cells (PGCL) for HP1γ deficient cells (Hayashi et al., 2011; Ohta et al., 2017), and neurons for HP1β deficient cells (Bibel et al., 2004). These experiments can contribute to our understanding of the homolog-specific roles, and potentially explain the striking differences in the mouse phenotypes of the individual HP1 KOs.

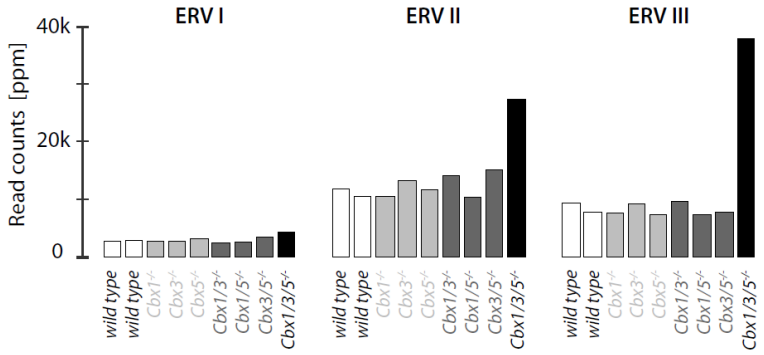



**Figure 20. Redundant role of HP1 proteins in the silencing of ERV elements.**

Bar graph displaying RNA-seq reads mapping to the ERV class I, II, and III elements. Counts were normalized to 1 million genome-mapping reads per library. *Cbx1*, *Cbx3*, and *Cbx5* genes encode for HP1α, HP1β, and HP1γ proteins, respectively. All mutant cell lines were derived from the same parental mES cell line through direct genome editing, and are hence fully isogenic.

## 2. RNA-mediated regulation of HP1 binding to chromatin

The HP1 hinge region acts as a flexible linker of the CD and the CSD, as a carrier of a NLS signal, and as a stabiliser of the interaction with the nucleosome. In addition, the positive charges of the basic hinge residues mediate RNA binding. The RNA interaction is best documented for the fission yeast HP1 homolog, Swi6 (Keller et al., 2012), followed by findings about the mammalian HP1 homolog, HP1α (Maison et al., 2011; Muchardt et al., 2002). Although RNA binding via the hinge seems to be evolutionary conserved, the role of this interaction could not differ more. RNA binding to the hinge of Swi6 triggers eviction from heterochromatin, while RNA interaction with the HP1α hinge promote targeting to heterochromatin. How can two opposite mechanisms be used by proteins with the same domain composition?

### 2.1. Sequence divergence of Swi6 and HP1α predicts different RNA-binding properties

The structured domains, CD and CSD, are well conserved between Swi6 and HP1α, but the NTE region and the hinge show only 10% sequence identity (Canzio et al., 2014). The Swi6 hinge is 83 amino acids long and contains 25 positively charged residues (KR25), while the HP1α hinge is 43 amino acids long with 13 positive residues (13KR). Despite the variation in length, the frequency of the KR residues is conserved between the two homologs. Examples of RNA binding facilitated by the KR-rich basic patches of unstructured domains have been previously documented, and proteome wide studies of mammalian RNA-binding proteins (RBP) revealed that such motifs are abundant amongst “unorthodox” RBPs (Castello et al., 2012; Järvelin et al., 2016; Kwon et al., 2013). RNA binding is typically mediated by the hydrogen bonds formed between RNA (phosphates, 2’-OH groups of ribose, and bases) and the -$NH_2$ groups of KR-residues and the -NH- groups of a side chain (Bahadur et al., 2008). Analogous disordered KR-patches were also found in DNA-binding proteins (DBPs), in which interaction with DNA largely depends on the phosphates and appears to be sequence-independent, serving rather as an affinity tuner of the main DNA-binding domain (Vuzman and Levy, 2012). It is not clear, however, how basic patches distinguish between RNA and DNA, or if and how they can recognise specific sequences.

The Swi6 hinge shows similar affinities for different RNA sequences but does not bind DNA (Keller et al., 2012). Similarly, HP1α was found to be a sequence-unspecific RNA binder with a very weak affinity for DNA (Muchardt et al., 2002). The preference of RNA over DNA can potentially originate from the higher stabilisation of the RNA:hinge complex due to the 2'-OH ribose groups, which are missing in DNA. Given the unspecificity of RNA binding by the hinge, it is possible that the affinity of the interaction linearly correlates with the number of basic patches. If so, we would expect the affinity of shorter linkers, like in HP1α, to be lower.

The NTE of Swi6 is about four times longer compared to the NTE of HP1α, and due to its length, it can fold back on the chromo domain, affecting its conformation. This property appears to be crucial in regulating the RNA-mediated eviction of Swi6 from chromatin. In the first step, RNA binding to the hinge triggers NTE folding to the chromo domain. In the second step, the NTE interaction with the CD outcompetes H3K9me3, resulting in the ejection of Swi6 from chromatin (personal communication, Claudia Keller). It is rather unlikely that HP1α with its short NTE can eject H3K9me3 in the same way. However, alternative mechanism may exist, including additional cofactors or PTMs.

## 2.2. Regulation of HP1α nucleic acid binding by PTMs

Immunofluorescence studies determined that the HP1α hinge region is required for proper localisation to heterochromatic foci. Mutation of the hinge KR residues results in dispersed HP1α binding across the nucleus (Muchardt et al., 2002), and so does the treatment of cells with RNase (Maison et al., 2002). These findings indicate RNA-dependent targeting of HP1α to heterochromatin rather than eviction. Indeed, SUMOylated HP1α was shown to interact specifically with major satellite repeats, guiding the recruitment of HP1α to pericentric chromatin (Maison et al., 2011). In this model, *de novo* recruitment of the SUMO-HP1α-RNA complex to pericentric regions is the initiating step of heterochromatin formation. In the following step, Suv39h is recruited and deposits H3K9 methylation, which creates a binding interface for unmodified HP1s, ultimately resulting in the spreading and compaction of heterochromatin. More recently, Suv39h1 was found to directly stimulate SUMOylation of HP1α, accelerating its targeting to *de novo* sites (Maison et al., 2016). This result indicates

that Suv39h1 recruitment is rather a parallel than a subsequent step in heterochromatin establishment, and further supports the paradigm of the interdependency of HP1s and HKMTases discussed in chapter 1.2.

The SUMO protein is half the size of HP1α, and it is quite conceivable that deposition of such a large modification will have a significant impact on the biophysical properties of the hinge region and surrounding domains. In addition, SUMO contains several basic patches that can potentially contribute to the specific RNA binding. Therefore, it is possible for HP1α to accommodate both RNA binding modes – unspecific binding by the unmodified hinge, and specific RNA binding by the SUMO-modified hinge.

The second PTM that is known to modulate the hinge is the phosphorylation of the NTE of HP1α introduced in chapter 1.3. (Larson et al., 2017). The unique behaviour of P-NTE-HP1α, oligomerisation and phase separation, can be abolished by mutation of the hinge KR residues. In addition, the chimeric HP1β protein that has P-NTE from HP1α cannot induce oligomerisation despite having a hinge region with 12 KR residues. However, the chimeric HP1β with P-NTE and hinge from HP1α, can form higher order oligomers. This is rather unexpected finding given the similarity of the hinge KR patches between the HP1 homologs, and suggests that certain KR residues might be essential and/or that other residues of the hinge are functionally equally important.

The P-NTE mediated oligomerisation followed by phase separation at higher concentrations was explained by the ability of phosphates on the NTE to make bridging contacts with the KR patch in the hinge region of a neighbouring dimer. Mechanistically, this is analogous to the nucleic acid binding described in chapter 2.1: the backbone phosphates of a nucleic acid create hydrogen bonds with the $-NH_2$ groups of KR residues. Indeed, oligomerisation and phase separation of unphosphorylated HP1α can be induced by DNA, and mutation of positive hinge residues or CSD dimerisation impairs this HP1α property.

Somewhat surprisingly, Larson *et al.* did not test if RNA can induce oligomerisation, or alternatively, if it can outcompete HP1 oligomerisation on DNA. In the latter case, higher local concentrations of RNA could trigger the release of HP1 dimers bound to RNA, inhibiting HP1 oligomerisation on chromatin and thus preventing phase separation. This could serve as a mechanism maintaining heterochromatin in a semi-permissive state as opposed to the

restricted state in phase-separated droplets, and/or as a mechanism preventing heterochromatin spreading into the neighbouring active regions, as is the case for Swi6.

### 2.3. HP1α binding to nucleosomes promotes RNA interaction and phase separation

During my PhD, I employed the biochemical methods to dissect the interaction of HP1α and other HP1 homologs with RNA. My work confirmed that HP1α interacts with RNA *in vitro*, but its affinity is extremely weak, approximately 40- to 80-fold weaker than that of Swi6. The weaker affinity of HP1α is not that surprising given the shorter length of its hinge accompanied by the smaller number of KR residues. Interestingly, RNA binding by HP1α can be stimulated by the GST tag, possibly due to a change in HP1α conformation, with the resulting affinity being about 5-fold stronger compared to untagged HP1α. As previously published, mutations of the hinge KR residues (KR13A) abolished the interaction with RNA (Muchardt et al., 2002). I also reproduced the unspecificity of RNA binding by the GST-HP1α protein, as I did not observe any preference between GC- or AU-rich substrates (Muchardt et al., 2002). In the next step, it would be interesting to test RNA affinity and specificity of the recombinant SUMO-HP1α. Is SUMO indeed promoting specific RNA binding?

In accordance with the previously published work, I did not observe RNA binding by the other two homologs, HP1β and HP1γ (Maison et al., 2011; Muchardt et al., 2002). The fact that HP1β and HP1γ cannot bind RNA is somewhat puzzling as their hinges contain similar number of KR residues as HP1α. HP1β and HP1γ where shown to associate with telomeric RNAs by RIP (Deng et al., 2009), but this does not prove a direct binding. A similar "uniqueness" of the HP1α hinge was also observed in the DNA oligomerisation experiments performed by Larson *et al.* (discussed above in 2.2.).

Next, I hypothesised that the nucleosome changes the conformation of HP1α in the same way as the GST tag, and I tested the RNA binding of untagged HP1α in a complex with the nucleosome. Indeed, HP1α bound to the nucleosome interacted with RNA with a similar affinity as to GST-HP1α, and this interaction was specifically abolished by the KR13A mutation. Interestingly, Larson *et al.* proposed in their work that unmodified HP1α (unHP1α)

alone in solution forms a closed dimer, in which the hinge is inaccessible. This would support my observations that unHP1α rarely interacts with RNA, unless I disrupt its conformation using a GST tag or binding to the nucleosome. Similarly, modification of the hinge region by SUMO most likely distorts the HP1α conformation, and thus, promotes hinge exposure and RNA binding.

What could be a possible physiological role of the HP1 protein binding to RNA on chromatin? Stabilisation of the binding or ejection? To test this, we designed an NMR experiment consisting of three steps: (1) recording the spectrum of HP1 alone, (2) titrating nucleosomes, and identifying chromo domain resonances that are affected upon binding, and (3) titrating RNA to the complex, detecting changes in the CD resonances responsible for the interaction with the nucleosome. This would allow me to distinguish between stable RNA interaction on the nucleosome and RNA-binding induced ejection. However, in the second step phase separation occurred, preventing me from finding the answer.

Larson *et al.* did not test if unHP1α can phase separate in the presence of nucleosomes, but they tested a DNA template, which indeed induced oligomerisation and phase separation (OPS). Why did I not see a similar effect in the EMSA assays? According to Larson *et al.* there is a critical protein concentration that is required for OPS to happen, and during the EMSA assays the concentrations are too low. A possible way to overcome the OPS effect and still be able to perform my desired experiment would be to immobilise nucleosomes on the surface, for example by the SPR method, and then measure the interaction with HP1 and RNA.

Interestingly, three independent studies reported just recently that satellite transcripts associated with pericentric heterochromatin stabilise the binding of Suv39h enzymes at these sites (Johnson et al., 2017; Shirai et al., 2017; Velazquez Camacho et al., 2017). These findings support the RNA-mediated stabilisation of HP1α binding at heterochromatin.

However, for now based on my data I can only conclude that unHP1α binding to nucleosomes promotes the RNA interaction and OPS. In order to explain the functional importance of these observations and deepen our understanding of the underling mechanisms e.g., the different effects of RNA and DNA, we and/or others, have to perform additional experiments, some of which are proposed in the next chapter.

## 2.4. Outlook

I propose to focus on investigating the roles of RNA binding by HP1α in mouse ES cells, and to use biochemistry rather as a complementary tool. First, we have to find an HP1α RNA binding mutant that does not affect affinity to H3K9me as is the case for the HP1α-KR13A mutant. For this, I would perform photo-cross-linking followed by mass-spectrometry for the recombinant GST-HP1α bound to RNA in order to identify the exact residues forming bonds with RNA (Kramer et al., 2014). I would use this information to design new RNA binding mutants, and test their affinity for RNA and H3K9me, at the end selecting the ones which impair only RNA binding. At this point, we could also test unHP1α and the newly generated mutants in EMSA assays with nucleosomes, and possibly with SPR as proposed in chapter 2.3.

In the second step, I propose to ectopically express HP1α and various RNA binding mutants in the triple HP1 KO cell line (tKO; Table 2). The use of the tKO has several benefits. For example, in this background the mutant HP1α proteins cannot be recruited to chromatin via an interaction with endogenous HP1 homologs. We can also use the growth defect of the tKO as an easy readout allowing us to test the rescue potential of different HP1α mutants. Ultimately, I would use genome-wide approaches to identify the binding and expression profiles of HP1α and HP1α RNA binding mutants. Next, I would try to identify cellular RNA binding substrates of HP1α by CLIP-seq. However, this method is known to work better for strong RNA binders. Thinking beyond RNA binding, it would also be interesting to include other HP1 mutants. What happens when we mutate the tandem array of serines in the NTE, CD binding to H3K9me, CSD dimerisation, or if we remove the CTE? I believe that these experiments combined with the experiments already performed or proposed in the chapter 1.3.1 would deepen our understanding of HP1s and their roles in the establishment and maintenance of heterochromatin.

## 3. DNA-sequence-dependent recruitment of HP1s

The chromo-shadow domain mediates homo- and hetero-dimerisation of HP1s, and interactions with the other proteins. The HP1α domain dissection study revealed that the vast majority of the protein interactions requires chromo-shadow dimerisation, and in some cases, like for the Ehmt1/Ehmt2 interaction, also the intact chromo domain (Nozawa et al., 2010; Sampath et al., 2007). The information presented in the previous chapters underlines the importance of CSD-mediated interactions in the recruitment of HP1 proteins to chromatin. HP1 binding to H3K9me-marked chromatin depends on interactions with the corresponding HKMTases, which restrict HP1 binding patterns. At first sight, the interaction with HKMTases may seem to provide specificity for HP1 binding, but HKMTases themselves lack specificity.

Where does the specificity of all chromatin-modifiers, -readers, -erasers, or -remodellers originate from? Initially, it has to come from the DNA and/or RNA sequence. This, of course, is not a new concept, even for the HP1s themselves. HP1 homologs can be targeted to specific sites in the genome via interaction with Trim28-KRAB-ZFs, an example of DNA-mediated specificity; or by interaction with major satellite transcripts in the case of SUMOylated HP1α, an example of RNA-mediated specificity. In both cases, HKMTases are recruited as well, methylating H3K9 residues, and thus, mediating the spreading of HP1s across the locus.

The existence of specificity factors in the heterochromatin recruitment pathways prompted me to investigate whether HP1 homologs could be recruited to euchromatin in a similar manner, especially as HP1γ was shown to bind to sites lacking H3K9me2/3 (Smallwood et al., 2012). Indeed, my PhD work revealed that HP1 proteins can be recruited to genomic loci in a DNA sequence-specific, H3K9 methylation-independent, manner via an interaction with the transcription factor Adnp. In the following chapters, I am going to discuss my proteomic analysis of HP1s in mouse ES cells, which uncovered several potential targeting factors, and in addition to Adnp, I will discuss another two transcription factors that may act as recruiters of HP1s.

### 3.1. Identification of potential HP1 targeting factors

To test my hypothesis of DNA-binding protein-mediated recruitment of HP1s to euchromatin, I had to first determine the interactomes of the individual HP1 homologs. Our analysis revealed a large number of DNA binders and transcription factors, some of which were specific to individual homologs. Interestingly, HP1β and HP1γ interactomes were more similar to each other than to that of HP1α. This was a bit unexpected given the previously reported localization patterns of the HP1 homologs. HP1α and HP1β were shown to predominantly bind to the pericentric heterochromatin with some staining in euchromatin (Lachner et al., 2001; Minc et al., 1999; Tachibana et al., 2005), while HP1γ was shown to be mostly dispersed in euchromatin (Minc et al., 2000; Tachibana et al., 2005). However, we have to take into account that immunofluorescence only provides a very broad overview of binding patterns, which do not necessarily reflect all functional states of HP1s, and in addition that there may be differences between pluripotent and differentiated cells. Indeed, Mattout *et al.* observed more dispersed binding of HP1β in mouse ES cells as compared to retinoic acid differentiated cells (Mattout et al., 2015). This, in combination with the interactome similarity, suggested that in mouse ES cells HP1β and HP1γ may both bind to euchromatic regions.

The second surprising finding was the low levels of Suv39h1 HKMTase I recovered in my TAP purifications. This is most likely due to the low solubility of the compacted heterochromatin and its associated factors, which require a more elaborate protein purification protocol than we used, e.g. including benzonase digestions. It is, therefore, likely that my proteomic experiments enriched for proteins associated with less compacted chromatin, which given my main interest in euchromatin-associated interactors was not an issue. Nevertheless, overall my proteomic experiments reproduced many previously identified HP1 interactors, and in addition, identified several new ones (see Introduction; (Nozawa et al., 2010; Rosnoblet et al., 2011)). Moreover, this is the first HP1-interactome dataset determined using the endogenously tagged proteins. Previous studies relied on protein overexpression or on potentially non-specific antibodies.

Ultimately, I proceeded to identifying potential HP1 targeting factors. I looked for proteins that bound to DNA, contained a PxVxL HP1-interaction motif, and had a high enrichments in the HP1γ or HP1β/HP1γ purifications. The proteins best satisfying these criteria were: Adnp, Adnp2, and Mga (Max gene-associated). Because out of these three factors Adnp was

the only one linked to HP1 localisation in chromatin (Mosch et al., 2011), I selected it as the best candidate, and moved on to tackle its role in HP1 targeting, which I summarise and discuss in the next chapter. In the later chapters, I introduce other two prospective candidates, Adnp2 and Mga, and discuss their potential roles in HP1 recruitment.

### 3.2. Adnp-mediated recruitment of HP1s to euchromatin

In the next steps, I focused on the characterisation of the Adnp-HP1 interaction, which eventually resulted in the discovery of a novel protein complex consisting of Adnp, HP1β and/or HP1γ, and the chromatin remodeller Chd4. Our consequent analysis of the genome-wide binding profiles of the individual components revealed over 15 000 ChAHP-bound sites that are devoid of H3K9me2/me3. Given my interest in histone mark-independent recruitment, this was a particularly exciting finding as we have two unspecific chromatin factors, a reader and remodeller, binding to chromatin in a complex with a DNA-specificity factor.

To test if Adnp is responsible for the recruitment of HP1, I performed HP1γ ChIP-seq in *Adnp* KO cells. My result confirmed my hypothesis, as I observed depletion of HP1γ binding at the ChAHP sites. The following question was whether it is indeed the DNA sequence recognised by Adnp that is targeting HP1γ and Chd4. For this, we computationally identified the Adnp binding motif, and I deleted it at one of the ChAHP target sites, at the promoter of the *Igfbp4* gene. As a result of the DNA motif deletion, all three subunits, Adnp, HP1γ and Chd4, failed to bind to the *Igfbp4* promoter, which was, as for the *Adnp*-KO, accompanied by the upregulation of *Igfbp4* mRNA levels. Note: Igfbp4, Insulin-like growth factor binding protein 4.

Interestingly, deletion of *Cbx1* and *Cbx3* was not sufficient to derepress the ChAHP targets, while deletion of all three HP1-encoding genes resulted in upregulation of the target genes. This suggests that HP1α, which is not a member of ChAHP under wild type conditions, can join the complex in the absence of the other two homologs. To ultimately verify this redundancy of HP1s within the complex, I propose to perform ChIP experiments, comparing HP1α binding to ChAHP sites in wild type and *Cbx1Cbx3* KO cells.

In summary, these results establish a novel DNA sequence-specific, H3K9 methylation-independent pathway for HP1 recruitment to chromatin, and demonstrate that H3K9 methylation is not required for repression of ChAHP target genes.

### 3.2.1. Silencing mechanism of the ChAHP complex

The canonical HP1-mediated silencing mechanism involves HKMTase depositing H3K9 methylation, which mediates the spreading of HP1 homologs along the locus, resulting in the formation of a broad heterochromatic domain that is sterically inaccessible to the transcriptional machinery. How can ChAHP mediate silencing independently of H3K9 methylation? Based on our ATAC-seq experiments, ChAHP binding to chromatin renders these sites inaccessible. This implies that ChAHP can locally restrict access to its sites, preventing other regulators, including transcriptional activators, from binding to DNA. Furthermore, my single step Adnp purifications, which are less stringent, identified additional Adnp interactors: H3K4me-demethylase Lsd1 and histone deacetylases Hdac1 and Hdac2. Recruitment of these enzymes to the ChAHP sites might contribute to silencing by keeping nucleosomes in a hypoacetylated state, and protected from the H3K4me-activatition mark. Bioinformatic analysis of Lsd1 and Hdac1/2 genome-wide binding profiles should reveal if these enzymes indeed associate with ChAHP sites. We can use already published datasets from mouse ES cells by Whyte and his colleagues (Whyte et al., 2012).

However, we should not fully exclude a potential role of H3K9me methylation later in development. The ChAHP sites might be in a temporary silent state, allowing for rapid induction when activating signals are strong and outcompete ChAHP, or alternatively, for rapid silencing in tissues where the corresponding genes have to be silenced permanently. In the latter case, HKMTase could be recruited, and together with HP1s create typical heterochromatin domains. I could test this hypothesis by differentiating my cells to a neuronal lineage, and performing H3K9me-ChIP on ChAHP-bound endodermal genes. If I do observe deposition of H3K9me, it would be interesting to induce HP1 depletion at the beginning of the differentiation, and test if the heterochromatisation of the corresponding loci fails in the absence of HP1s.

Finally, in addition to characterising ChAHP, I propose to investigate potential existence of alternative targeting complexes, which is discussed in the next chapters.

### 3.3. Adnp2 as a prospective targeting factor of HP1s

As the name suggests Adnp2 is a homolog of Adnp, and thus, it is my next top candidate for TF-mediated recruitment of HP1s. In the following paragraphs, I am going to summarise what is known about Adnp2, and later, I am going to discuss its possible roles and its redundancy with Adnp, ultimately speculating about the existence of an alternative ChAHP complex, ChAHP2. In parallel, I am going to propose experiments that can help us to verify those ideas.

The Adnp2 homolog shows 20% sequence identity compared to Adnp, which may seem low at first sight, yet the majority of major features are conserved between the two proteins (Fig. 21). Adnp2 contains a conserved N-terminus, nine N-terminal zinc fingers, a C-terminal homeodomain, and a C-terminal PxVxL pentapeptide motif; however, it lacks an H3K9 mimic found in Adnp, and instead contains an additional C-terminal extension of unknown function. Similarly to Adnp, all of the Adnp2 ZFs are the C2H2 type but they show different levels of conservation, which might affect their sequence specificity compared to Adnp; three ZFs are conserved (1, 8, 9), and six degenerate (2, 3, 5, 6, 7) (Bateman et al., 2017). Despite variation in the conservation of the individual ZFs between the two homologs, they appear to be distributed in a similar manner, a cluster of four at N-terminus and a cluster of five closer to the homeodomain.

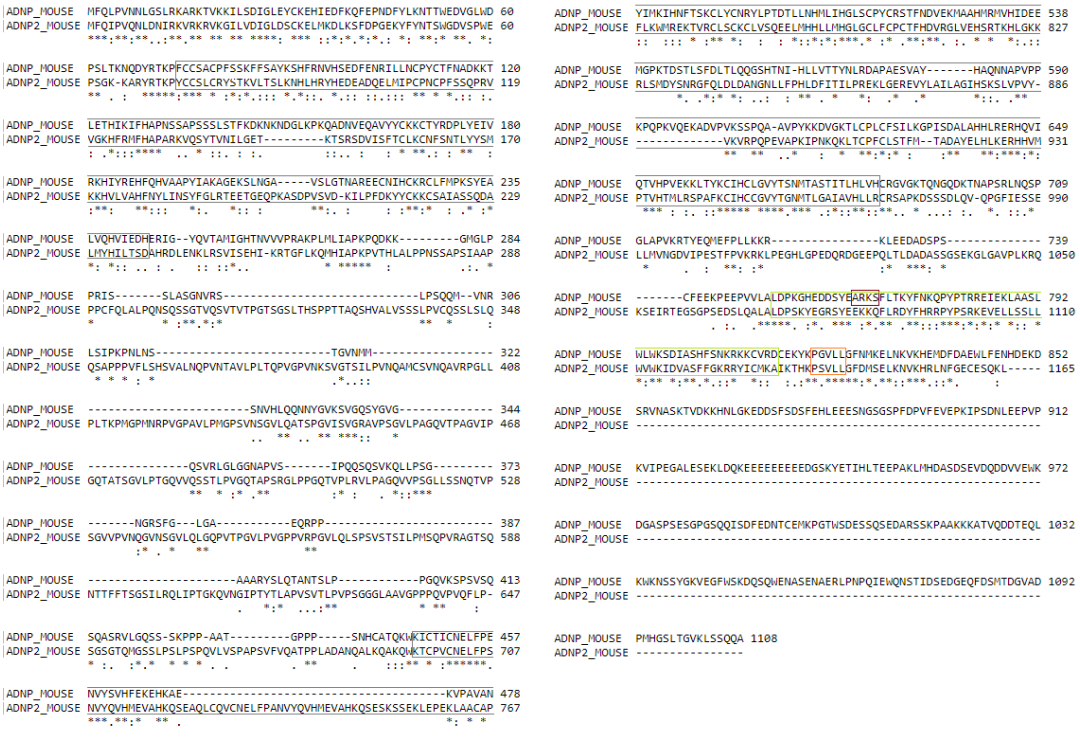


**Figure 21. Uniprot clustal aligment of Adnp and Adnp2 homologs.**

The two homologs show 20% sequence identity, and share N-terminal cluster of four zinc fingers (grey), followed by the second cluster of five zinc fingers (grey), a C-terminal homeodomain (green), and a PxVxL motif (orange). Adnp2 homolog lacks an H3K9 mimic (red), but contains C-terminal extension missing in Adnp.

Literature studying Adnp2 and its roles is extremely sparse. One of the few available studies tested *Adnp2* expression levels in mouse tissues, finding it to be ubiquitously expressed in distinct tissues with the highest expression levels in brain, and cerebral cortex in particular (Kushnir et al., 2008). The upregulation of *Adnp2* in neuronal tissues appears to already be induced during embryonic development. Similar expression patterns were observed for *Adnp* mRNAs (see Introduction). In addition, a *de novo* mutation in the *ADNP2* gene was identified in whole-genome sequencing in one of the patients with the behavioural disorder (Chung et al., 2015), and *ADNP2* mRNA levels appear to be upregulated in patients with schizophrenia (Dresner et al., 2011). Altogether, this suggest a possible role of Adnp2 in neurogenesis as is the case for Adnp, but given the limited information it is impossible to make any conclusions about Adnp2 function.

### 3.3.1. Potential redundancy of Adnp homologs and the existence of ChAHP2

The homeodomains of both homologs appear to be highly conserved, which suggests similar DNA motif recognition, but our data show that Adnp does not require a functional homeodomain for binding to ChAHP sites. Adnp binding to ChAHP sites appears to be dependent on the ZFs, which are not as highly conserved between the two homologs. ZFs can be in theory provide specificity driving the two homologs to distinct/partially overlapping sites, while the homeodomain might be important rather for stabilisation of DNA binding. I can compare the genome-wide binding of the two homologs, if I endogenously tag the *Adnp2* gene and perform ChIP-sequencing. Will I see separate, partly overlapping or overlapping binding? This would already indicate to what extent the two homologs are redundant in mouse ES cells.

My data show that Adnp is not essential for the survival of mESCs but its depletion leads to upregulation of a subset of lineage specifying genes, and to changes in morphology of the cells accompanied by a reduction in alkaline phosphatase staining. This is consistent with the effects seen in *Adnp* KO mice, where embryos fail to express the correct lineage markers

(Pinhasov et al., 2003). However, it is surprising that the *Adnp* KO embryos die only at E8.5-E9.5 if the phenotype is, at least to some extent, present already in the pluripotent state. It is possible that Adnp and Adnp2 are partly redundant in the early stages of development, and their roles become essential in later stages. Thus, I propose to delete *Adnp2* in wild type and *Adnp* KO mESCs. This, in combination with the proposed ChIP-seq experiment, will provide a good basis for investigation of their redundancy in mESCs.

Moreover, I can already draw some conclusions about the two homologs based on my proteomics data, which would, in my opinion, rather suggest partial redundancy. As mentioned above, HP1β and HP1γ both co-purify Adnp and Adnp2, which is not that surprising given that both homologs contain a PxVxL motif. However, HP1β seems to favour Adnp2, while HP1γ shows higher enrichments of Adnp. In addition, HP1γ binds preferentially to the genomic ChAHP sites compared to HP1β, and the recruitment of HP1β can be stimulated by deleting *Cbx3*. What sites does HP1β favour in the wild type condition?

Perhaps in wild type cells, HP1β might be preferentially targeted to Adnp2 genomic sites? And if so, can Chd4 join Adnp2 and HP1β, forming a ChAHP2 complex? Again, I can look for some clues in my data. First, I know that a single step purification of heterozygously tagged Chd4 recovers Adnp, Adnp2, HP1β and HP1γ (Fig. 22). Second, from our *in vitro* complex reconstitution we know that HP1s do not directly interact with Chd4, and that the N-terminus of Adnp is required for the interaction with Chd4. My preliminary ChAHP crosslinking-MS/MS experiments point towards the N-terminal 60aa of Adnp as a site of the Chd4 interaction (data not shown). Interestingly, looking at the alignment of Adnp with Adnp2 we can see that this N-terminal part is one of the best conserved regions in the two proteins (Fig. 21). Third, from my Adnp TAP experiments I know that Adnp does not interact with Adnp2, meaning they cannot co-purify each other. These three points in combination with the very high Adnp2 enrichments in Chd4 TAPs indicate that Chd4 could indeed directly interact with Adnp2. But in order to get an ultimate answer we have to perform high stringency TAP purifications with the endogenously tagged Adnp2 protein, followed by *in vitro* reconstitution. For a start, we can use my ChIP-seq data to search for the Chd4-HP1 co-bound sites that are independent of Adnp.

In summary, I speculate that there are two partially redundant ChAHP complexes in mouse ES cells. ChAHP consisting of Chd4, Adnp and HP1γ, and ChAHP2 consisting of Chd4,

Adnp2 and HP1β. HP1β and HP1γ are promiscuous components, which is a reason why I recover both Adnp and Adnp2 in their corresponding TAPs. The close similarity of both complexes suggests partial redundancy, and would explain why in *Adnp* KO cells low levels of HP1γ persist at ChAHP sites. Those could be redundant targets of the ChAHP2 complex, to which HP1γ can be recruited as well. This would also explain why I see rather a small subset of ChAHP bound genes being upregulated in *Adnp* KO. Of course, I can test all of those assumptions performing the experiments I proposed in the previous paragraphs.

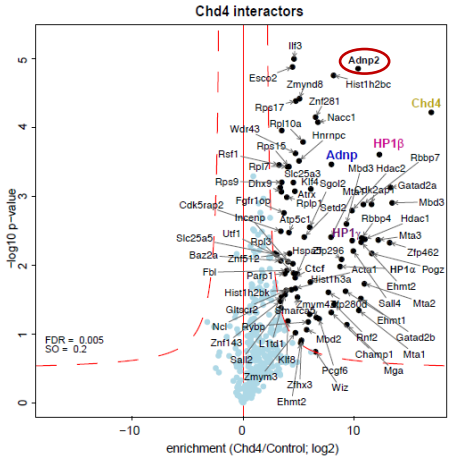



**Figure 22. Chd4 interactome in mouse ES cells.**

AP-LC-MS/MS of endogenously FLAG/Avi-tagged Chd4. Protein purification was performed in the presence of 350mM NaCl. Parental mES cell line serves as background control. n=3 biological replicates (i.e. 3 independent *Chd4*$^{FlagAvi/-}$ mES cell lines).

## 3.4. Mga as a prospective recruiter of HP1s

Another transcription factor highly enriched in my TAP-LC-MS/MS experiments that could act as a specific HP1 recruiter is Mga. Mga is a 330kDa protein with a basic helix-loop-helix DNA binding domain and a PxVxL motif, PQVFL at position 1593aa-1597aa (Bateman et al., 2017). Mga is known to form a heterodimer with another transcription factor recovered in my purifications, although at much lower levels, Max (Myc-associated factor X) (Hurlin et al., 1999). Mga/Max heterodimer (MMh) is a known component of the previously identified non-canonical PRC1 complex, PRC1.6. In addition to MMh, PRC1.6 consists of Ehmt1/2,

HP1γ, HP1β, E2f6 (E2 factor 6), Pcgf6 (Polycomb group ring finger 6), Ring2 (Ring finger protein 2), L3mbtl2 (Lethal 3 malignant brain tumour-like protein 2), and Tfdp1 (Transcription factor dimerisation partner 1), all of which I co-purified; plus several other variable components, which do not appear in my purifications (Gao et al., 2012; Ogawa et al., 2002; Qin et al., 2012; Trojer et al., 2011).

MMh could act as a sequence specifier of HP1 binding in the context of the PRC1.6 complex. The presence of HKMTases in the complex also indicates the presence of H3K9me2 at those sites. I hypothesise that this is another example of HKMTase-HP1 binding specified by DNA sequence, as in the case of Setdb1-Trim28-KRAB-ZFs.

Indeed just recently, MMh was shown to be responsible for the sequence specific recruitment of several PRC1.6 components, Pcgf6, Ring2 and L3mbtl2, to the TSS of meiosis- and germ cell-related genes in mouse ES cells; and depletion of Pcgf6, Ring2 or Max resulted in derepression of the bound genes (Endoh et al., 2017). Endoh *et al.* did not test if HP1 and/or Ehmt1/2 binding is lost upon MMh depletion. However, they looked at the expression profiles of *Cbx1Cbx3* double KO mouse ES cells, and did not observe any upregulation of meiosis- or germ cell-related genes, concluding that HP1s are not required for silencing.

These findings are quite intriguing considering the *Cbx3* KO mice phenotype in the germline. It would be interesting to test if Ehmt1/2 and HP1β/γ binding to PRC1.6 sites is dependent on MMh, and if I see upregulation of meiosis- and germ cell-related genes in my tKO cell line. The redundancy of HP1s I have observed so far in ES cells (ERV silencing and within ChAHP) could also apply in the context of the PRC1.6 complex, especially as HP1α also interacts with Mga in my TAP experiments. A homolog-specific function of HP1γ may be required only at the later stages of differentiation of germ cells during mouse development; this brings me back to the experiments I proposed in the first chapter of this discussion (1.3.1).

# 4. Concluding remarks

The generally accepted canonical pathway of HP1 recruitment via H3K9me neither explains different localisation patterns of the HP1 homologs, nor how they can distinguish between H3K9me2 and H3K9me3, nor how they are recruited to sites lacking H3K9me. In this final chapter, I present my personal perspective on HP1 recruitment, based on my work and the literature discussed in this thesis. I propose that the role of the H3K9 methylation mark is not to recruit HP1 proteins, but to increase affinity of binding after recruitment, and to mediate HP1 spreading along the locus. HP1s can be recruited via interactions with RNA and/or other proteins: (a) major satellite transcripts providing specificity to pericentric chromatin, (b) already bound HP1s and HKMTases, and (c) DNA-binders and transcription factors targeting HP1s based on their DNA-sequence specificity. In case of the last option, the interaction is not necessarily direct, which is further discussed in chapter 4.2.

## 4.1. RNA-regulated binding of HP1α to pericentric chromatin

I postulate that there are two modes of HP1α-RNA interaction, both affecting binding to chromatin (Fig. 23). The first is RNA-mediated recruitment of HP1α to pericentric chromatin that depends on SUMOylation of the hinge, which facilitates specific RNA-binding, and ultimately results in *de novo* recruitment of SUMO-HP1α and Suv39h1 to pericentric chromatin (Fig. 23A) (Maison et al., 2011, 2016). This is followed by recruitment of additional unmodified HP1s and Suv39h1 enzymes, and results in spreading and, ultimately, establishment of pericentric heterochromatin (Fig. 23B). Suv39h1 interacts with major satellite transcripts that associate with pericentric heterochromatin, and this stabilises its retention at the locus (Johnson et al., 2017; Shirai et al., 2017; Velazquez Camacho et al., 2017). Similarly, HP1α bound to nucleosomes can interact with major satellite transcripts, but the effect of such an interaction is unclear (my work, Fig. 23B). There are two options: (1) stabilisation of binding to nucleosomes, and/or (2) eviction of HP1α from chromatin (as in case of Swi6 (Keller et al., 2012)). I speculate that low levels of RNA stabilise HP1α binding, but high levels (e.g. at the borders of heterochromatin) could induce eviction. A similar effect of RNA concentration was, in fact, observed for Swi6 itself (Keller et al., 2012).

Interestingly, this pathway appears to be unique to the HP1α homolog, which is neither essential during mouse development nor in the later adult stages. This indicates that there are alternative mechanisms of pericentric heterochromatin establishment and maintenance, which require further investigation. Another open question that remains to be investigated is whether similar RNA-mediated regulation of HP1 binding to pericentric heterochromatin also exists in different chromatin contexts, for example in ERV repeats that are found spread across the euchromatic regions.

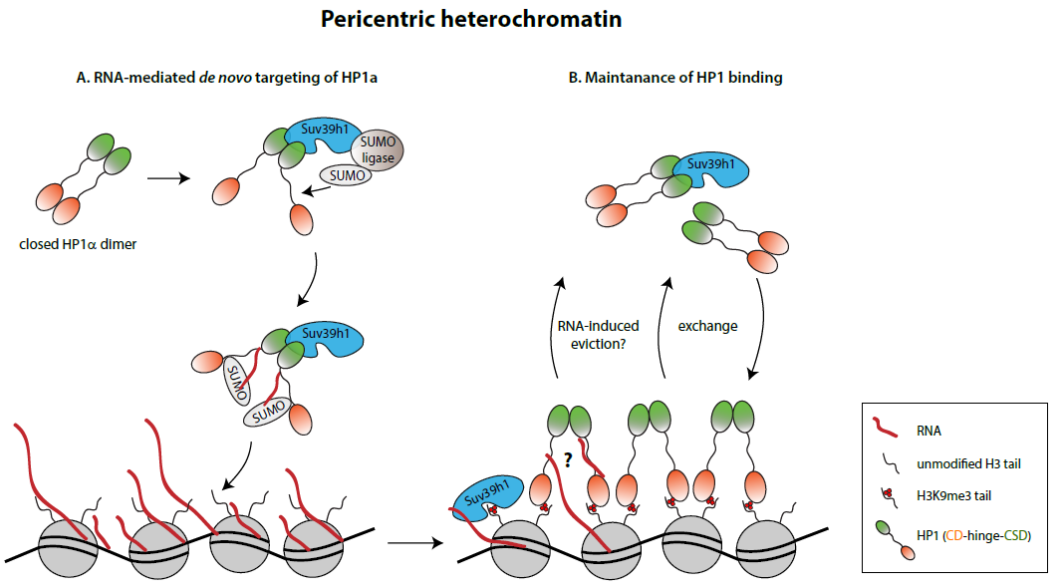



**Figure 23. Simplified model of HP1α targeting to pericentric heterochromatin.**

(A) A scheme depicting RNA-mediated *de novo* targeting of HP1α to pericentric chromatin. Top left, the nucleoplasmic HP1α dimer is in a closed conformation that is incompatible with RNA binding. Suv39h1-stimulated SUMOylation of the HP1α hinge induces a change in conformation, which exposes the hinge and facilitates RNA binding (open HP1α dimer). Specific binding of the major satellite (pericentric) transcripts targets the SUMO-HP1α-Suv39h1 complex to the pericentric chromatin. Suv39h1 deposits H3K9 trimethylation to which HP1α binds, and this initiates heterochromatin formation and spreading (B).

(B) A scheme depicting HP1α binding to pericentric heterochromatin. Suv39h1-deposited H3K9me3 increases the affinity of HP1α binding and mediates HP1α spreading along the locus. Nucleoplasmic HP1α and Suv39h1 can bind directly to chromatin via interactions with already bound HP1α and/or Suv39h1. Major satellite transcripts stabilise Suv39h1 association with pericentric heterochromatin. HP1α bound to nucleosomes is in an open conformation, and can interact with major satellite transcripts, but the effect of this interaction is not clear. It can, as in the case of Suv39h1, stabilise retention on chromatin or it can lead to eviction of HP1α to the nucleoplasm.

This model is based on (Johnson et al., 2017; Keller et al., 2012; Larson et al., 2017; Maison et al., 2011, 2016; Shirai et al., 2017; Velazquez Camacho et al., 2017), and my work presented in this thesis. Scales are approximate.

## 4.2. DNA-sequence-specific targeting of HP1s to silent euchromatin

HP1 proteins can be recruited to distinct loci in euchromatin where they contribute to silencing. For the sites associated with H3K9 methylation, it is, indeed, tempting to conclude that the histone mark acts as a recruiter. However, if we take a closer look at figure 24, we see that in almost all cases HP1s closely interact with DNA specificity factors. What would be an evolutionary purpose of these interactions if not recruitment?

For the ERV class I-II, HP1s interact directly with Trim28, which gains its specificity from KRAB-ZFs (Fig. 24A). For the ChAHP sites, HP1s interact directly with the transcription factor Adnp, which is essential for HP1 recruitment (Fig. 24C). The importance of HP1 targeting to the Adnp sites is underlined by my work, which revealed that Adnp with mutations found in Helsmoortel-Van der Aa syndrome patients fails to interact with HP1s, and thus, target them to chromatin. For the PRC1.6 sites, HP1s interact with the recruiting transcription factor Mga, which contains a PxVxL motif, indicating it is a direct interaction (Fig. 24D). Finally for the ERV class III, there is no known recruiter of the Ehmt1/2 heterodimer, and therefore, it would be unfair to make any conclusions about HP1 recruitment or the order of binding (Fig. 24B). However, my proteomic datasets uncovered multiple zinc finger proteins, some of which were also shown to interact with Ehmt2 (Maier et al., 2015; Zhang et al., 2016). ZFs that have especially high enrichments in my experiments are Zfp462, Zfp518b, and Zfp280c/d. Out of these, Zfp518b appears to be the most interesting candidate for targeting Ehmt2 to the ERVs class III: (1) it is one of the top enriched proteins in all HP1 TAPs, and all three homologs bind to ERVs, and (2) Zfp518b KD was shown to reduce the levels of H3K9me2 (Maier et al., 2015). It would be interesting to test if Zfp518b or other ZFs recruit HP1s and Ehmt1/2 to the specific sites in the genome.

Finally, coming to the role of H3K9me, my PhD work demonstrated that HP1 homologs do not require H3K9me for silencing of the ChAHP sites. The only effect we see as a consequence of lacking HKMTase and H3K9me at those loci, is that HP1s cannot spread (Fig. 24C). This is apparent from my ChIP-seq data where we see narrow HP1 peaks at the ChAHP sites, and wide HP1 peaks at the H3K9me3-marked sites. As discussed in chapter 3.2 of this section, the difference between the H3K9me-HP1- and ChAHP-silenced chromatin could be the stability of the repressed states - permanent and temporary silencing, respectively.

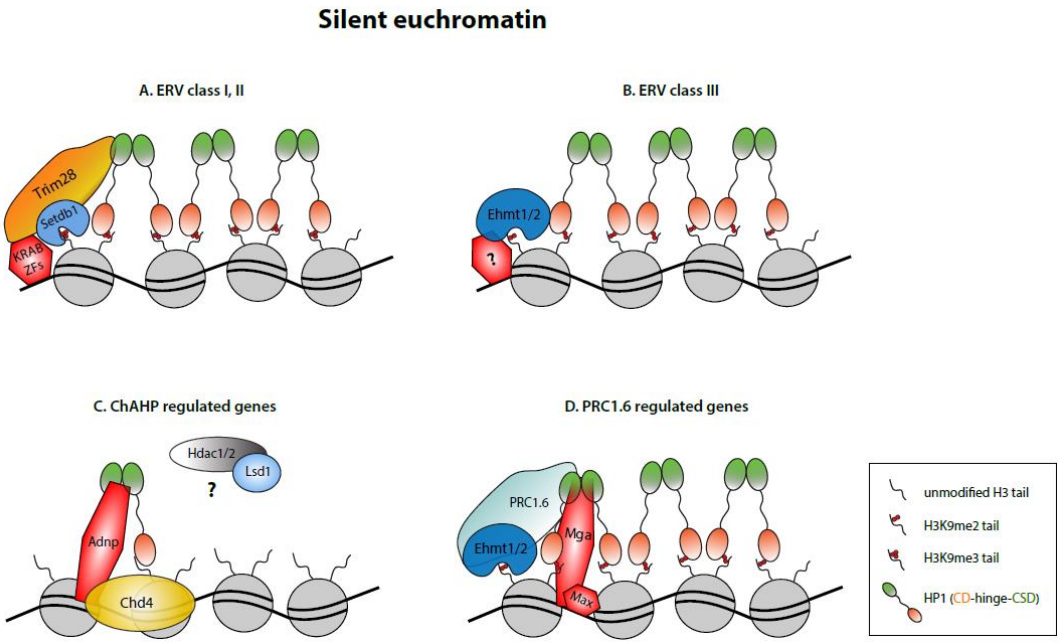



**Figure 24. Simplified model of HP1 targeting to silent euchromatin.**

(A) HP1 homologs are recruited to the ERV class I and II via direct interaction with Trim28-KRAB-ZFs. Trim28 recruits Setdb1, which deposits H3K9me3, and mediates HP1 spreading.

(B) HP1 homologs can be recruited to the ERV class III, which is silenced by an Ehmt1/2 heterodimer that deposits H3K9 dimethylation. The mechanism of the Ehmt1/2 and HP1 recruitment is not clear, but HP1s interact with Ehmt1/2 via the chromo-shadow domain and/or chromo domain.

(C) HP1 homologs are recruited by the transcription factor Adnp in a DNA-sequence-specific, H3K9me-independent, manner. HP1, Adnp, and the chromatin remodeller Chd4 form a complex named ChAHP. HP1γ associates preferentially with the complex. The complex may also associate with additional repressive enzymes such as Lsd1 and/or Hdac1/2. The lack of H3K9me at these sites prevents HP1 spreading, but does not affect silencing.

(D) HP1s and Ehmt1/2 are components of the PRC1.6 complex, which is recruited to meiosis- and germ cell-related genes. PRC1.6 appears to be targeted by two transcription factors, Mga and Max, which form a heterodimer. The Mga/Max-dependency of HP1 or Ehmt1/2 binding to the PRC1.6 sites has not been tested. Mga contains a PxVxL-HP1 interaction motif.

(A), (B), (D) The presence of HKMTases and, thus, H3K9 methylation mediates spreading of HP1 along the locus, while in (C) HP1 binding is locally restricted.

This model is based on (Endoh et al., 2017; Maksakova et al., 2011, 2013; Ogawa et al., 2002), and my work presented in this thesis. Scales are approximate.

# *References*

Aagaard, L., Laible, G., Selenko, P., Schmid, M., Dorn, R., Schotta, G., Kuhfittig, S., Wolf, A., Lebersorger, A., Singh, P.B., et al. (1999). Functional mammalian homologues of the Drosophila PEV-modifier Su(var)3-9 encode centromere-associated proteins which complex with the heterochromatin component M31. EMBO J. *18*, 1923–1938.

Abe, K., Naruse, C., Kato, T., Nishiuchi, T., Saitou, M., and Asano, M. (2011). Loss of heterochromatin protein 1 gamma reduces the number of primordial germ cells via impaired cell cycle progression in mice. Biol Reprod *85*, 1013–1024.

Adams, C.C., and Workman, J.L. (1995). Binding of disparate transcriptional activators to nucleosomal DNA is inherently cooperative. Mol. Cell. Biol. *15*, 1405–1421.

Akhtar, A., and Gasser, S.M. (2007). The nuclear envelope and transcriptional control. Nat. Rev. Genet. *8*, 507–517.

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

Allis, C.D., Caparros, M.-L., Jenuwein, T., Reinberg, D., and Lachner, M. (2015). Epigenetics (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press).

Amor, D.J., Kalitsis, P., Sumer, H., and Andy Choo, K.H. (2004). Building the centromere: from foundation proteins to 3D organization. Trends Cell Biol. *14*, 359–368.

Andersen, P.R., Tirian, L., Vunjak, M., and Brennecke, J. (2017). A heterochromatin-dependent transcription machinery drives piRNA expression. Nature *549*, 54–59.

Aravin, A.A., Naumova, N.M., Tulin, A. V., Vagin, V. V., Rozovsky, Y.M., and Gvozdev, V.A. (2001). Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline. Curr. Biol. *11*, 1017–1027.

Aucott, R., Bullwinkel, J., Yu, Y., Shi, W., Billur, M., Brown, J.P., Menzel, U., Kioussis, D., Wang, G., Reisert, I., et al. (2008). HP1-beta is required for development of the cerebral neocortex and neuromuscular junctions. J. Cell Biol. *183*, 597–606.

Aydin, E., Kloos, D.P., Gay, E., Jonker, W., Hu, L., Bullwinkel, J., Brown, J.P., Manukyan, M., Giera, M., Singh, P.B., et al. (2015). A hypomorphic Cbx3 allele causes prenatal growth restriction and perinatal energy homeostasis defects. J. Biosci. *40*, 325–338.

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

Bahadur, R.P., Zacharias, M., and Janin, J. (2008). Dissecting protein-RNA recognition sites. Nucleic Acids Res. *36*, 2705–2716.

Ball, L.J., Murzina, N. V, Broadhurst, R.W., Raine, A.R.C., Archer, S.J., Stott, F.J., Murzin, A.G., Singh, P.B., Domaille, P.J., and Laue, E.D. (1997). Structure of the chromatin binding (chromo) domain from mouse modifier protein 1. EMBO J. *16*, 2473–2481.

Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. Cell Res. *21*, 381–395.

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

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

Bassan, M., Zamostiano, R., Davidson, A., Pinhasov, A., Giladi, E., Perl, O., Bassan, H., Blat, C.,

Gibney, G., Glazner, G., et al. (1999). Complete sequence of a novel protein containing a femtomolar-activity- dependent neuroprotective peptide. J. Neurochem. *72*, 1283–1293.

Basta, J., and Rauchman, M. (2017). The Nucleosome Remodeling and Deacetylase Complex in Development and Disease. In Translating Epigenetics to the Clinic, (NIH Public Access), pp. 37–72.

Bateman, A., Martin, M.J., O'Donovan, C., Magrane, M., Alpi, E., Antunes, R., Bely, B., Bingley, M., Bonilla, C., Britto, R., et al. (2017). UniProt: The universal protein knowledgebase. Nucleic Acids Res. *45*, 158–169.

Berger, S.L., Kouzarides, T., Shiekhattar, R., and Shilatifard, A. (2009). An operational definition of epigenetics. Genes Dev. *23*, 781–783.

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

Bibel, M., Richter, J., Schrenk, K., Tucker, K.L., Staiger, V., Korte, M., Goetz, M., and Barde, Y.-A. (2004). Differentiation of mouse embryonic stem cells into a defined neuronal lineage. Nat. Neurosci. *7*, 1003–1009.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes Dev. *16*, 6–21.

Bird, A. (2007). Perceptions of epigenetics. Nature *447*, 396–398.

Bird, A.P. (1978). Use of restriction enzymes to study eukaryotic DNA methylation. II. The symmetry of methylated sites supports semi-conservative copying of the methylation pattern. J. Mol. Biol. *118*, 49–60.

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

Blattner, F., Plunket III, G., Bloch, C., Perna, N.N., Burland, V., Riley, M., Collado-Vides, J., Glasner, J., Rode, C., Mayhew, G., et al. (1997). The Complete Genome Sequence of Escherichia coli K-12. Science (80-. ). *277*, 1453–1462.

Bönisch, C., and Hake, S.B. (2012). Histone H2A variants in nucleosomes and chromatin: More or less stable? Nucleic Acids Res. *40*, 10719–10741.

Bouazoune, K., Mitterweger, A., Längst, G., Imhof, A., Akhtar, A., Becker, P.B., and Brehm, A. (2002). The dMi-2 chromodomains are DNA binding modules important for ATP-dependent nucleosome mobilization. EMBO J. *21*, 2430–2440.

Boyer, L.A., Latek, R.R., and Peterson, C.L. (2004). The SANT domain: a unique histone-tail-binding module? Nat. Rev. Mol. Cell Biol. *5*, 158–163.

Brasher, S. V, Smith, B.O., Fogh, R.H., Nietlispach, D., Thiru, A., Nielsen, P.R., Broadhurst, R.W., Ball, L.J., Murzina, N. V, and Laue, E.D. (2000). The structure of mouse HP1 suggests a unique mode of single peptide recognition by the shadow chromo domain dimer. EMBO J. *19*, 1587–1597.

Brown, J.P., Bullwinkel, J., Baron-Lühr, B., Billur, M., Schneider, P., Winking, H., and Singh, P.B. (2010). HP1gamma function is required for male germ cell survival and spermatogenesis. Epigenetics Chromatin *3*, 1-9.

Brownell, J.E., Zhou, J., Ranalli, T., Kobayashi, R., Edmondson, D.G., Roth, S.Y., and Allis, C.D. (1996). Tetrahymena histone acetyltransferase A: A homolog to yeast Gcn5p linking histone acetylation to gene activation. Cell *84*, 843–851.

Bühler, M., Verdel, A., and Moazed, D. (2006). Tethering RITS to a Nascent Transcript Initiates RNAi- and Heterochromatin-Dependent Gene Silencing. Cell *125*, 873–886.

Canzio, D., Liao, M., Naber, N., Pate, E., Larson, A., Wu, S., Marina, D.B., Garcia, J.F., Madhani, H.D., Cooke, R., et al. (2013). A conformational switch in HP1 releases auto-inhibition to drive heterochromatin assembly. Nature *496*, 377–381.

Canzio, D., Larson, A., and Narlikar, G.J. (2014). Mechanisms of functional promiscuity by HP1 proteins. Trends Cell Biol. *24*, 377–386.

Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E., Humphreys, D.T., Preiss, T., Steinmetz, L.M., et al. (2012). Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins. Cell *149*, 1393–1406.

Cavalli, G., and Misteli, T. (2013). Functional implications of genome topology. Nat. Struct. Mol. Biol. *20*, 290–299.

Chen, D., Ma, H., Koh, S.S., Huang, S.-M., Schurter, B.T., Aswas, D.W., and Stallcup, M.R. (1999). Regulation of Transcription by a Protein Methyltransferase. Science *284*, 2174–2177.

Cheutin, T., McNairn, A.J., Jenuwein, T., Gilbert, D.M., Singh, P.B., and Misteli, T. (2003). Maintenance of Stable Heterochromatin Domains by Dynamic HP1 Binding. Science *299*, 721–725.

Cheutin, T., Gorski, S.A., May, K.M., Singh, P.B., and Misteli, T. (2004). In vivo dynamics of Swi6 in yeast: evidence for a stochastic model of heterochromatin. Mol. Cell. Biol. *24*, 3157–3167.

Chinwalla, A.T., Cook, L.L., Delehaunty, K.D., Fewell, G.A., Fulton, L.A., Fulton, R.S., Graves, T.A., Hillier, L.W., Mardis, E.R., McPherson, J.D., et al. (2002). Initial sequencing and comparative analysis of the mouse genome. Nature *420*, 520–562.

Chuang, L.S.-H., Ian, H.-I., Koh, T.-W., Ng, H.-H., Xu, G., and Li, B.F.L. (1997). Human DNA-(Cytosine-5) Methyltransferase-PCNA Complex as a Target for p21WAF1. Science *277*, 1996–2000.

Chung, J.H., Cai, J., Suskin, B.G., Zhang, Z., Coleman, K., and Morrow, B.E. (2015). Whole-Genome Sequencing and Integrative Genomic Analysis Approach on Two 22q11.2 Deletion Syndrome Family Trios for Genotype to Phenotype Correlations. Hum. Mutat. *36*, 797–807.

Cirillo, L.A., McPherson, C.E., Bossard, P., Stevens, K., Cherian, S., Shim, E.Y., Clark, K.L., Burley, S.K., and Zaret, K.S. (1998). Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome. EMBO J. *17*, 244–254.

Clapier, C.R., and Cairns, B.R. (2014). Fundamentals of Chromatin: Chromatin Remodeling Complexes. Springer. 69–146.

Clapier, C.R., Iwasa, J., Cairns, B.R., and Peterson, C.L. (2017). Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. Nat. Rev. Mol. Cell Biol. *18*, 407–422.

Clayton, A.L., Rose, S., Barratt, M.J., and Mahadevan, L.C. (2000). Phosphoacetylation of histone H3 on c-fos- and c-jun-associated nucleosomes upon gene activation. EMBO J. *19*, 3714–3726.

Cosma, M.P., Tanaka, T., and Nasmyth, K. (1999). Ordered recruitment of transcription and chromatin remodeling factors to a cell cycle- and developmentally regulated promoter. Cell *97*, 299–311.

Cowieson, N.P., Partridge, J.F., Allshire, R.C., and McLaughlin, P.J. (2000). Dimerisation of a chromo shadow domain and distinctions from the chromodomain as revealed by structural analysis. Curr. Biol. *10*, 517–525.

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

Dang, W., and Bartholomew, B. (2007). Domain Architecture of the Catalytic Subunit in the ISW2-Nucleosome Complex. Mol. Cell. Biol. *27*, 8306–8317.

Davis, R.L., Weintraub, H., and Lassar, A.B. (1987). Expression of a Single Transfected cDNA Converts Fibmblasts to Myoblasts. Cell *51*, 987–1000.

Delmas, V., Stokes, D.G., and Perry, R.P. (1993). A mammalian DNA-binding protein that contains a chromodomain and an SNF2/SWI2-like helicase domain. Proc. Natl. Acad. Sci. *90*, 2414–2418.

Deng, Z., Norseen, J., Wiedmer, A., Riethman, H., and Lieberman, P.M. (2009). TERRA RNA binding to TRF2 facilitates heterochromatin formation and ORC recruitment at telomeres. Mol. Cell *35*, 403–413.

Denslow, S.A., and Wade, P.A. (2007). The human Mi-2/NuRD complex and gene regulation. Oncogene *26*, 5433–5438.

Dhalluin, C., Carlson, J.E., Zeng, L., He, C., Aggarwal, A.K., and Zhou, M.M. (1999). Structure and ligand of a histone acetyltransferase bromodomain. Nature *399*, 491–496.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380.

Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., et al. (2012). Landscape of transcription in human cells. Nature *489*, 101–108.

Dresner, E., Agam, G., and Gozes, I. (2011). Activity-dependent neuroprotective protein (ADNP) expression level is correlated with the expression of the sister protein ADNP2: Deregulation in schizophrenia. Eur. Neuropsychopharmacol. *21*, 355–361.

Durrin, L.K., Mann, R.K., Kayne, P.S., Grunstein, M., Bradbury, E.M., and Grunstein, M. (1991). Yeast histone H4 N-terminal sequence is required for promoter activation in vivo. Cell *65*, 1023–1031.

Eissenberg, J.C., and Elgin, S.C. (2000). The HP1 protein family: Getting a grip on chromatin. Curr. Opin. Genet. Dev. *10*, 204–210.

Eissenberg, J.C., James, T.C., Foster-Hartnett, D.M., Hartnett, T., Ngan, V., and Elgin, S.C. (1990). Mutation in a heterochromatin-specific chromosomal protein is associated with suppression of position-effect variegation in Drosophila melanogaster. Proc. Natl. Acad. Sci. U. S. A. *87*, 9923–9927.

Endoh, M., Endo, T.A., Shinga, J., Hayashi, K., Farcas, A., Ma, K.W., Ito, S., Sharif, J., Endoh, T., Onaga, N., et al. (2017). PCGF6-PRC1 suppresses premature differentiation of mouse embryonic stem cells by regulating germ cell-related genes. Elife *6*, e21064.

Erdel, F., Schubert, T., Marth, C., Langst, G., and Rippe, K. (2010). Human ISWI chromatin-remodeling complexes sample nucleosomes via transient binding reactions and become immobilized at active sites. Proc. Natl. Acad. Sci. *107*, 19873–19878.

Eskeland, R., Eberharter, A., and Imhof, A. (2007). HP1 binding to chromatin methylated at H3K9 is enhanced by auxiliary factors. Mol. Cell. Biol. *27*, 453–465.

Fan, Y., Nikitina, T., Morin-Kensicki, E.M., Zhao, J., Magnuson, T.R., Woodcock, C.L., and Skoultchi, A.I. (2003). H1 linker histones are essential for mouse development and affect nucleosome spacing in vivo. Mol. Cell. Biol. *23*, 4559–4572.

Felsenfeld, G., and Groudine, M. (2003). Controlling the double helix. Nature *421*, 448–453.

Festenstein, R., Pagakis, S.N., Hiragami, K., Lyon, D., Verreault, A., Sekkali, B., and Kioussis, D. (2003). Modulation of Heterochromatin Protein 1 Dynamics in Primary Mammalian Cells. Science *299*, 719 LP-721.

Filion, G.J., van Bemmel, J.G., Braunschweig, U., Talhout, W., Kind, J., Ward, L.D., Brugman, W., de Castro, I.J., Kerkhoven, R.M., Bussemaker, H.J., et al. (2010). Systematic Protein Location

Mapping Reveals Five Principal Chromatin Types in Drosophila Cells. Cell *143*, 212–224.

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

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

Friedman, J.R., Fredericks, W.J., Jensen, D.E., Speicher, D.W., Huang, X.P., Neilson, E.G., and Rauscher, F.J. (1996). KAP-1, a novel corepressor for the highly conserved KRAB repression domain. Genes Dev. *10*, 2067–2078.

Fujita, N., Jaye, D.L., Geigerman, C., Akyildiz, A., Mooney, M.R., Boss, J.M., and Wade, P.A. (2004). MTA3 and the Mi-2/NuRD complex regulate cell fate during B lymphocyte differentiation. Cell *119*, 75–86.

Fukagawa, T., and Earnshaw, W.C. (2014). Neocentromeres. Curr. Biol. *24*, 946–947.

Fuks, F., Hurd, P.J., Deplus, R., and Kouzarides, T. (2003). The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. Nucleic Acids Res. *31*, 2305–2312.

Galupa, R., and Heard, E. (2015). X-chromosome inactivation: new insights into cis and trans regulation. Curr. Opin. Genet. Dev. *31*, 57–66.

Gao, Z., Zhang, J., Bonasio, R., Strino, F., Sawai, A., Parisi, F., Kluger, Y., and Reinberg, D. (2012). PCGF Homologs, CBX Proteins, and RYBP Define Functionally Distinct PRC1 Family Complexes. Mol. Cell *45*, 344–356.

García-Cao, M., O'Sullivan, R., Peters, A.H.F.M., Jenuwein, T., and Blasco, M.A. (2004). Epigenetic regulation of telomere length in mammalian cells by the Suv39h1 and Suv39h2 histone methyltransferases. Nat. Genet. *36*, 94–99.

Gaszner, M., and Felsenfeld, G. (2006). Insulators: exploiting transcriptional and epigenetic mechanisms. Nat. Rev. Genet. *7*, 703–713.

Godde, J.S., and Widom, J. (1992). Chromatin structure of Schizosaccharomyces pombe. J. Mol. Biol. *226*, 1009–1025.

Goldberg, A.D., Banaszynski, L.A., Noh, K.-M., Lewis, P.W., Elsaesser, S.J., Stadler, S., Dewell, S., Law, M., Guo, X., Li, X., et al. (2010). Distinct Factors Control Histone Variant H3.3 Localization at Specific Genomic Regions. Cell *140*, 678–691.

Grüne, T., Brzeski, J., Eberharter, A., Clapier, C.R., Corona, D.F.V., Becker, P.B., and Müller, C.W. (2003). Crystal structure and functional analysis of a nucleosome recognition module of the remodeling factor ISWI. Mol. Cell *12*, 449–460.

Halder, G., Callaerts, P., and Gehring, W. (1995). Induction of ectopic eyes by targeted expression of the eyeless gene in Drosophila. Science *267*, 1788–1792.

Hale, T.K., Contreras, A., Morrison, A.J., and Herrera, R.E. (2006). Phosphorylation of the Linker Histone H1 by CDK Regulates Its Binding to HP1alpha. Mol. Cell *22*, 693–699.

Hansen, K.H., Bracken, A.P., Pasini, D., Dietrich, N., Gehani, S.S., Monrad, A., Rappsilber, J., Lerdrup, M., and Helin, K. (2008). A model for transmission of the H3K27me3 epigenetic mark. Nat. Cell Biol. *10*, 1291–1300.

Harshman, S.W., Young, N.L., Parthun, M.R., and Freitas, M.A. (2013). H1 histones: current perspectives and challenges. Nucleic Acids Res. *41*, 9593–9609.

Hathaway, N.A., Bell, O., Hodges, C., Miller, E.L., Neel, D.S., and Crabtree, G.R. (2012). Dynamics and memory of heterochromatin in living cells. Cell *149*, 1447–1460.

Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S., and Saitou, M. (2011). Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. Cell *146*, 519–532.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat. Genet. *39*, 311–318.

Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., et al. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature *459*, 108–112.

Hellauer, K., Sirard, E., and Turcotte, B. (2001). Decreased expression of specific genes in yeast cells lacking histone H1. J. Biol. Chem. *276*, 13587–13592.

Helsmoortel, C., Vulto-van Silfhout, A.T., Coe, B.P., Vandeweyer, G., Rooms, L., van den Ende, J., Schuurs-Hoeijmakers, J.H.M., Marcelis, C.L., Willemsen, M.H., Vissers, L.E.L.M., et al. (2014). A SWI/SNF-related autism syndrome caused by de novo mutations in ADNP. Nat. Genet. *46*, 380–384.

Hendrich, B., and Bird, A. (1998). Identification and Characterization of a Family of Mammalian Methyl-CpG Binding Proteins. Mol. Cell. Biol. *18*, 6538–6547.

Higo, S., Asano, Y., Kato, H., Yamazaki, S., Nakano, A., Tsukamoto, O., Seguchi, O., Asai, M., Asakura, M., Asanuma, H., et al. (2010). Isoform-specific intermolecular disulfide bond formation of heterochromatin protein 1 (HP1). J. Biol. Chem. *285*, 31337–31347.

Hiragami-Hamada, K., Shinmyozu, K., Hamada, D., Tatsu, Y., Uegaki, K., Fujiwara, S., and Nakayama, J. -i. (2011). N-Terminal Phosphorylation of HP1alpha Promotes Its Chromatin Binding. Mol. Cell. Biol. *31*, 1186–1200.

Ho, L., and Crabtree, G.R. (2010). Chromatin remodelling during development. Nature *463*, 474–484.

Højfeldt, J.W., Agger, K., and Helin, K. (2013). Histone lysine demethylases as targets for anticancer therapy. Nat. Rev. Drug Discov. *12*, 917–930.

Holliday, R. & Pugh, J.E. (1975). DNA modification mechanisms and gene activity during development. Science (80-. ). *187*, 226–232.

Huang, C., Su, T., Xue, Y., Cheng, C., Lay, F.D., McKee, R.A., Li, M., Vashisht, A., Wohlschlegel, J., Novitch, B.G., et al. (2017). CBX3 maintains lineage specificity during neural differentiation. Genes Dev. *31*, 241–246.

Hurlin, P.J., Steingrımsson, E., Copeland, N.G., Jenkins, N.A., and Eisenman, R.N. (1999). Mga, a dual-specificity transcription factor that interacts with Max and contains a T-domain DNA-binding motif. EMBO J. *18*, 7019–7028.

Ivanova, A. V, Bonaduce, M.J., Ivanov, S. V, and Klar, a J. (1998). The chromo and SET domains of the Clr4 protein are essential for silencing in fission yeast. Nat. Genet. *19*, 192–195.

Iwafuchi-Doi, M., Donahue, G., Kakumanu, A., Watts, J.A., Mahony, S., Pugh, B.F., Lee, D., Kaestner, K.H., and Zaret, K.S. (2016). The Pioneer Transcription Factor FoxA Maintains an Accessible Nucleosome Configuration at Enhancers for Tissue-Specific Gene Activation. Mol. Cell *62*, 79–91.

Jacobs, S.A., and Khorasanizadeh, S. (2002). Structure of HP1 Chromodomain Bound to a Lysine 9–Methylated Histone H3 Tail. Science (80-. ). *295*, 2080–2083.

Jacobs, S.A., Taverna, S.D., Zhang, Y., Briggs, S.D., Li, J., Eissenberg, J.C., Allis, C.D., and Khorasanizadeh, S. (2001). Specificity of the HP1 chromo domain for the methylated N-terminus of histone H3. EMBO J. *20*, 5232–5241.

James, T.C., and Elgin, S.C. (1986). Identification of a nonhistone chromosomal protein associated with heterochromatin in Drosophila melanogaster and its gene. Mol. Cell. Biol. *6*, 3862–3872.

Järvelin, A.I., Noerenberg, M., Davis, I., and Castello, A. (2016). The new (dis)order in RNA regulation. Cell Commun. Signal. *14*, 9.

Jedrusik, M.A., and Schulze, E. (2001). H1.1 in C. elegans chromatin silencing. Development *128*, 1069–1080.

Johnson, W.L., Yewdell, W.T., Bell, J.C., McNulty, S.M., Duda, Z., O'Neill, R.J., Sullivan, B.A., and Straight, A.F. (2017). RNA-dependent stabilization of SUV39H1 at constitutive heterochromatin. Elife *6*, e25299.

Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat. Rev. Genet. *13*, 484–492.

Kadoch, C., and Crabtree, G.R. (2015). Mammalian SWI/SNF chromatin remodeling complexes and cancer: Mechanistic insights gained from human genomics. Sci. Adv. *1*, e1500447–e1500447.

Kapranov, P., Willingham, A.T., and Gingeras, T.R. (2007). Genome-wide transcription and the implications for genomic organization. Nat. Rev. Genet. *8*, 413–423.

Karytinos, A., Forneris, F., Profumo, A., Ciossani, G., Battaglioli, E., Binda, C., and Mattevi, A. (2009). A novel mammalian flavin-dependent histone demethylase. J. Biol. Chem. *284*, 17775–17782.

Kato, M., Kato, Y., Nishida, M., Hayakawa, T., Haraguchi, T., Hiraoka, Y., H Inoue, Y., and Yamaguchi, M. (2007). Functional domain analysis of human HP1 isoforms in Drosophila. Cell Struct. Funct. *32*, 57–67.

Keller, C., Adaixo, R., Stunnenberg, R., Woolcock, K.J., Hiller, S., and Bühler, M. (2012). HP1(Swi6) mediates the recognition and destruction of heterochromatic RNA transcripts. Mol. Cell *47*, 215–227.

Keller, C., Kulasegaran-Shylini, R., Shimada, Y., Hotz, H.-R., and Bühler, M. (2013). Noncoding RNAs prevent spreading of a repressive histone mark. Nat. Struct. Mol. Biol. *20*, 994–1000.

Keppler, B.R., and Archer, T.K. (2008). Chromatin-modifying enzymes as therapeutic targets – Part 1. Expert Opin. Ther. Targets *12*, 1301–1312.

Kilic, S., Bachmann, A.L., Bryan, L.C., and Fierz, B. (2015). Multivalency governs HP1α association dynamics with the silent chromatin state. Nat. Commun. *6*, 1–11.

Klose, R.J., Kallin, E.M., and Zhang, Y. (2006). JmjC-domain-containing proteins and histone demethylation. Nat. Rev. Genet. *7*, 715–727.

Kornberg, R.D. (1974). Chromatin Structure: A Repeating Unit of Histones and DNA. Science (80-. ). *184*, 868–871.

Kornberg, R.D., and Thomas, J.O. (1974). Chromatin structure; oligomers of the histones. Science *184*, 865–868.

Kourmouli, N., Theodoropoulos, P.A., Dialynas, G., Bakou, A., Politou, A.S., Cowell, I.G., Singh, P.B., and Georgatos, S.D. (2000). Dynamic associations of heterochromatin protein 1 with the nuclear envelope. EMBO J. *19*, 6558–6568.

Kouzarides, T. (2007). Chromatin Modifications and Their Function. Cell *128*, 693–705.

Kramer, K., Sachsenberg, T., Beckmann, B.M., Qamar, S., Boon, K.-L., Hentze, M.W., Kohlbacher, O., and Urlaub, H. (2014). Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins. Nat. Methods *11*, 1064–1070.

Kuo, M.H., Brownell, J.E., Sobel, R.E., Ranalli, T.A., Cook, R.G., Edmondson, D.G., Roth, S.Y., and Allis, C.D. (1996). Transcription-linked acetylation by Gcn5p of histones H3 and H4 at specific lysines. Nature *383*, 269–272.

Kushnir, M., Dresner, E., Mandel, S., and Gozes, I. (2008). Silencing of the ADNP-family member, ADNP2, results in changes in cellular viability under oxidative stress. J. Neurochem. *105*, 537–545.

Kwon, S.C., Yi, H., Eichelbaum, K., Föhr, S., Fischer, B., You, K.T., Castello, A., Krijgsveld, J., Hentze, M.W., and Kim, V.N. (2013). The RNA-binding protein repertoire of embryonic stem cells. Nat. Struct. Mol. Biol. *20*, 1122–1130.

de la Serna, I.L., Ohkawa, Y., and Imbalzano, A.N. (2006). Chromatin remodelling in mammalian differentiation: lessons from ATP-dependent remodellers. Nat. Rev. Genet. *7*, 461–473.

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature *410*, 116–120.

Lai, A.Y., and Wade, P.A. (2011). Cancer biology and NuRD: a multifaceted chromatin remodelling complex. Nat. Rev. Cancer *11*, 588–596.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860–921.

Larson, A.G., Elnatan, D., Keenen, M.M., Trnka, M.J., Johnston, J.B., Burlingame, A.L., Agard, D.A., Redding, S., and Narlikar, G.J. (2017). Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. Nature *547*, 236–240.

Lechner, M.S., Begg, G.E., Speicher, D.W., and Rauscher, F.J. (2000). Molecular determinants for targeting heterochromatin protein 1-mediated gene silencing: direct chromoshadow domain-KAP-1 corepressor interaction is essential. Mol. Cell. Biol. *20*, 6449–6465.

Lechner, M.S., Schultz, D.C., Negorev, D., Maul, G.G., and Rauscher, F.J. (2005). The mammalian heterochromatin protein 1 binds diverse nuclear proteins through a common motif that targets the chromoshadow domain. Biochem. Biophys. Res. Commun. *331*, 929–937.

Lee, J.T., Strauss, W.M., Dausman, J.A., and Jaenisch, R. (1996). A 450 kb Transgene Displays Properties of the Mammalian X-Inactivation Center. Cell *86*, 83–94.

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

LeRoy, G., Weston, J.T., Zee, B.M., Young, N.L., Plazas-Mayorca, M.D., and Garcia, B.A. (2009). Heterochromatin Protein 1 Is Extensively Decorated with Histone Code-like Post-translational Modifications. Mol. Cell. Proteomics *8*, 2432–2442.

Li, H., Rauch, T., Chen, Z.X., Szabó, P.E., Riggs, A.D., and Pfeifer, G.P. (2006). The histone methyltransferase SETDB1 and the DNA methyltransferase DNMT3A interact directly and localize to promoters silenced in cancer cells. J. Biol. Chem. *281*, 19489–19500.

Li, X., Wang, W., Wang, J., Malovannaya, A., Xi, Y., Li, W., Guerra, R., Hawke, D.H., Qin, J., and Chen, J. (2015). Proteomic analyses reveal distinct chromatin- associated and soluble transcription factor complexes. Mol Syst Biol *11*, 1–16.

Li, X.-Y., Thomas, S., Sabo, P.J., Eisen, M.B., Stamatoyannopoulos, J.A., and Biggin, M.D. (2011). The role of chromatin accessibility in directing the widespread, overlapping patterns of Drosophila transcription factor binding. Genome Biol. *12*, 1–17.

Lomberk, G., Wallrath, L.L., and Urrutia, R. (2006a). The Heterochromatin Protein 1 family. Genome Biol. *7*, 228.1-228.8.

Lomberk, G., Bensi, D., Fernandez-Zapico, M.E., and Urrutia, R. (2006b). Evidence for the existence of an HP1-mediated subcode within the histone code. Nat. Cell Biol. *8*, 407–415.

Lorch, Y., LaPointe, J.W., and Kornberg, R.D. (1987). Nucleosomes inhibit the initiation of

transcription but allow chain elongation with the displacement of histones. Cell *49*, 203–210.

Loyola, A., Tagami, H., Bonaldi, T., Roche, D., Quivy, J.P., Imhof, A., Nakatani, Y., Dent, S.Y.R., and Almouzni, G. (2009). The HP1α–CAF1–SetDB1-containing complex provides H3K9me1 for Suv39-mediated K9me3 in pericentric heterochromatin. EMBO Rep. *10*, 769–775.

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

Luscombe, N.M., Austin, S.E., Berman, H.M., and Thornton, J.M. (2000). An overview of the structures of protein-DNA complexes. Genome Biol. *1*, 1–19.

Mahadevan, L.C., Willis, A.C., and Barratt, M.J. (1991). Rapid histone H3 phosphorylation in response to growth factors, phorbol esters, okadaic acid, and protein synthesis inhibitors. Cell *65*, 775–783.

Maier, V.K., Feeney, C.M., Taylor, J.E., Creech, A.L., Qiao, J.W., Szanto, A., Das, P.P., Chevrier, N., Cifuentes-Rojas, C., Orkin, S.H., et al. (2015). Functional proteomics defines a PRC2-G9A interaction network and reveals ZNF518B as a G9A regulator. Mol Cell Proteomics *14*, 1435–1446.

Maison, C., Bailly, D., Peters, A.H.F.M., Quivy, J.-P., Roche, D., Taddei, A., Lachner, M., Jenuwein, T., and Almouzni, G. (2002). Higher-order structure in pericentric heterochromatin involves a distinct pattern of histone modification and an RNA component. Nat. Genet. *30*, 329–334.

Maison, C., Bailly, D., Roche, D., Montes de Oca, R., Probst, A. V, Vassias, I., Dingli, F., Lombard, B., Loew, D., Quivy, J.-P., et al. (2011). SUMOylation promotes de novo targeting of HP1α to pericentric heterochromatin. Nat. Genet. *43*, 220–227.

Maison, C., Bailly, D., Quivy, J.-P., and Almouzni, G. (2016). The methyltransferase Suv39h1 links the SUMO pathway to HP1α marking at pericentric heterochromatin. Nat. Commun. *7*, 12224.

Maksakova, I.A., Goyal, P., Bullwinkel, J., Brown, J.P., Bilenky, M., Mager, D.L., Singh, P.B., and Lorincz, M.C. (2011). H3K9me3-binding proteins are dispensable for SETDB1/H3K9me3-dependent retroviral silencing. Epigenetics Chromatin *4*, 1-18.

Maksakova, I.A., Thompson, P.J., Goyal, P., Jones, S.J., Singh, P.B., Karimi, M.M., and Lorincz, M.C. (2013). Distinct roles of KAP1, HP1 and G9a/GLP in silencing of the two-cell-specific retrotransposon MERVL in mouse ES cells. Epigenetics Chromatin *6*, 1–16.

Malik, H.S., and Henikoff, S. (2003). Phylogenomics of the nucleosome. Nat. Struct. Biol. *10*, 882–891.

Mandel, S., and Gozes, I. (2007). Activity-dependent neuroprotective protein constitutes a novel element in the SWI/SNF chromatin remodeling complex. J. Biol. Chem. *282*, 34448–34456.

Mandel, S., Rechavi, G., and Gozes, I. (2007). Activity-dependent neuroprotective protein (ADNP) differentially interacts with chromatin to regulate genes essential for embryogenesis. Dev. Biol. *303*, 814–824.

Marfella, C.G.A., and Imbalzano, A.N. (2007). The Chd family of chromatin remodelers. Mutat. Res. *618*, 30–40.

Margueron, R., Justin, N., Ohno, K., Sharpe, M.L., Son, J., Drury III, W.J., Voigt, P., Martin, S.R., Taylor, W.R., De Marco, V., et al. (2009). Role of the polycomb protein EED in the propagation of repressive histone marks. Nature *461*, 762–767.

Matsui, T., Leung, D., Miyashita, H., Maksakova, I.A., Miyachi, H., Kimura, H., Tachibana, M., Lorincz, M.C., and Shinkai, Y. (2010). Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. Nature *464*, 927–931.

Mattout, A., Aaronson, Y., Sailaja, B.S., Raghu Ram, E. V., Harikumar, A., Mallm, J.P., Sim, K.H., Nissim-Rafinia, M., Supper, E., Singh, P.B., et al. (2015). Heterochromatin Protein 1beta (HP1beta) has distinct functions and distinct nuclear distribution in pluripotent versus differentiated cells.

Genome Biol. *16*, 213.1-213.21.

Matzke, M., Kanno, T., Daxinger, L., Huettel, B., and Matzke, A.J. (2009). RNA-mediated chromatin-based silencing in plants. Curr. Opin. Cell Biol. *21*, 367–376.

Melcher, M., Schmid, M., Aagaard, L., Selenko, P., Laible, G., and Jenuwein, T. (2000). Structure-Function Analysis of SUV39H1 Reveals a Dominant Role in Heterochromatin Organization, Chromosome Segregation, and Mitotic Progression. Mol. Cell. Biol. *20*, 3728–3741.

Mendez, D.L., Mandt, R.E., and Elgin, S.C.R. (2013). Heterochromatin protein 1a (HP1a) partner specificity is determined by critical amino acids in the chromo shadow domain and C-terminal extension. J. Biol. Chem. *288*, 22315–22323.

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

Minc, E., Allory, Y., Worman, H.J., Courvalin, J.C., and Buendia, B. (1999). Localization and phosphorylation of HP1 proteins during the cell cycle in mammalian cells. Chromosoma *108*, 220–234.

Minc, E., Courvalin, J.C., and Buendia, B. (2000). HP1gamma associates with euchromatin and heterochromatin in mammalian nuclei and chromosomes. Cytogenet. Cell Genet. *90*, 279–284.

Mishima, Y., Watanabe, M., Kawakami, T., Jayasinghe, C.D., Otani, J., Kikugawa, Y., Shirakawa, M., Kimura, H., Nishimura, O., Aimoto, S., et al. (2013). Hinge and chromoshadow of HP1α participate in recognition of K9 methylated histone H3 in nucleosomes. J. Mol. Biol. *425*, 54–70.

Mizuguchi, G., Shen, Y., Landry, J., Wei-Hua, W., Sen, S., and Carl, W. (2004). ATP-Driven Exchange of Histone H2AZ Variant Catalyzed by SWR1 Chromatin Remodeling Complex. Science (80-. ). *303*, 343–348.

Mosch, K., Franz, H., Soeroes, S., Singh, P.B., and Fischle, W. (2011). HP1 recruits activity-dependent neuroprotective protein to H3K9me3 marked pericentromeric heterochromatin for silencing of major satellite repeats. PLoS One *6*, e15894.

Muchardt, C., Guilleme, M., Seeler, J.-S., Trouche, D., Dejean, A., and Yaniv, M. (2002). Coordinated methyl and RNA binding is required for heterochromatin localization of mammalian HP1alpha. EMBO Rep. *3*, 975–981.

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

Munari, F., Soeroes, S., Zenn, H.M., Schomburg, A., Kost, N., Schröder, S., Klingberg, R., Rezaei-Ghaleh, N., Stützer, A., Gelato, K.A., et al. (2012). Methylation of lysine 9 in histone H3 directs alternative modes of highly dynamic interaction of heterochromatin protein hHP1β with the nucleosome. J. Biol. Chem. *287*, 33756–33765.

Musselman, C.A., Mansfield, R.E., Garske, A.L., Davrazou, F., Kwan, A.H., Oliver, S.S., O'Leary, H., Denu, J.M., Mackay, J.P., and Kutateladze, T.G. (2009). Binding of the CHD4 PHD2 finger to histone H3 is modulated by covalent modifications. Biochem. J. *423*, 179–187.

Nakayama, J., Rice, J.C., Strahl, B.D., Allis, C.D., and Grewal, S.I.S. (2001). Role of Histone H3 Lysine 9 Methylation in Epigenetic Control of Heterochromatin Assembly. Science *292*, 110–113.

Naruse, C., Fukusumi, Y., Kakiuchi, D., and Asano, M. (2007). A novel gene trapping for identifying genes expressed under the control of specific transcription factors. Biochem. Biophys. Res. Commun. *361*, 109–115.

Nielsen, A.L., Oulad-Abdelghani, M., Ortiz, J.A., Remboutsika, E., Chambon, P., and Losson, R. (2001). Heterochromatin formation in mammalian cells: Interaction between histones and HP1

Proteins. Mol. Cell *7*, 729–739.

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

Nora, E.P., Goloborodko, A., Valton, A.L., Gibcus, J.H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L.A., and Bruneau, B.G. (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. Cell *169*, 930–944.

Nozawa, R.-S., Nagao, K., Masuda, H.-T., Iwasaki, O., Hirota, T., Nozaki, N., Kimura, H., and Obuse, C. (2010). Human POGZ modulates dissociation of HP1alpha from mitotic chromosome arms through Aurora B activation. Nat. Cell Biol. *12*, 719–727.

O'Shaughnessy-Kirwan, A., Signolet, J., Costello, I., Gharbi, S., and Hendrich, B. (2015). Constraint of gene expression by chromatin remodelling protein CHD4 facilitates lineage specification. Development *142*, 2586–2597.

Ogawa, H., Ishiguro, K., Gaubatz, S., Livingston, D.M., and Nakatani, Y. (2002). A Complex with Chromatin Modifiers That Occupies E2F- and Myc-Responsive Genes in G0 Cells. Science (80-. ). *296*, 1132–1136.

Ohta, H., Kurimoto, K., Okamoto, I., Nakamura, T., Yabuta, Y., Miyauchi, H., Yamamoto, T., Okuno, Y., Hagiwara, M., Shirane, K., et al. (2017). In vitro expansion of mouse primordial germ cell-like cells recapitulates an epigenetic blank slate. EMBO J. *36*, 1888–1907.

Papamichos-Chronakis, M., Watanabe, S., Rando, O.J., and Peterson, C.L. (2011). Global regulation of H2A.Z localization by the INO80 chromatin-remodeling enzyme is essential for genome integrity. Cell *144*, 200–213.

Parker, M.W., Bello, M. Lo, and Federici, G. (1990). Crystallization of glutathione S-transferase from human placenta. J. Mol. Biol. *213*, 221–222.

Paro, R., and Hognesst, D.S. (1991). The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of Drosophila. Dev. Biol. *88*, 263–267.

Patterton, H.G., Landel, C.C., Landsman, D., Peterson, C.L., and Simpson, R.T. (1998). The biochemical and phenotypic characterization of Hho1p, the putative linker histone H1 of Saccharomyces cerevisiae. J. Biol. Chem. *273*, 7268–7276.

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

Peters, A.H.F.M., Kubicek, S., Mechtler, K., O'Sullivan, R.J., Derijck, A.A.H.A., Perez-Burgos, L., Kohlmaier, A., Opravil, S., Tachibana, M., Shinkai, Y., et al. (2003). Partitioning and Plasticity of Repressive Histone Methylation States in Mammalian Chromatin. Mol. Cell *12*, 1577–1589.

Pinhasov, A., Mandel, S., Torchinsky, A., Giladi, E., Pittel, Z., Goldsweig, A.M., Servoss, S.J., Brenneman, D.E., and Gozes, I. (2003). Activity-dependent neuroprotective protein: A novel gene essential for brain formation. Dev. Brain Res. *144*, 83–90.

Platero, J.S., Hartnett, T., and Eissenberg, J.C. (1995). Functional analysis of the chromo domain of HP1. EMBO J. *14*, 3977–3986.

Polo, S.E., Roche, D., and Almouzni, G. (2006). New Histone Incorporation Marks Sites of UV Repair in Human Cells. Cell *127*, 481–493.

Pradhan, S.K., Su, T., Yen, L., Jacquet, K., Huang, C., Côté, J., Kurdistani, S.K., and Carey, M.F. (2016). EP400 Deposits H3.3 into Promoters and Enhancers during Gene Activation. Mol. Cell *61*, 27–38.

Prasanth, S.G., Shen, Z., Prasanth, K. V, and Stillman, B. (2010). Human origin recognition complex is essential for HP1 binding to chromatin and heterochromatin organization. Proc. Natl. Acad. Sci.

*107*, 15093–15098.

Ptashne, M. (2007). On the use of the word "epigenetic." Curr. Biol. *17*, 233–236.

Ptashne, M. (2013). Epigenetics: Core misconcept. Proc. Natl. Acad. Sci. *110*, 7101–7103.

Qin, J., Whyte, W.A., Anderssen, E., Apostolou, E., Chen, H.H., Akbarian, S., Bronson, R.T., Hochedlinger, K., Ramaswamy, S., Young, R.A., et al. (2012). The polycomb group protein L3mbtl2 assembles an atypical PRC1-family complex that is essential in pluripotent stem cells and early development. Cell Stem Cell *11*, 319–332.

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B.D., Sun, Z.W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C.P., Allis, C.D., et al. (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature *406*, 593–599.

Rice, J.C., Briggs, S.D., Ueberheide, B., Barber, C.M., Shabanowitz, J., Hunt, D.F., Shinkai, Y., and Allis, C.D. (2003). Histone Methyltransferases Direct Different Degrees of Methylation to Define Distinct Chromatin Domains. Mol. Cell *12*, 1591–1598.

Riggs, A.D. (1975). X inactivation, differentiation, and DNA methylation. Cytogenet. Genome Res. *14*, 9–25.

Riggs, A.D., and Porter, T.N. (1996). Overview of Epigenetic Mechanisms. On Epigenetic Mechanisms of Gene Regulation, (Cold Spring Harbor Laboratory Press), pp. 29–45.

Robinson, P.J., and Rhodes, D. (2006). Structure of the "30nm" chromatin fibre: A key role for the linker histone. Curr. Opin. Struct. Biol. *16*, 336–343.

Rosnoblet, C., Vandamme, J., Völkel, P., and Angrand, P.-O. (2011). Analysis of the human HP1 interactome reveals novel binding partners. Biochem. Biophys. Res. Commun. *413*, 206–211.

Rowe, H.M., Jakobsson, J., Mesnard, D., Rougemont, J., Reynard, S., Aktas, T., Maillard, P. V, Layard-Liesching, H., Verp, S., Marquis, J., et al. (2010). KAP1 controls endogenous retroviruses in embryonic stem cells. Nature *463*, 237–240.

Ryan, D.P., Sundaramoorthy, R., Martin, D., Singh, V., and Owen-Hughes, T. (2011). The DNA-binding domain of the Chd1 chromatin-remodelling enzyme contains SANT and SLIDE domains. EMBO J. *30*, 2596–2609.

Sampath, S.C., Marazzi, I., Yap, K.L., Sampath, S.C., Krutchinsky, A.N., Mecklenbräuker, I., Viale, A., Rudensky, E., Zhou, M.-M., Chait, B.T., et al. (2007). Methylation of a histone mimic within the histone methyltransferase G9a regulates protein complex assembly. Mol. Cell *27*, 596–608.

Santenard, A., Ziegler-Birling, C., Koch, M., Tora, L., Bannister, A.J., and Torres-Padilla, M.-E. (2010). Heterochromatin formation in the mouse embryo requires critical residues of the histone variant H3.3. Nat. Cell Biol. *12*, 853–862.

Sarraf, S.A., and Stancheva, I. (2004). Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. Mol. Cell *15*, 595–605.

Schmitges, F.W., Prusty, A.B., Faty, M., Stützer, A., Lingaraju, G.M., Aiwazian, J., Sack, R., Hess, D., Li, L., Zhou, S., et al. (2011). Histone Methylation by PRC2 Is Inhibited by Active Chromatin Marks. Mol. Cell *42*, 330–341.

Schotta, G., Ebert, A., Krauss, V., Fischer, A., Hoffmann, J., Rea, S., Jenuwein, T., Dorn, R., and Reuter, G. (2002). Central role of Drosophila SU(VAR)3-9 in histone H3-K9 methylation and heterochromatic gene silencing. EMBO J. *21*, 1121–1131.

Schotta, G., Lachner, M., Sarma, K., Ebert, A., Sengupta, R., Reuter, G., Reinberg, D., and Jenuwein, T. (2004). A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. Genes Dev. *18*, 1251–1262.

Schubert, H.L., Wittmeyer, J., Kasten, M.M., Hinata, K., Rawling, D.C., Héroux, A., Cairns, B.R., and Hill, C.P. (2013). Structure of an actin-related subcomplex of the SWI/SNF chromatin remodeler. Proc. Natl. Acad. Sci. U. S. A. *110*, 3345–3350.

Schultz, D.C., Ayyanathan, K., Negorev, D., Maul, G.G., and Rauscher Iii, F.J. (2002). SETDB1 : a novel KAP - 1 - associated histone H3 , lysine 9 - specific methyltransferase that contributes to HP1 - mediated silencing of euchromatic genes by KRAB zinc - finger proteins. Genes Dev. *16*, 919–932.

Schuster, E.F., and Stöger, R. (2002). CHD5 defines a new subfamily of chromodomain-SW12/SNF2-like helicases. Mamm. Genome *13*, 117–119.

Shema, E., Jones, D., Shoresh, N., Donohue, L., Ram, O., and Bernstein, B.E. (2016). Single-molecule decoding of combinatorially modified nucleosomes. Science *352*, 717–721.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., Casero, R.A., and Shi, Y. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell *119*, 941–953.

Shimada, A., Dohke, K., Sadaie, M., Shinmyozu, K., Nakayama, J.I., Urano, T., and Murakami, Y. (2009). Phosphorylation of Swi6/HP1 regulates transcriptional gene silencing at heterochromatin. Genes Dev. *23*, 18–23.

Shirai, A., Kawaguchi, T., Shimojo, H., Muramatsu, D., Ishida-Yonetani, M., Nishimura, Y., Kimura, H., Nakayama, J., and Shinkai, Y. (2017). Impact of nucleic acid and methylated H3K9 binding activities of Suv39h1 on its heterochromatin assembly. Elife *6*, e25317.

Shur, I., and Benayahu, D. (2005). Characterization and functional analysis of CReMM, a novel chromodomain helicase DNA-binding protein. J. Mol. Biol. *352*, 646–655.

Sigalov, E., Fridkin, M., Brenneman, D.E., and Gozes, I. (2001). VIP-Related Protection Against Iodoacetate Toxicity in Pheochromocytoma ( PC12 ) Cells. J. Mol. Neurosci. *15*, 147–154.

Singh, P.B., Miller, J.R., Pearce, J., Kothary, R., Burton, R.D., Paro, R., and James, T.C. (1991). A sequence motif found in a Drosophila heterochromatin protein is conserved in animals and plants. Nucleic Acids Res. *19*, 789–794.

Skourti-Stathaki, K., Kamieniarz-Gdula, K., and Proudfoot, N.J. (2014). R-loops induce repressive chromatin marks over mammalian gene terminators. Nature *516*, 436–439.

Slattery, M., Zhou, T., Yang, L., Dantas Machado, A.C., Gordân, R., and Rohs, R. (2014). Absence of a simple code: How transcription factors read the genome. Trends Biochem. Sci. *39*, 381–399.

Smallwood, A., Estève, P.O., Pradhan, S., and Carey, M. (2007). Functional cooperation between HP1 and DNMT1 mediates gene silencing. Genes Dev. *21*, 1169–1178.

Smallwood, A., Hon, G.C., Jin, F., Henry, R.E., Espinosa, J.M., and Ren, B. (2012). CBX3 regulates efficient RNA processing genome-wide. Genome Res. *22*, 1426–1436.

Smothers, J.F., and Henikoff, S. (2001). The hinge and chromo shadow domain impart distinct targeting of HP1-like proteins. Mol. Cell. Biol. *21*, 2555–2569.

Song, F., Chen, P., Sun, D., Wang, M., Dong, L., Liang, D., Xu, R.-M., Zhu, P., and Li, G. (2014). Cryo-EM Study of the Chromatin Fiber Reveals a Double Helix Twisted by Tetranucleosomal Units. Science *344*, 376–380.

Soufi, A., Donahue, G., and Zaret, K.S. (2012). Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. Cell *151*, 994–1004.

Soufi, A., Garcia, M.F., Jaroszewicz, A., Osman, N., Pellegrini, M., and Zaret, K.S. (2015). Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. Cell *161*, 555–568.

Sripathy, S.P., Stevens, J., and Schultz, D.C. (2006). The KAP1 Corepressor Functions To

Coordinate the Assembly of De Novo HP1-Demarcated Microenvironments of Heterochromatin Required for KRAB Zinc Finger Protein-Mediated Transcriptional Repression. Mol. Cell. Biol. *26*, 8623–8638.

Stedman, E., and Stedman, E. (1950). Cell Specificity of Histones. Nature *166*, 780–781.

van Steensel, B., and Belmont, A.S. (2017). Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. Cell *169*, 780–791.

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

Strahl, B.D., Ohba, R., Cook, R.G., and Allis, C.D. (1999). Methylation of histone H3 at lysine 4 is highly conserved and correlates with transcriptionally active nuclei in Tetrahymena. Proc. Natl. Acad. Sci. *96*, 14967–14972.

Struhl, K. (1999). Fundamentally different logic of gene regulation in eukaryotes and prokaryotes. Cell *98*, 1–4.

Stunnenberg, R., Kulasegaran-Shylini, R., Keller, C., Kirschmann, M.A., Gelman, L., and Bühler, M. (2015). H3K9 methylation extends across natural boundaries of heterochromatin in the absence of an HP1 protein. EMBO J. *34*, 2789–2803.

Tachibana, M., Sugimoto, K., Nozaki, M., Ueda, J., Ohta, T., Ohki, M., Fukuda, M., Takeda, N., Niida, H., Kato, H., et al. (2002). G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. Genes Dev. *16*, 1779–1791.

Tachibana, M., Ueda, J., Fukuda, M., Takeda, N., Ohta, T., Iwanari, H., Sakihama, T., Kodama, T., Hamakubo, T., and Shinkai, Y. (2005). Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9. Genes Dev. *19*, 815–826.

Tagami, H., Ray-Gallet, D., Almouzni, G., and Nakatani, Y. (2004). Histone H3.1 and H3.3 Complexes Mediate Nucleosome Assembly Pathways Dependent or Independent of DNA Synthesis. Cell *116*, 51–61.

Takahashi, K., and Yamanaka, S. (2006). Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. Cell *126*, 663–676.

Talbert, P.B., and Henikoff, S. (2010). Histone variants — ancient wrap artists of the epigenome. Nat. Rev. Mol. Cell Biol. *11*, 264–275.

Talbert, P.B., and Henikoff, S. (2016). Histone variants on the move: substrates for chromatin dynamics. Nat. Rev. Mol. Cell Biol. *18*, 115–126.

Tan, M., Luo, H., Lee, S., Jin, F., Yang, J.S., Montellier, E., Buchou, T., Cheng, Z., Rousseaux, S., Rajagopal, N., et al. (2011). Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. Cell *146*, 1016–1028.

Taunton, J., Hassig, C.A., and Schreiber, S.L. (1996). A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. Science (80-. ). *272*, 408–411.

Thiru, A., Nietlispach, D., Mott, H.R., Okuwaki, M., Lyon, D., Nielsen, P.R., Hirshberg, M., Verreault, A., Murzina, N. V, and Laue, E.D. (2004). Structural basis of HP1/PXVXL motif peptide interactions and HP1 localisation to heterochromatin. EMBO J. *23*, 489–499.

Trojer, P., and Reinberg, D. (2007). Facultative Heterochromatin: Is There a Distinctive Molecular Signature? Mol. Cell *28*, 1–13.

Trojer, P., Cao, A.R., Gao, Z., Li, Y., Zhang, J., Xu, X., Li, G., Losson, R., Erdjument-Bromage, H., Tempst, P., et al. (2011). L3MBTL2 Protein Acts in Concert with PcG Protein-Mediated Monoubiquitination of H2A to Establish a Repressive Chromatin Structure. Mol. Cell *42*, 438–450.

Tschiersch, B., Hofmann, A., Krauss, V., Dorn, R., Korge, G., and Reuter, G. (1994). The protein

encoded by the Drosophila position-effect variegation suppressor gene Su(var)3-9 combines domains of antagonistic regulators of homeotic gene complexes. EMBO J. *13*, 3822–3831.

Vakoc, C.R., Mandat, S.A., Olenchock, B.A., and Blobel, G.A. (2005). Histone H3 lysine 9 methylation and HP1gamma are associated with transcription elongation through mammalian chromatin. Mol. Cell *19*, 381–391.

Velazquez Camacho, O., Galan, C., Swist-Rosowska, K., Ching, R., Gamalinda, M., Karabiber, F., De La Rosa-Velazquez, I., Engist, B., Koschorz, B., Shukeir, N., et al. (2017). Major satellite repeat RNA stabilize heterochromatin retention of Suv39h enzymes by RNA-nucleosome association and RNA:DNA hybrid formation. Elife *6*, e25293.

Venter, J., Adams, M., Myers, E., Li, P., Mural, R., Sutton, G., Smith, H., Yandell, M., Evans, C., Holt, R., et al. (2001). The Sequence of the Human Genome. Science *291*, 1304.

Vermeulen, M., Eberl, H.C., Matarese, F., Marks, H., Denissov, S., Butter, F., Lee, K.K., Olsen, J. V, Hyman, A.A., Stunnenberg, H.G., et al. (2010). Quantitative Interaction Proteomics and Genome-wide Profiling of Epigenetic Histone Marks and Their Readers. Cell *142*, 967–980.

Verschure, P.J., Kraan, I. Van Der, Leeuw, W. De, Vlag, J. Van Der, Carpenter, A.E., and Belmont, A.S. (2005). In Vivo HP1 Targeting Causes Large-Scale Chromatin Condensation and Enhanced Histone Lysine Methylation In Vivo HP1 Targeting Causes Large-Scale Chromatin Condensation and Enhanced Histone Lysine Methylation †. Mol. Cell. Biol. *25*, 4552–4564.

Voigt, P., LeRoy, G., Drury, W.J., Zee, B.M., Son, J., Beck, D.B., Young, N.L., Garcia, B.A., and Reinberg, D. (2012). Asymmetrically modified nucleosomes. Cell *151*, 181–193.

Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I.S., and Martienssen, R.A. (2002). Regulation of Heterochromatic Silencing and Histone H3 Lysine-9 Methylation by RNAi. Science *297*, 1833–1837.

Voullaire, L.E., Slater, H.R., Petrovic, V., and Choo, K.H. (1993). A functional marker centromere with no detectable alpha-satellite, satellite III, or CENP-B protein: activation of a latent centromere? Am. J. Hum. Genet. *52*, 1153–1163.

Vuzman, D., and Levy, Y. (2012). Intrinsically disordered regions as affinity tuners in protein–DNA interactions. Mol. BioSyst. *8*, 47–57.

Waddington, C. (1942). The epigenotype. Endeavour.

Wang, H.B., and Zhang, Y. (2001). Mi2, an auto-antigen for dermatomyositis, is an ATP-dependent nucleosome remodeling factor. Nucleic Acids Res *29*, 2517–2521.

Wang, G., Ma, A., Chow, C.M., Horsley, D., Brown, N.R., Cowell, I.G., and Singh, P.B. (2000). Conservation of heterochromatin protein 1 function. Mol. Cell. Biol. *20*, 6970–6983.

Wang, Z., Zang, C., Rosenfeld, J.A., Schones, D.E., Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Peng, W., Zhang, M.Q., et al. (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. Nat. Genet. *40*, 897–903.

Whyte, W.A., Bilodeau, S., Orlando, D.A., Hoke, H.A., Frampton, G.M., Foster, C.T., Cowley, S.M., and Young, R.A. (2012). Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. Nature *482*, 221.

Wood, V., Gwilliam, R., Rajandream, M.-A., Lyne, M., Lyne, R., Stewart, A., Sgouros, J., Peat, N., Hayles, J., Baker, S., et al. (2002). The genome sequence of Schizosaccharomyces pombe. Nature *415*, 871–880.

Woodage, T., Basrai, M.A., Baxevanis, A.D., Hieter, P., and Collins, F.S. (1997). Characterization of the CHD family of proteins. Proc. Natl. Acad. Sci. *94*, 11472–11477.

Workman, J.L., and Kingston, R.E. (1998). Alteration of nucleosome structure as a mechanism of transcriptional regulation. Annu. Rev. Biochem. *67*, 545–579.

Xiao, H., Sandaltzopoulos, R., Wang, H.M., Hamiche, A., Ranallo, R., Lee, K.M., Fu, D., and Wu, C. (2001). Dual function of largest NURF subunit NURF 301 in nucleosome sliding and transcriptional intereactions. Mol Cell *8*, 531–543.

Yamada, T., Fischle, W., Sugiyama, T., Allis, C.D., and Grewal, S.I.S. (2005). The nucleation and maintenance of heterochromatin by a histone deacetylase in fission yeast. Mol. Cell *20*, 173–185.

Yamamoto, K., and Sonoda, M. (2003). Self-interaction of heterochromatin protein 1 is required for direct binding to histone methyltransferase, SUV39H1. Biochem. Biophys. Res. Commun. *301*, 287–292.

Yamane, K., Toumazou, C., Tsukada, Y. ichi, Erdjument-Bromage, H., Tempst, P., Wong, J., and Zhang, Y. (2006). JHDM2A, a JmjC-Containing H3K9 Demethylase, Facilitates Transcription Activation by Androgen Receptor. Cell *125*, 483–495.

Yang, X.-J., and Seto, E. (2007). HATs and HDACs: from structure, function and regulation to novel strategies for therapy and prevention. Oncogene *26*, 5310–5318.

Ye, Q., and Worman, H.J. (1996). Interaction between an integral protein of the nuclear envelope inner membrane and human chromodomain proteins homologous to Drosophila HP1. J. Biol. Chem. *271*, 14653–14656.

Yildirim, O., Li, R., Hung, J.-H., Chen, P.B., Dong, X., Ee, L.-S., Weng, Z., Rando, O.J., and Fazzio, T.G. (2011). Mbd3/NURD Complex Regulates Expression of 5-Hydroxymethylcytosine Marked Genes in Embryonic Stem Cells. Cell *147*, 1498–1510.

Yuan, G.-C., Liu, Y.-J., Dion, M.F., Slack, M.D., Wu, L.F., Altschuler, S.J., and Rando, O.J. (2005). Genome-Scale Identification of Nucleosome Positions in S. cerevisiae. Science *309*, 626–630.

Zamostiano, R., Pinhasov, A., Gelber, E., Steingart, R.A., Seroussi, E., Giladi, E., Bassan, M., Wollman, Y., Eyre, H.J., Mulley, J.C., et al. (2001). Cloning and characterazation of the human activity-dependent neuroprotective protein. J. Biol. Chem. *276*, 708–714.

Zaret, K.S., and Carroll, J.S. (2011). Pioneer transcription factors: establishing competence for gene expression. Genes Dev. *25*, 2227–2241.

Zaret, K.S., and Mango, S.E. (2016). Pioneer transcription factors, chromatin dynamics, and cell fate control. Curr. Opin. Genet. Dev. *37*, 76–81.

Zhang, T., Termanis, A., Ozkan, B., Bao, X.X., Culley, J., de Lima Alves, F., Rappsilber, J., Ramsahoye, B., and Stancheva, I. (2016). G9a/GLP Complex Maintains Imprinted DNA Methylation in Embryonic Stem Cells. Cell Rep. *15*, 77–85.

Zhao, Y., and Garcia, B.A. (2015). Comprehensive catalog of currently documented histone modifications. Cold Spring Harb. Perspect. Biol. *7*, 1–20.

Zhao, T., Heyduk, T., and Eissenberg, J.C. (2001). Phosphorylation Site Mutations in Heterochromatin Protein 1 (HP1) Reduce or Eliminate Silencing Activity. J. Biol. Chem. *276*, 9512–9518.

Zlatanova, J., Bishop, T.C., Victor, J.-M., Jackson, V., and van Holde, K. (2009). The Nucleosome Family: Dynamic and Growing. Structure *17*, 160–171.

# *Acknowledgements*

First and foremost, I would like to thank to my principal supervisor and advisor, Marc Bühler, for giving me the opportunity to undertake my PhD in his laboratory. I particularly appreciate his open door policy, which meant that he was always there, excited to discuss my projects. In addition, I am grateful for his patience, support and enthusiasm during the past years that culminated in this thesis. Next, I would like to thank to the members of my thesis committee, Antoine H. Peters, Jörg Betschinger, Wolfgang Fischle, and Wolf Reik, for taking their time, contributing ideas and suggestions during the thesis committee meetings and for evaluating the thesis. My thanks go also to Dirk Schübeler for chairing my PhD defence.

I would like to express my thanks to everyone who contributed to and/or helped with my PhD projects; particularly to Claudia Keller who supervised me in the first years, teaching me biochemistry; to Ricardo Aidaxo, Morgan Callon, Antoine Clery, Mauro Zurini, Sebastian Hiller, and to Beat Fierz for collaborating on my first project; and to Susan Gasser's lab, Jeff Chao's lab and Nico Thomae's lab for sharing their equipment. Next, I would to thank everyone who contributed to the second project, which resulted in the manuscript that is part of this thesis: Sarah H. Carl, Anja Basters, Fabio Mohn, Daniel Hess, Vytautas Iesmantavicius, Lisa Lampersberger, Yukiko Shimada, Nicolas H. Thomä, Jörg Betschinger and his lab members, and special big thanks goes to Matyas Flemr for being tissue culture and genome editing guru, and for sharing his insights. I would also like to thank to Alex Tuck for helping me out with deep-seq analysis, and especially for proofreading my thesis.

I would like to thank to all the former and current members of the Bühler lab for creating a stimulating working atmosphere that was at the same time very friendly and fun. Former members: Raghavendran Kulasegaran, Julien Bethune, Daniele Oberti, Katrina Woolcock, Claudia Keller, Rieka Stunnenberg, Katarzyna Kowalik, Rodrigo Villasenor Molino, Adriano Biasini, Alice Wenger, Silke Lochs, Lisa Lampersberger and Scott Berry. Current members: Yukiko Shimada, Tahsin Kuzdere, Ausma Termanis, Matyas Flemr, Lukas Kaaji, Philip Knuckles, Alex Tuck, Lea Duempelmann, Valentin Flury, Aneliya Rankova, Nathalie Laschet, Fabio Mohn, and Gizem Incekara. Future member: Aparna Pandey who is joining us soon to continue working on ChAHP, best of luck in your PhD!

Further thanks goes to the great FMI facilities and their members, particularly to Protein Analysis Facility, Genomics Facility, Protein Structure Facility, FACS, and to our Technical Facilities.

Additionally, I would like to thank to Elida Keller and Piera Cicchetti for being incredibly caring and supportive PhD counsellors.

Personally, I am very grateful for having a chance to work in such a great lab, which is outgoing and crazy and fun and never boring, special thanks goes to Yuki for tolerating it, and often even encouraging it! Thank you all for all the BBQs, Rhine swimming, hiking, skiing, movie nights, fondue evenings, coffee breaks, and weird off topic discussions in our office. Specially, I thank Rieka and Vale for being such great friends, to Matyas for not leaving me behind in the mountains, and to Ani for our long lunch breaks.

Finally, I would like to thank to Dima for all his support and encouragement

# *Appendix*

LETTER

https://doi.org/10.1038/s41586-018-0153-8

# Activity-dependent neuroprotective protein recruits HP1 and CHD4 to control lineage-specifying genes

Veronika Ostapcuk[1,2], Fabio Mohn[1], Sarah H. Carl[1,3,5], Anja Basters[1,5], Daniel Hess[1], Vytautas Iesmantavicius[1], Lisa Lampersberger[1,4], Matyas Flemr[1], Aparna Pandey[1,2], Nicolas H. Thomä[1], Joerg Betschinger[1] & Marc Bühler[1,2]*

**De novo mutations in *ADNP*, which encodes activity-dependent neuroprotective protein (ADNP), have recently been found to underlie Helsmoortel–Van der Aa syndrome, a complex neurological developmental disorder that also affects several other organ functions[1]. ADNP is a putative transcription factor that is essential for embryonic development[2]. However, its precise roles in transcriptional regulation and development are not understood. Here we show that ADNP interacts with the chromatin remodeller CHD4 and the chromatin architectural protein HP1 to form a stable complex, which we refer to as ChAHP. Besides mediating complex assembly, ADNP recognizes DNA motifs that specify binding of ChAHP to euchromatin. Genetic ablation of ChAHP components in mouse embryonic stem cells results in spontaneous differentiation concomitant with premature activation of lineage-specific genes and in a failure to differentiate towards the neuronal lineage. Molecularly, ChAHP-mediated repression is fundamentally different from canonical HP1-mediated silencing: HP1 proteins, in conjunction with histone H3 lysine 9 trimethylation (H3K9me3), are thought to assemble broad heterochromatin domains that are refractory to transcription. ChAHP-mediated repression, however, acts in a locally restricted manner by establishing inaccessible chromatin around its DNA-binding sites and does not depend on H3K9me3-modified nucleosomes. Together, our results reveal that ADNP, via the recruitment of HP1 and CHD4, regulates the expression of genes that are crucial for maintaining distinct cellular states and assures accurate cell fate decisions upon external cues. Such a general role of ChAHP in governing cell fate plasticity may explain why *ADNP* mutations affect several organs and body functions and contribute to cancer progression[1,3,4]. Notably, we found that the integrity of the ChAHP complex is disrupted by nonsense mutations identified in patients with Helsmoortel–Van der Aa syndrome, and this could be rescued by aminoglycosides that suppress translation termination[5]. Therefore, patients might benefit from therapeutic agents that are being developed to promote ribosomal read-through of premature stop codons[6,7].**

ADNP contains nine N-terminal zinc-fingers and a C-terminal homeobox domain, strongly suggesting transcription factor activity[8]. Although originally associated with neuronal function[9], ADNP is essential for embryonic development in mice: *Adnp*-deficient mouse embryos exhibit neural tube closure defects and die at days 8.5–9.5 of gestation. Two studies in knockout mouse embryos identified potential ADNP target genes that are implicated in cell differentiation and the maintenance of stem cells[2,10].

To dissect the molecular activity of ADNP, we exploited mouse embryonic stem (ES) cells[11]. We first inserted a Flag-AviTag at the endogenous *Adnp* gene[12] (Extended Data Fig. 1a–c) and performed chromatin immunoprecipitation coupled to next-generation sequencing (ChIP–seq) to interrogate putative ADNP–DNA interactions genome-wide. This revealed 15,026 sites that are significantly enriched for ADNP (Fig. 1a, b and Supplementary Table 1). Notably, most (61%) of the peaks were found in introns or proximal of annotated transcription start sites. The remaining peaks were located promoter distal in intergenic regions (Extended Data Fig. 1d, e). To analyse the function of ADNP, we generated homozygous *Adnp* knockout mouse ES cells (Extended Data Fig. 2a, b). Compared with wild-type ES cells, $Adnp^{-/-}$ cells displayed gross morphological changes and appeared to differentiate spontaneously as they started spreading out of characteristically densely packed ES cell colonies (Fig. 1c, d). In addition, $Adnp^{-/-}$ cells displayed heterogeneous activity of the pluripotency associated marker alkaline phosphatase (Fig. 1d). Transcriptome profiling by RNA sequencing (RNA-seq) revealed that most of the genes with altered mRNA levels in $Adnp^{-/-}$ cells were upregulated (Extended Data Fig. 2c). Many genes bound by ADNP and displaying increased expression in the absence of ADNP encode known lineage specification factors, such as GATA4, GATA6, BMP1 or SOX17 (Supplementary Table 2). For example, *Gata4* is expressed predominantly in mesoderm- and endoderm-derived tissues[13], and forced *Gata4* expression in mouse ES cells induces differentiation towards extra-embryonic endoderm[14]. Moreover, genes upregulated in $Adnp^{-/-}$ cells were enriched for Gene Ontology terms related to differentiation and development (Extended Data Fig. 2d and Supplementary Table 3). We also observed a group of genes that were upregulated both in $Adnp^{-/-}$ cells as well as in extra-embryonic endoderm stem-cell lines, which can be differentiated from mouse ES cells[15] (Extended Data Fig. 2e). To gain further insight into the biological role of ADNP, we differentiated wild-type and $Adnp^{-/-}$ ES cells towards neuronal precursor cells (Fig. 1c, d) using an established differentiation protocol[16]. $Adnp^{-/-}$ ES cells formed smaller embryoid bodies and showed increased cell death after differentiation when compared to wild-type cells (Fig. 1d). *Nanog* and *Oct4* (also known as *Pou5f1*) expression was downregulated in both wild-type and $Adnp^{-/-}$ cells, indicating successful exit from pluripotency (Fig. 1e). However, whereas $Adnp^{+/+}$ cells started expressing neural markers such as *Pax6* and *Ngn2* (also known as *Neurog2*) over the course of differentiation, $Adnp^{-/-}$ cells failed to induce neural genes (Fig. 1f). Instead, the expression of *Gata4* and *Sox17* was specifically induced in $Adnp^{-/-}$ cells (Fig. 1g), indicating misspecification towards the endodermal lineage under conditions that normally induce neuronal fate. This ES cell phenotype is reminiscent of $Adnp^{-/-}$ mouse embryos, which show a developmental delay, fail to induce *Pax6* and suffer from defective neural tube closure[2]. Thus, ADNP is required to restrain the expression of lineage-specifying genes in ES cells and for specification towards the neuronal lineage upon external differentiation cues.

These results are consistent with previously reported repressive activity of ADNP when artificially targeted to a reporter gene[17]. Furthermore, ADNP was shown to co-immunoprecipitate with the SWI/SNF chromatin remodelling complex[18] or with proteins of the heterochromatin protein 1 (HP1) family[10,19]. To unambiguously identify ADNP-interacting proteins in mouse ES cells, we subjected ADNP tagged endogenously with a Flag-AviTag to tandem-affinity purification coupled to liquid chromatography tandem mass spectrometry

[1]Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland. [2]University of Basel, Basel, Switzerland. [3]Swiss Institute of Bioinformatics, Basel, Switzerland. [4]University of Vienna, Vienna, Austria. [5]These authors contributed equally: Sarah H. Carl, Anja Basters. *e-mail: marc.buehler@fmi.ch

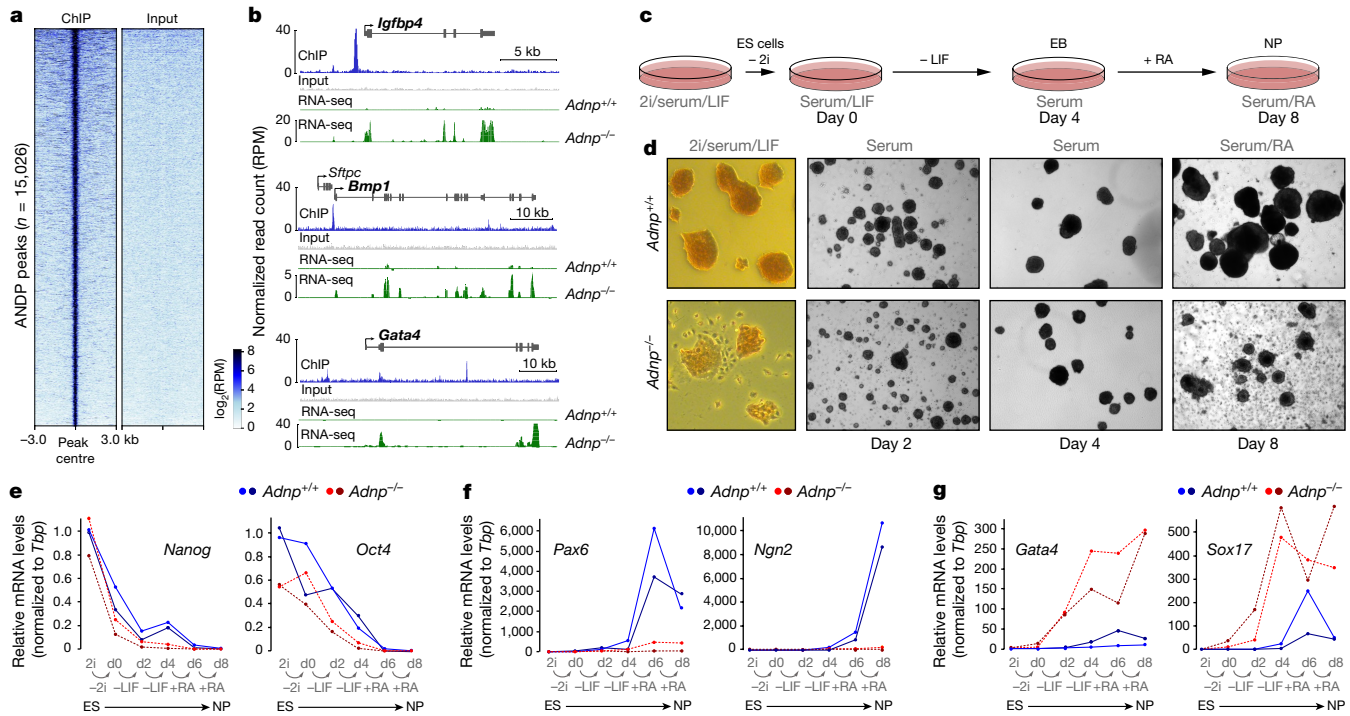



**Fig. 1 | ADNP binds and represses lineage-specifying genes. a**, Heat map of ADNP ChIP–seq enrichment across all significant peaks ($n = 15{,}026$) in the mouse genome. Each row represents a 6-kb window centred on peak midpoints, sorted by the ADNP ChIP signal. Input signals for the same windows are shown on the right. Average peak intensity of $n = 3$ biological replicates. RPM, reads per million. **b**, UCSC genome browser shots of three endoderm specification factors (*Igfbp4*, *Bmp1* and *Gata4*). ChIP–seq profiles for ADNP and input, and RNA-seq profiles for wild-type ($Adnp^{+/+}$) and ADNP-knockout ($Adnp^{-/-}$) mouse ES cells. Both ChIP–seq and RNA-seq profiles are normalized for library size. The experiment was repeated three times. **c**, Wild-type ES cells differentiate to neural progenitors (NP) in response to external cues. Consecutive withdrawal of 2i and leukaemia inhibitory factor (LIF) results in the formation of cellular aggregates (embryoid bodies, EB), which further differentiate into neural progenitors by the addition of retinoic acid (RA) at day 4. **d**, Phase-contrast images (original magnification, ×5) of $Adnp^{+/+}$ and $Adnp^{-/-}$ mouse ES cells stained with alkaline phosphatase when grown in 2i, serum and LIF (day 0), and during differentiation towards the neuronal lineage (days 2, 4 and 8). The experiment was repeated three times. **e**–**g**, mRNA expression profiles of genes specifying pluripotent cells (**e**; *Nanog* and *Oct4*), or cells of the neural (**f**; *Pax6* and *Ngn2*) or the endodermal (**g**; *Gata4* and *Sox17*) lineages from two independent differentiation experiments (light and dark coloured dots). Values normalized to *Tbp* mRNA are shown relative to the $Adnp^{+/+}$ parental cell line (in 2i/serum/LIF medium) for each replicate. Biological replicates were performed using independent mouse ES cell lines for each tagged protein. d, day.

(TAP–LC–MS/MS). Besides ADNP, we observed highly notable enrichment of HP1β, HP1γ and CHD4, but not SWI/SNF complex subunits. These interactions were preserved under 500 mM NaCl, showing that ADNP stably interacts with CHD4 and the HP1β and HP1γ proteins in ES cells (Fig. 2a). To corroborate this, we inserted a Flag-AviTag into the endogenous *Cbx1*, *Cbx3* and *Cbx5* genes, which encode the three mammalian HP1 isoforms HP1β, HP1γ and HP1α, respectively[12] (Extended Data Fig. 3a). Both ADNP and CHD4 were highly enriched in HP1β and HP1γ purifications (Fig. 2b and Extended Data Fig. 3). By contrast, CHD4 did not co-purify with HP1α (Extended Data Fig. 3b, e), and ADNP was 100-fold and 235-fold less abundant than in HP1β and HP1γ purifications, respectively (Extended Data Fig. 3g).

To verify that ADNP, HP1 and CHD4 form a stable complex via direct protein–protein interactions, we set out to reconstitute complex formation in vitro with recombinant human ADNP, HP1γ and CHD4 (Extended Data Fig. 4). Co-lysis of cells expressing HP1γ, ADNP and CHD4 resulted in the formation of a trimeric complex, which was preserved after streptavidin affinity purification, anion-exchange chromatography, and size-exclusion chromatography (SEC) (Fig. 2c, d). Subsequent experiments with full-length and truncated variants of the proteins (Extended Data Fig. 4) revealed that ADNP is at the core of the complex and interacts with CHD4 via its N terminus and with the chromoshadow domain (CSD) of HP1 via its C-terminal domain (Fig. 2e), probably through the PXVXL (in which X denotes any amino acid) motif[17]. In conclusion, CHD4, ADNP and HP1β/γ form a stable protein complex, which we refer to as ChAHP.

Next we performed ChIP–seq with endogenously tagged HP1α, HP1β and HP1γ and consulted a published dataset[20] for CHD4. Corroborating the biochemistry, all of the ADNP peaks ($n = 15{,}026$, Fig. 1) showed enrichment for CHD4 and HP1β/γ (Fig. 2f). Of the HP1 isoforms, the average HP1γ occupancy was the highest, HP1β was moderately enriched, and HP1α was barely detectable at those sites (Fig. 2f and Extended Data Fig. 5a–d). This confirms our TAP–LC–MS/MS results and indicates that HP1γ is the dominant isoform in ChAHP, whereas HP1β is present in a minor fraction of ChAHP complexes or forms sub-stoichiometric heterodimers with HP1γ. In line with a partial redundancy of HP1β and HP1γ, we observed that the average HP1β occupancy on all ChAHP-bound sites was greatly increased in the absence of HP1γ, whereas HP1γ occupancy remained similar in the absence of HP1β (Extended Data Fig. 5c). Thus, HP1γ is the predominant member of ChAHP in ES cells.

HP1 proteins recognize and bind to methylated H3K9 through the chromodomain[21,22], indicating that HP1 might target ChAHP to H3K9 methylated nucleosomes. Consistent with previous immunostaining experiments[17], we observed slight ADNP and CHD4 association with H3K9me3-marked chromatin. However, most of the highly enriched ADNP, and respective ChAHP peaks, were located in euchromatin (Fig. 2f). Consistent with repressive activity of ChAHP, histone modifications associated with active transcription were also absent (Extended Data Fig. 5e, f). In line with an H3K9me3-independent recruitment of ChAHP, HP1γ with mutations in the chromodomain that abolish H3K9me binding still bound to ChAHP target genes (Extended Data Fig. 5g). By contrast, the binding of HP1γ was lost at all ChAHP-bound sites in the absence of ADNP, whereas HP1γ bound to genomic regions with H3K9me3-modified nucleosomes remained largely unaffected in $Adnp^{-/-}$ cells (Fig. 3a). These results suggest that ADNP targets

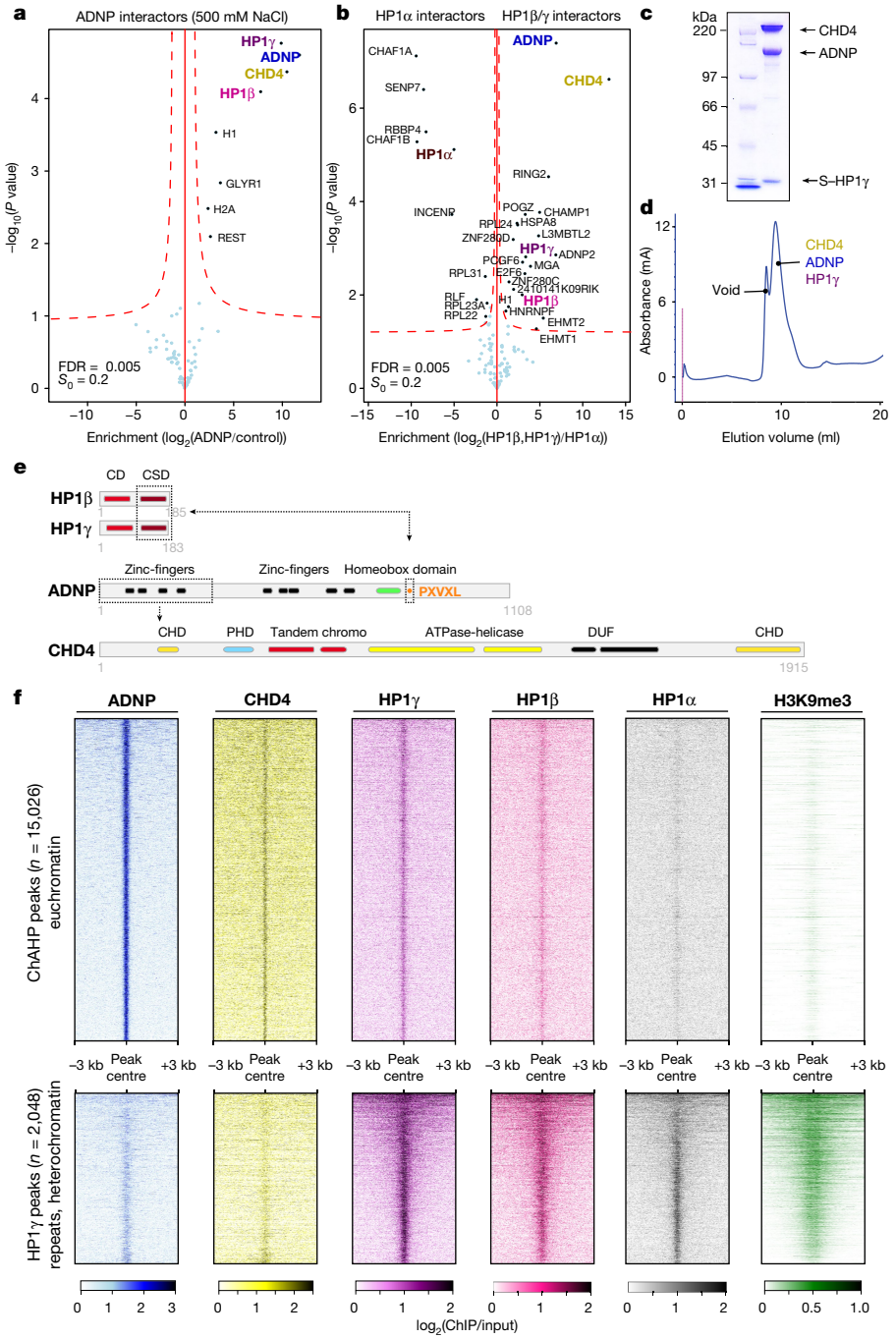



**Fig. 2 | ADNP mediates ChAHP complex formation. a**, TAP–LC–MS/MS of ADNP endogenously tagged with Flag-AviTag. Protein purification was performed in the presence of 500 mM NaCl. Parental mouse ES cell line serves as background control. $n = 3$ biological replicates. FDR, false discovery rate. **b**, TAP–LC–MS/MS of HP1α, HP1β and HP1γ endogenously tagged with Flag-AviTag. Protein purification was performed in the presence of 350 mM NaCl. Proteins that interact predominantly with HP1α (left), or HP1β or HP1γ (right) are indicated by UniProt names. $n = 3$ biological replicates. Statistical analysis was done with Perseus (see Methods). Mass spectrometry raw data are deposited with ProteomeXchange. **c**, **d**, In vitro reconstitution of the ChAHP complex. ADNP, CHD4 and HP1γ were expressed in Hi5 insect cells. Strep-tagged HP1γ (S–HP1γ) was pulled down with co-purifying ADNP and CHD4, followed by separation on size-exclusion chromatography (SEC) (Extended Data Fig. 4b, c). The fraction containing purified ChAHP was loaded on SDS–PAGE (**c**) and reinjected on SEC (**d**). For gel source data, see Supplementary Fig. 1. All experiments were performed at least twice. **e**, Scheme depicting ChAHP subunit interactions (see Extended Data Fig. 4). ADNP N-terminal zinc-fingers are necessary for the interaction with yet-to-be-determined CHD4 residues. The PXVXL motif in ADNP mediates the interaction with the CSD of HP1. Protein domains as predicted by InterPro. CD, chromodomain. **f**, Heat map of ADNP, CHD4, HP1γ, HP1β, HP1α and H3K9me3 ChIP–seq enrichment across all euchromatic sites bound by ChAHP (top) or heterochromatic sites bound by HP1γ (bottom). Each row represents a 6-kb window centred on the ADNP or HP1γ peak midpoint, respectively. Rows are sorted by ADNP (top) or HP1γ (bottom) ChIP enrichment. Average peak intensity of $n = 3$ biological replicates. Biological replicates were performed using independent mouse ES cell lines for each tagged protein.

the complex to euchromatic sites in a sequence-specific manner. Motif analysis of ADNP-bound loci revealed several significant DNA motifs. The highest-enriched motif (CGCCCYCTNSTG) was present in 63% of peaks ($P = 1 \times 10^{-10,538}$), and several motifs often co-occurred at bound genomic loci (Extended Data Fig. 6). To prove that ChAHP is indeed recruited via sequence-specific binding of ADNP, we deleted the predicted ADNP-binding motif (Δmotif) at the endogenous *Igfbp4* locus (Fig. 3b). Validating GCCCCCTGGAG as an ADNP-binding site, ADNP enrichment was specifically lost at the *Igfbp4* locus in $Igfbp4^{\Delta motif/\Delta motif}$ cells, whereas another ChAHP target gene (*Bmp1*) remained unaffected (Fig. 3c). Importantly, the binding of CHD4 and HP1γ was also depleted at the *Igfbp4* but not the *Bmp1* locus in $Igfbp4^{\Delta motif/\Delta motif}$ cells (Fig. 3d, e). Consistent with ChAHP-mediated target repression, we observed significantly increased *Igfbp4* but not *Bmp1* mRNA levels in $Igfbp4^{\Delta motif/\Delta motif}$ cells (Fig. 3f).

Identification of ChAHP strongly suggests that ADNP exerts its repressive function with the help of HP1. Indeed, $Cbx1^{-/-}Cbx3^{-/-}Cbx5^{-/-}$ triple-knockout cells revealed a distinct group of genes that was also upregulated in $Adnp^{-/-}$ cells. This was not evident in *Cbx* single- or double-knockout cells (Extended Data Fig. 7 and Supplementary Table 4). This suggests functional replacement by HP1α in the absence of HP1β and HP1γ, even though it only weakly interacts with ADNP and is not highly enriched at ChAHP target genes (Extended Data Figs. 3g, 5d). The fact that overall gene expression was not greatly affected if at least one HP1 isoform was present provides a general indication that HP1 isoforms can act partially redundantly to repress target genes in ES cells (Extended Data Fig. 7b).

The requirement of HP1 for ChAHP-mediated repression prompted us to revisit *ADNP* mutations found in patients with Helsmoortel–Van der Aa syndrome. Most are frameshift or nonsense mutations that result in C-terminally truncated ADNP that lacks the homeobox domain and the HP1 interaction motif[1] (Extended Data Fig. 1b). This suggests that mutant ADNP fails to assemble functional ChAHP and/or to bind its target genes. To test this, we introduced a patient-specific nonsense

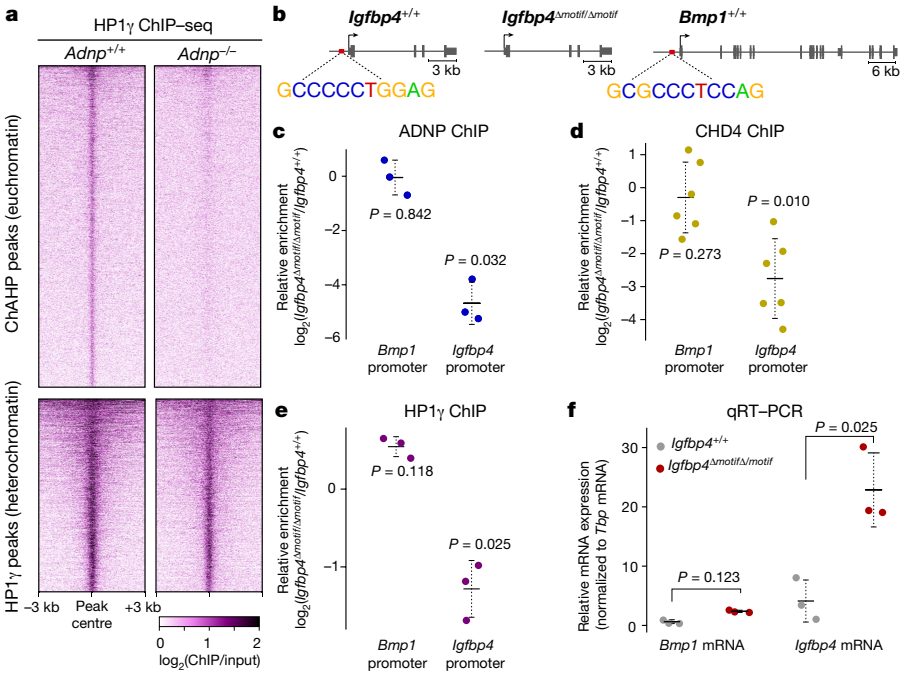



**Fig. 3 | DNA sequence specifies ChAHP association with euchromatin.** **a**, Heat map of HP1γ ChIP–seq enrichment in euchromatic ChAHP-bound sites (top) or heterochromatic HP1-bound sites (bottom) in $Adnp^{+/+}$ and $Adnp^{-/-}$ mouse ES cells. Each row represents a 6-kb window centred on the respective peak signal. Rows are sorted by mean ChIP enrichment. Average peak intensity of $n = 3$ biological replicates (that is, three independent ES cell lines). **b**, Schemes depicting the location of ADNP-binding motifs in the *Igfbp4* (left) and *Bmp1* (right) genes and the *Igfbp4* locus with the motif deletion $Igfbp4^{\Delta motif/\Delta motif}$ (middle). **c**, ChIP–qPCR measuring ADNP enrichments at *Igfbp4* and *Bmp1* promoters in $Igfbp4^{\Delta motif/\Delta motif}$ cells compared to the parental line ($Igfbp4^{+/+}$). $n = 3$ biological replicates. **d**, ChIP–qPCR measuring CHD4 enrichments at *Igfbp4* and *Bmp1* promoters in $Igfbp4^{\Delta motif/\Delta motif}$ cells compared to the parental line. $n = 6$ biological replicates. **e**, ChIP–qPCR measuring HP1γ enrichments at *Igfbp4* and *Bmp1* promoters in $Igfbp4^{\Delta motif/\Delta motif}$ cells compared to the parental line. $n = 3$ biological replicates. **f**, qRT–PCR measurement of *Igfbp4* and *Bmp1* mRNA levels in wild-type and $Igfbp4^{\Delta motif/\Delta motif}$ cells. $n = 3$ biological replicates. *P* values in **c**–**f** were calculated using two-tailed unpaired unequal variances *t*-tests. Centre values denote the mean; error bars denote s.d.

mutation upstream of the homeobox domain at amino acid position 718 ($Adnp^{PTC718}$; Extended Data Fig. 8a; corresponds to Tyr719 in human ADNP). $ADNP^{PTC718}$ failed to co-purify with HP1β and HP1γ, whereas the interaction with CHD4 remained preserved (Extended Data Fig. 8b). Consistent with the requirement of HP1 for silencing, we observed increased expression of ADNP-target genes in cells that express $ADNP^{PTC718}$ (Extended Data Fig. 8c). The $ADNP^{PTC718}$ protein still bound its target site at the *Igfbp4* locus (Extended Data Fig. 8d), indicating that the homeobox domain is dispensable for DNA binding but might assist in target repression. These results demonstrate that patients with nonsense mutations in the *ADNP* gene cannot assemble fully functional ChAHP complexes. To test whether this could potentially be restored pharmacologically, we treated $ADNP^{PTC718}$-expressing cells with gentamycin or paromomycin, two aminoglycoside antibiotics that promote translational read-through[5]. Indeed, gentamycin treatment promoted read-through of PTC718 (Extended Data Fig. 8e, f) and rescued the interaction between $ADNP^{PTC718}$ and HP1β or HP1γ. Although less effectively, HP1γ was also retrieved from samples that were treated with paromomycin (Extended Data Fig. 8b). Thus, a therapeutic approach that promotes ribosomal read-through of premature stop codons could be of considerable medical benefit. However, the discovery of new nonsense suppressors will be inevitable, because the clinical utility of aminoglycoside therapy is limited by low efficacy and serious toxicities[6].

Finally, we set out to investigate the molecular function of ChAHP. Inspired by the role of CHD4 in nucleosome remodelling and HP1 in heterochromatin assembly, we investigated a possible role of ChAHP in regulating local chromatin accessibility by the ATAC-seq method (assay for transposase-accessible chromatin using sequencing)[23]. Many transcription factors, such as NRF1[24], generate local accessible regions at their DNA-binding sites (Fig. 4a). Unexpectedly, we did not observe such footprints for ADNP. Instead, chromatin across ChAHP-bound

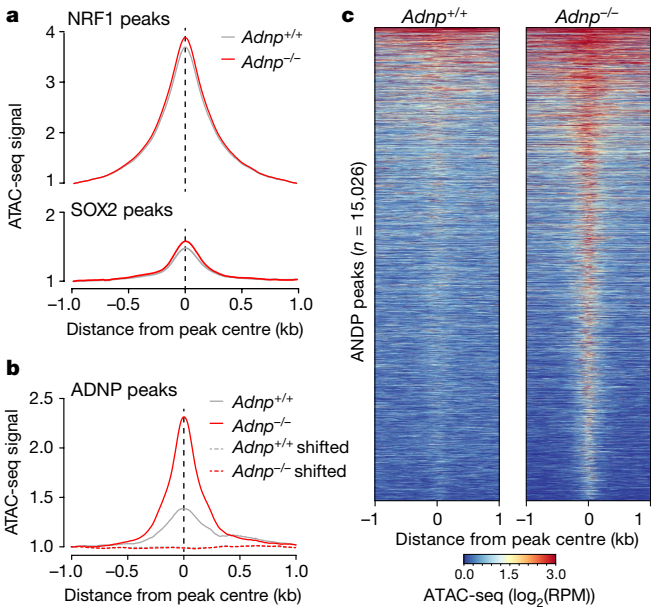



**Fig. 4 | ChAHP obstructs chromatin accessibility.** **a**, Average accessibility of loci bound by unrelated transcription factors NRF1 (top) and SOX2 (bottom) in $Adnp^{+/+}$ (grey) and $Adnp^{-/-}$ (red) mouse ES cell lines measured by ATAC-seq. Profiles represent averaged biological replicates ($n = 4$). **b**, Average accessibility of loci bound by ADNP and 'random' control loci (ADNP peaks shifted by 10 kb; dashed lines) in $Adnp^{+/+}$ (grey) and $Adnp^{-/-}$ (red) mouse ES cells. Profiles represent averaged biological replicates ($n = 4$). **c**, Heat map showing ATAC-seq read coverage in a 2-kb window around all ADNP peaks normalized by library depth in $Adnp^{+/+}$ (left) and $Adnp^{-/-}$ (right) ES cells.

loci was largely devoid of an ATAC-seq signal (Fig. 4b, c). This suggests that ChAHP is bound to chromatin with impaired accessibility, or conversely, that the binding of ChAHP renders chromatin inaccessible. Notably, all ChAHP-bound sites became readily accessible in the absence of ADNP, whereas ChAHP-independent control loci such as NRF1- or SOX2-binding sites, as well as a 'random' set of genomic loci (ADNP peaks shifted by 10 kb) showed no difference in accessibility (Fig. 4a–c). Notably, the opening of chromatin in $Adnp^{-/-}$ cells was restricted to a few hundred base pairs around ChAHP-binding sites, and the surrounding regions remained inaccessible (Fig. 4c). Thus, rather than assembling broad, inaccessible domains of chromatin, ChAHP denies direct access to its cognate DNA-binding sites.

In summary, we have discovered ChAHP, a gene-regulatory complex that consists of the chromatin remodeller CHD4, the DNA-binding factor ADNP, and the heterochromatin proteins HP1β and HP1γ. By locally restricting access to DNA, ChAHP prevents endodermal gene transcription in mouse ES cells and during neuroectodermal differentiation. This stabilizes cellular states and ensures correct lineage specification. Although ChAHP could directly interfere with transcribing RNA polymerase, we favour a model in which ChAHP prevents the binding of other regulatory factors, such as transcriptional activators, to DNA. Although the exact mode of action of ChAHP remains to be determined, such a model would be consistent with the notion that ChAHP also binds outside gene bodies and promoters.

## Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0153-8.

Received: 6 September 2017; Accepted: 13 April 2018;
Published online: 23 May 2018

1. Helsmoortel, C. et al. A SWI/SNF-related autism syndrome caused by de novo mutations in ADNP. *Nat. Genet.* **46**, 380–384 (2014).
2. Pinhasov, A. et al. Activity-dependent neuroprotective protein: a novel gene essential for brain formation. *Brain Res. Dev. Brain Res.* **144**, 83–90 (2003).
3. Gozes, I., Yeheskel, A. & Pasmanik-Chor, M. Activity-dependent neuroprotective protein (ADNP): a case study for highly conserved chordata-specific genes shaping the brain and mutated in cancer. *J. Alzheimers Dis.* **45**, 57–73 (2015).
4. Gozes, I. et al. The compassionate side of neuroscience: Tony Sermone's undiagnosed genetic journey–ADNP mutation. *J. Mol. Neurosci.* **56**, 751–757 (2015).
5. Hermann, T. Aminoglycoside antibiotics: old drugs and new therapeutic approaches. *Cell. Mol. Life Sci.* **64**, 1841–1852 (2007).
6. Peltz, S. W., Morsy, M., Welch, E. M. & Jacobson, A. Ataluren as an agent for therapeutic nonsense suppression. *Annu. Rev. Med.* **64**, 407–425 (2013).
7. Welch, E. M. et al. PTC124 targets genetic disorders caused by nonsense mutations. *Nature* **447**, 87–91 (2007).
8. Zamostiano, R. et al. Cloning and characterization of the human activity-dependent neuroprotective protein. *J. Biol. Chem.* **276**, 708–714 (2001).
9. Bassan, M. et al. Complete sequence of a novel protein containing a femtomolar-activity-dependent neuroprotective peptide. *J. Neurochem.* **72**, 1283–1293 (1999).
10. Mandel, S., Rechavi, G. & Gozes, I. Activity-dependent neuroprotective protein (ADNP) differentially interacts with chromatin to regulate genes essential for embryogenesis. *Dev. Biol.* **303**, 814–824 (2007).
11. Niwa, H. Mouse ES cell culture system as a model of development. *Dev. Growth Differ.* **52**, 275–283 (2010).
12. Flemr, M. & Bühler, M. Single-step generation of conditional knockout mouse embryonic stem cells. *Cell Reports* **12**, 709–716 (2015).
13. Molkentin, J. D. The zinc finger-containing transcription factors GATA-4, -5, and -6. Ubiquitously expressed regulators of tissue-specific gene expression. *J. Biol. Chem.* **275**, 38949–38952 (2000).
14. Fujikura, J. et al. Differentiation of embryonic stem cells is induced by GATA factors. *Genes Dev.* **16**, 784–789 (2002).
15. Cho, L. T. Y. et al. Conversion from mouse embryonic to extra-embryonic endoderm stem cells reveals distinct differentiation capacities of pluripotent stem cell states. *Development* **139**, 2866–2877 (2012).
16. Bibel, M. et al. Differentiation of mouse embryonic stem cells into a defined neuronal lineage. *Nat. Neurosci.* **7**, 1003–1009 (2004).
17. Mosch, K., Franz, H., Soeroes, S., Singh, P. B. & Fischle, W. HP1 recruits activity-dependent neuroprotective protein to H3K9me3 marked pericentromeric heterochromatin for silencing of major satellite repeats. *PLoS ONE* **6**, e15894 (2011).
18. Mandel, S. & Gozes, I. Activity-dependent neuroprotective protein constitutes a novel element in the SWI/SNF chromatin remodeling complex. *J. Biol. Chem.* **282**, 34448–34456 (2007).
19. Vermeulen, M. & et al. Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. *Cell* **142**, 967–980 (2010).
20. de Dieuleveult, M. et al. Genome-wide nucleosome specificity and function of chromatin remodellers in ES cells. *Nature* **530**, 113–116 (2016).
21. Bannister, A. J. et al. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410**, 120–124 (2001).
22. Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. & Jenuwein, T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410**, 116–120 (2001).
23. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
24. Domcke, S. et al. Competition between DNA methylation and transcription factors determines binding of NRF1. *Nature* **528**, 575–579 (2015).

**Acknowledgements** We thank A. Peters and D. Schübeler for discussion and feedback on the manuscript, and H. Pickersgill for editing assistance. We thank Y. Shimada and N. Laschet for technical support. We would like to thank the FMI Functional Genomics facility for assistance in library construction and next generation sequencing, and J. Seebacher for discussions. This work was supported by funds from the Swiss National Science Foundation (SNF).

**Reviewer information** *Nature* thanks P. Wade and the other anonymous reviewer(s) for their contribution to the peer review of this work.

**Author contributions** V.O. designed and performed experiments, analysed data, generated cell lines and prepared figures. M.F. generated cell lines and advised on experimental design. D.H. and V.I. acquired and analysed mass spectrometry data. Bioinformatic and computational analysis was performed by S.H.C. and F.M. A.B. performed in vitro biochemistry experiments. V.O. and L.L. performed immunoprecipitation and mass spectrometry with patient-specific *Adnp* mutants, and L.L. performed ChIP experiments for the patient-specific *Adnp* mutants. A.P. performed ChIP experiments for the *Cbx3* chromodomain mutant. F.M. generated and analysed ATAC-seq data. N.H.T. and J.B. advised on the experimental design. M.B. conceived and supervised the study, and secured funding. M.B. and F.M. wrote the manuscript. All authors discussed the results and commented on the manuscript.

**Competing interests** The authors declare competing financial interests: a patent application has been filed (EP17191642.2). The Friedrich Miescher Institute for Biomedical Research (FMI) receives significant financial contributions from the Novartis Research Foundation. Published research reagents from the FMI are shared with the academic community under a Material Transfer Agreement (MTA) having terms and conditions corresponding to those of the UBMTA (Uniform Biological Material Transfer Agreement).

**Additional information**
**Extended data** is available for this paper at https://doi.org/10.1038/s41586-018-0153-8.
**Supplementary information** is available for this paper at https://doi.org/10.1038/s41586-018-0153-8.
**Reprints and permissions information** is available at http://www.nature.com/reprints.
**Correspondence and requests for materials** should be addressed to M.B.
**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized, and investigators were not blinded to allocation during experiments and outcome assessment.

**Cell culture and genome editing.** Mouse ES cells (129×C57BL/6[25] were cultured on gelatin-coated dishes in ES medium containing DMEM (Gibco 21969-035), supplemented with 15% fetal bovine serum (FBS; Gibco), 1× non-essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), 2 mM L-glutamine (Gibco), 0.1 mM 2-mercaptoethanol (Sigma), 50 mg ml$^{-1}$ penicillin, 80 mg ml$^{-1}$ streptomycin, 3 μM glycogen synthase kinase (GSK) inhibitor (Calbiochem, D00163483), 10 μM MEK inhibitor (Tocris, PD0325901), and homemade LIF, at 37 °C in 5% $CO_2$.

The genome editing was performed as previously published[12] in the absence of GSK and MEK inhibitors in the above-described ES medium.

**Generation of endogenously tagged ES cell lines.** For endogenous gene tagging using TALENs, Rosa26:BirA-V5-expressing cells (cMB053 or cMB063) were transfected with 400 ng TALEN-EED, 400 ng TALEN-KKR, 100 ng pRRP reporter and 1,000 ng of donor single-stranded oligodeoxynucleotide encoding the tag sequence. The single-stranded oligodeoxynucleotides were synthesized as Ultramers by Integrated DNA Technologies and their sequences are listed along with TALEN sequences in Supplementary Table 6. All transfections were carried out using Lipofectamine 3000 reagent (Invitrogen) at a 3 μl:1 μg DNA ratio in OptiMEM medium (Invitrogen). Transfected cells were selected by adding puromycin (2 μg ml$^{-1}$) to the ES medium 24 h after transfection. After 36 h of selection, surviving cells were sparsely seeded for clonal expansion. The resulting clones were individually picked, split and screened by western blot for desired tag integration. See Supplementary Information for the list of tagged cell lines generated in this study.

**Straight knockout ES cell line generation.** *Cbx1*$^{-/-}$ mouse ES cells were generated using TALENs that target the first and last coding exon, resulting in a deletion of approximately 6,000 bp (exon 2–exon 6).

*Adnp*$^{-/-}$ mouse ES cells were generated using *Cas9* and TALENs that target the first and last coding exon, resulting in a deletion of approximately 7,000 bp (exon 2–exon 4). The *Cas9* sgRNA sequence was cloned into the SpCas9-2A-mCherry plasmid[26]. Sequences of TALENs and *Cas9* gRNA can be found in Supplementary Table 6. See Supplementary Information for the list of straight knockout cell lines generated in this study.

**Conditional ES cell line generation.** The *Cbx3*$^{fl/fl}$ cell line was generated as described[12]. For the *Cbx5*$^{fl/fl}$ conditional cell line, a mouse ES cell line containing an integration of the CreERT2 recombinase fusion in the *Rosa26* locus (cMB052 or cMB063) was transfected with TALENs cutting before and after the third exon. Single-stranded oligodeoxynucleotides with corresponding homology arms and *loxP* sites for integration were also included in the transfection mix (see Supplementary Table 6 for sequences). Clones were screened for homozygous integrations for both *loxP* sites. A cell line with both bi-allelic *loxP* integrations was tested for recombination efficiency by treating the cells with 0.1 μM 4-hydroxytamoxifen (4-OHT; Sigma) followed by western blot or quantitative PCR with reverse transcription (qRT–PCR).

**Transient expression experiments in ES cells.** The full-length *Adnp* cDNA was cloned into the mammalian expression vector pEFaFB (promoter of elongation factor-1 alpha; ATG-3×Flag-Avi-GOI-2A-puromycin), creating pEFaFB-Adnp, which was then used as a template to mutate codon 718 (TAT to TAA) using the QuikChange Lightning Site-Directed Mutagenesis (SDM) Kit (Agilent), with the final construct encoding pEFaFB-ADNP$^{PTC718}$. Alternatively, *Adnp*$^{PTC718}$ cDNA was cloned into the pEFaCFB vector (C-terminal 3×Flag-AviTag).

The *Cbx3* cDNA was cloned into pEFaFB, and then used as a template for SDM mutating chromodomain residues Trp43 and Phe46 to Ala (*Cbx3* CDmut)[27].

For ChIP experiments, 5 × 10$^6$ *Adnp*$^{-/-}$ (cMB377) cells in a 10-cm dish were reverse transfected with 10 μg pEFaFB-Adnp or pEFaFB-AdnpPTC718. Alternatively, *Cbx1*$^{-/-}$*Cbx3*$^{-/-}$*Cbx5*$^{-/-}$ triple-knockout (4-OHT-treated cMB282) cells were transfected with 10 μg pEFaFB-Cbx3 or pEFaFB-Cbx3-CDmut. Cells were collected 48 h after transfection, and further processed according to the ChIP protocol (below). For affinity purification experiments, 6 × 10$^6$ cells in a 15-cm dish were reverse transfected with 10 μg pEFaFB-AdnpPTC718 or pEFaCFB-AdnpPTC718. Twenty hours after transfection, cells were forward-transfected with 5 μg pEFaFB-AdnpPTC718. Finally, 48 h from the first transfection, cells were treated with 2 mg ml$^{-1}$ gentamycin (Sigma, G1914) or paromomycin (Sigma, P9297) for 24 h. Cells were then collected and processed according to the Affinity purification protocol (bellow). Transfections were carried out using Lipofectamine 3000 reagent (Invitrogen) at a 3 μl:1 μg DNA ratio in OptiMEM medium (Invitrogen), and 10 μM ROCK inhibitor (Tocris, Y27632) for increased cell survival.

**Differentiation of ES cells to neuronal precursors.** The cMB263 (*Adnp*$^{+/+}$) and cMB267 (*Adnp*$^{-/-}$) ES cell lines were differentiated as previously described[28], except that no feeder cells were used. Instead, cells were grown in ES cell medium containing 2i, as described above.

**Western blotting.** Cells were grown to confluency on 6-well plates, collected in PBS, pelleted by 2 min centrifugation at 400*g*, and pellets were then resuspended in 100 μl protein extraction buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1% Triton X-100, 0.5 mM EDTA, and 5% glycerol) supplemented with protease inhibitor cocktail (PIC; Roche), 1 mM PMSF, and 1 mM dithiothreitol (DTT). Proteins were extracted for 30 min on ice, the lysates were centrifuged at 16,000*g* for 20 min at 4 °C, and the protein concentration in the supernatant was determined using the BioRad protein assay. For western blotting, 20 μg of protein was resolved on NuPAGE-Novex Bis-Tris 4–12% gradient gels (Invitrogen), which were semi-dry transferred on polyvinylidene fluoride (PVDF) membrane, blocked for 30 min in 2.5% non-fat dry milk in TBS plus 0.05% Tween 20 (TBST), and stained with primary antibodies at 4 °C overnight. The primary antibodies used for western blotting were mouse anti-Flag (1:1,000, Sigma clone M2), goat-anti-HP1α (1:1,000, Abcam, ab77256), mouse-anti-HP1α (1:1,000, Millipore, mab3446), rat-anti-HP1β (1:500, Serotec, MCA1946), mouse-anti-HP1γ (1:2,000, Cell Signaling Technology), mouse-anti-CHD4 (1:1,000, Abcam, ab70469), rabbit-anti-MTA2 (1:1,000, Bethyl, A300-395A-T), rabbit-anti-GATAD2B (1:1,000, Bethyl, A301-283A-T), rabbit-anti-MBD3 (1:1,000, Bethyl, A302-528A-T) and rat-anti-tubulin (1:5,000, Abcam clone YL1/2). Signal was detected with corresponding horseradish peroxidase (HRP)-conjugated secondary antibodies and Immobilon Western Chemiluminiscent HRP Substrate (Millipore). For streptavidin staining, membranes were blocked after transfer in 2% bovine serum albumin (BSA) in TBST and incubated with streptavidin–HRP (1:20,000, Sigma) for 30 min at room temperature, followed by signal development as above.

**ChIP.** A confluent 10-cm culture dish of ES cells (approximately 2 × 10$^7$ cells) were cross-linked for 7 min at room temperature, with 1% final formaldehyde solution (Sigma, F8775) added directly to the ES medium. Cross-linking was quenched by the addition of glycine to a final concentration of 0.125 mM and incubation at 4 °C for 10 min; cells were then washed twice with PBS. Cells were collected in 1 ml PBS with PIC (Roche) and spun at 600*g* for 5 min at 4 °C. Cells were then resuspended in 5 ml wash solution I (10 mM Tris pH 8, 10 mM EDTA, 0.5 mM EGTA, 0.25% Triton X-100), incubated for 10 min on ice, then spun at 1,200*g* for 5 min at 4 °C. The remaining nuclear pellet was then resuspended in 5 ml wash solution II (10 mM Tris pH 8, 1 mM EDTA, 0.5 mM EGTA and 200 mM NaCl) and incubated for 5 min on ice, then spun at 1,200*g* for 5 min at 4 °C. Cell pellet was subsequently washed in 900 μl sonication buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA and 0.1% SDS) without disturbing the pellet, and finally resuspended in the sonication buffer supplemented with PIC. Chromatin was then sonicated in Covaris 1-ml tubes for 15 min with the following settings: duty cycle: 5%, peak incident power: 140 W, cycles per burst: 200, temperature (bath): 4 °C.

**Beads preparation.** For Bio-ChIP (ChIP for proteins tagged with the Flag-Avi tag), 40 μl Dynabeads Stepavidin (Thermo Fisher) per sample, or alternatively 40 μl Protein-G Dynabeads (Thermo Fisher) per sample for ChIP with protein-specific antibodies (Ab-ChIP), were washed twice for 5 min in 0.5 ml blocking buffer (PBS, 0.5% Tween and 0.5% BSA). Streptavidin Dynabeads were then washed twice with immunoprecipitaiton buffer (50 mM HEPES pH 7.5, 150 mM NaCl, 1 mM EDTA, 0.1% SDS, 0.1% sodium deoxycholate and 1% Triton X-100) and stored on ice. Protein-G Dynabeads were incubated for 1 h at room temperature in blocking buffer with the desired antibody. Beads were then washed twice in blocking buffer and stored on ice. For CHD4 ChIP, 10 μg mouse anti-CHD4 (Abcam, ab70469, 3F2/4) conjugated to Protein G was used.

**Immunoprecipitation and washes.** For immunoprecipitation analyses, 10 μl (1%) was kept as the input sample, and 40 μl pre-blocked Dynabeads were added to 1 ml of sonicated chromatin in immunoprecipitation buffer and incubated overnight at 4 °C on a rotating wheel. Beads were collected on a magnetic rack for 2–3 min to remove supernatant between each step, and washed as follows: for Bio-ChIP, twice for 10 min with 2% SDS in TE buffer (10 mM Tris, pH 8, 1 mM EDTA), once for 10 min with high salt buffer (50 mM HEPES pH 7.5, 1 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate and 500 mM NaCl), once for 10 min with DOC buffer (250 mM LiCl, 0.5%, NP-40, 0.5% deoxycholate, 1 mM EDTA and 10 mM Tris pH 8) and twice for 10 min with 1 ml TE buffer. For Ab-ChIP, beads were washed five times with immunoprecipitation buffer, twice with DOC buffer, and twice with TE buffer. Beads were then resuspended in 300 μl elution buffer (1% SDS and 100 mM $NaHCO_3$) and 6 μl RNaseA (10 mg ml$^{-1}$ stock) and incubated at 37 °C for 30 min while mixing. Elution buffer was adjusted with 6 μl 0.5 M EDTA, 12 μl 1 M Tris pH 8 and 2.5 μl Proteinase K (10 mg ml$^{-1}$, Roche). Beads were incubated for 3 h at 55 °C and then overnight at 65 °C with mixing to de-crosslink. The same procedure was followed for input samples including RNase and proteinase K digestion. DNA was purified using AMPure XP beads (Beckman Coulter). Quantification was performed with Qubit dsDNA high-sensitivity assay (Thermo Fisher).

**ChIP–qPCR and ChIP–seq.** DNA was subjected to qPCR analysis (as described for qRT–PCR, below) using ChIP primers described in Supplementary Information. For ChIP–seq sample preparation, library construction was performed using the NEBNext Ultra kit (New England Biolabs) following

manufacturer recommendations. Libraries were sequenced on Illumina HiSeq 2500 machines, with 50-bp single-end sequencing.

**qRT–PCR and RNA-seq.** For qRT–PCR experiments, total RNA was extracted from ES cells with the Absolutely RNA Microprep Kit (Stratagene). Total RNA (500 ng) was reverse transcribed with the Primescript RT kit (Clontech). qRT–PCR was performed on a CFX96 Real-Time PCR System (Bio-Rad) using the SsoAdvanced SYBR Green Supermix (Bio-Rad, 172–5264). Relative RNA levels were calculated from $C_t$ values according to the $\Delta C_t$ method and normalized to *Tbp* mRNA levels where applicable. For RNA-seq, total RNA (isolated as above) was subjected to ribosomal RNA depletion using the Ribozero kit (Illumina) followed by library construction using the ScriptSeq V2 library preparation kit (Illumina).

**ATAC-seq.** ATAC-seq was performed following a previously published protocol[23] using 50,000 $Adnp^{+/+}$ or $Adnp^{-/-}$ mouse ES cells. The experiment was performed in biological replicates using two independent isogenic cell lines for each genotype. Libraries were paired-end sequenced (2 × 75 bp) using an Illumina NextSeq 500 device.

**Affinity purification for LC–MS/MS.** All affinity purifications in this work were performed according to the following protocol with the exception of $ADNP^{PTC718}$ affinity purifications (see below). Cells were grown to confluency on 10-cm dishes, collected in PBS, and pelleted by centrifugation at 400*g* for 2 min. All subsequent steps were performed on ice or at 4 °C. Pellets were resuspended in 3 ml of nuclear extract buffer 1 (NEB1; 20 mM HEPES, 10 mM KCl, 1 mM EDTA, 0.1 mM $Na_3VO_4$, 0.2% NP-40, 10% glycerol, 1 mM DTT and 1× PIC) followed by centrifugation at 1,000*g* for 3 min. Pellets were resuspended in 1 ml NEB1 buffer and incubated on ice for 10 min, followed by dounce homogenization. Isolated nuclei were collected by centrifugation at 1,000*g* for 15 min, and carefully washed twice with 1 ml NEB1 without disturbing the pellet. Pellets were then resuspended in 0.5 ml of nuclear extract buffer 2 (NEB2; 20 mM HEPES, 10 mM KCl, 1 mM EDTA, 0.1 mM $Na_3VO_4$, 350 mM NaCl, 20% glycerol, 1 mM DTT and 1× PIC), dounce homogenized (20× up and down), incubated for 30 min, and finally spun at 16,000*g* for 30 min. Protein concentration was determined using the Bradford assay, and approximately 250 μg of nuclear extract was used per affinity purification. The protein lysates were adjusted to affinity purification buffer (350 mM or 500 mM NaCl, 20 mM Tris-HCl, pH 7.5, 0.3% NP-40, 1 mM EDTA, 10% glycerol, 1 mM DTT and 1× PIC), added to 20 μl anti-Flag-M2 Dynabeads (Sigma), and incubated overnight rotating at 4 °C. Dynabeads were washed the next day in affinity purification buffer (4 × 10 min), followed by 3 × 15-min elutions of bound proteins with 3×Flag peptide (final concentration 0.3 mg $ml^{-1}$ in affinity purification buffer, Sigma). Next, elutions were pooled and added to the washed Stepavidin Dynabeads (Thermo Fisher), and incubated overnight rotating at 4 °C. Streptavidin Dynabeads were washed the next day with affinity purification buffer (4 × 10 min), followed by a wash with affinity purification buffer without NP40. For single-step affinity purification, Flag purification was omitted, and lysates were directly applied to the Streptavidin Dynabeads. The enriched proteins were digested directly on the Dynabeads with 0.1 mg $ml^{-1}$ trypsin in digestion buffer (50 mM Tris pH 8.0, 1 mM $CaCl_2$ and 1 mM TCEP).

For $ADNP^{PTC718}$ affinity purification, cells from 2 × 15-cm dishes per replicate were used; collection and nuclear lysate isolation were as described above. Next, 400 μg of nuclear lysates were used for single-step purification with Streptavidin Dynabeads in affinity purification buffer (2 h incubation at 4 °C), followed by washes (see above). The enriched proteins were digested directly on the Dynabeads with 0.2 μg Lys-C in 5 μl digestion buffer (3 M guanidium chloride, 20 mM EPPS, pH 8.5, 10 mM CAA and 5 mM TCEP) for 2 h at room temperature. Next, samples were diluted with 50 mM HEPES, pH 8.5, and digested with 0.2 μg trypsin overnight at 37 °C. The next day, 0.2 μg fresh trypsin was added, and samples were incubated for an additional 5 h at 37 °C.

**Mass spectrometry.** *Analysis of affinity purification.* The generated peptides (see 'Affinity purification for LC–MS/MS') were acidified with TFA to a final concentration of 0.8% and analysed by LC–MS/MS with an EASY-nLC 1000 using the two-column set-up (Thermo Scientific). The peptides were loaded with 0.1% formic acid, 2% acetonitrile in $H_2O$ onto a peptide trap (Acclaim PepMap 100, 75 μm × 2 cm, C18, 3 μm, 100 Å) at a constant pressure of 80 MPa. Peptides were separated, at a flow rate of 150 nl $min^{-1}$ with a linear gradient of 2–6% buffer B in buffer A in 3 min followed by an linear increase from 6 to 22% in 40 min, 22–28% in 9 min, 28–36% in 8 min, 36–80% in 1 min and the column was finally washed for 14 min at 80% buffer B in buffer A (buffer A: 0.1% formic acid; buffer B: 0.1% formic acid in acetonitrile) on a 50 μm × 15 cm ES801 C18, 2 μm, 100 Å column (Thermo Scientific) mounted on a DPV ion source (New Objective) connected to a Orbitrap Fusion (Thermo Scientific). The data were acquired using 120,000 resolution for the peptide measurements in the Orbitrap and a top T (3 s) method with HCD fragmentation for each precursor and fragment measurement in the ion trap according the recommendation of the manufacturer (Thermo Scientific).

Protein identification and relative quantification of the proteins was done with MaxQuant version 1.5.3.8 using Andromeda as search engine[29] and label-free quantification (LFQ[30]) as described previously[31]. The mouse subset of the UniProt version 2015_01 combined with the contaminant DB from MaxQuant was searched and the protein and peptide FDR values were set to 0.01. All MaxQuant parameters can be found in the uploaded parameterfile: mqpar.xml (deposited in the PRIDE repository, see Data availability).

Statistical analysis was done in Perseus (version 1.5.2.6)[29,30,32]. Results were filtered to remove reverse hits, contaminants and peptides found in only one sample. Missing values were imputed and potential interactors were determined using *t*-test and visualized by a volcano plot. Significance lines corresponding to a given FDR have been determined by a permutation-based method[33]. Threshold values (FDR) were selected between 0.005 and 0.05 and $S_0$ (curve bend) between 0.2 and 2.0 and are shown in the corresponding figures. Results were exported from Perseus and visualized using statistical computing language R.

*iBAQ.* Intensity based absolute quantification (iBAQ) was done as described previously[34] to evaluate protein abundances in the ChAHP complexes of the different pull-down reactions.

*PRM data acquisition.* Parallel reaction monitoring (PRM) analyses were performed using the same LC–MS system and gradient as described above. The acquisition method consisted of acquiring one MS spectrum at 120,000 resolution from 375 to 1,575 Da followed by 21 PRM spectra. An isolation window of 1.6 Da, a resolution of 240,000, and an automatic gain control value of $10^5$ was used. Fragmentation was performed with a stepped HCD collision energy of 30 ± 5, and MS/MS scans were acquired with a scan range from 110 to 1,800.

*PRM data analysis.* The acquired PRM data were processed using Skyline 4.1[35]. The transition selection was systematically verified and adjusted when necessary to ensure that no co-eluting contaminant distorted quantification based on traces co-elution (retention time) and the correlation between the relative intensities of the endogenous fragment ion traces, and their counterparts from the library.

MASCOT 2.5 was used in the Decoy mode to search the Swissprot mouse version 2015_01 including common contaminants. The enzyme specificity was set to trypsin allowing for up to three incomplete cleavage sites. Carbamidomethylation of cysteine (+57.0245) was set as a fixed modification, oxidation of methionine (+15.9949 Da) and acetylation of the protein N terminus (+42.0106 Da) were set as variable modifications. Parent ion mass tolerance was set to 10 p.p.m. and fragment ion mass tolerance to 0.6 Da. The results were validated with the program Scaffold Version 4.4 (Proteome Software). Protein identifications were accepted if they could be established at greater than 0.1% FDR rate as calculated in Scaffold.

**SEC of nuclear lysates.** Nuclear lysates were isolated as described above ('Affinity purification for LC–MS/MS' section) from 3 × 10-cm dishes of $Adnp^{Flag\text{-}AviTag/Flag\text{-}AviTag}$ ES cells (cMB264). Nuclear lysates were then concentrated to 250 μl final volume using Amicon Ultra 0.5 ml Centrifugical Filters (3 kDa, Millipore), and fractionated by SEC on a Superose 6 HR 10/300 resin by fast protein liquid chromatography (AKTA; Amersham-Pharmacia Biotech). The predicted size exclusion maximum for this resin is 40 MDa, with a void volume of 7.35 ml. The column was equilibrated in 2 column volumes of gel filtration (GF) buffer (250 mM NaCl, 50 mM Tris-HCl pH 7.5, 1 mM DTT, 1× PIC) before sample loading. A high-molecular-mass protein column standard was used to define the column resolution (Sigma). Protein peaks were detected by UV monitoring. Thyroglobulin (669,000 Da) peaked in fractions 9 and 10. Before loading, each nuclear lysate was adjusted to the appropriate column conditions and centrifuged at 100,000*g* for 30 min. A 200 μl of lysate was loaded onto the column and collected into 350-μl fractions; fractions were then subjected to trichloroacetic acid (TCA) precipitation for western blot analysis. For TCA precipitation, the sample volume was adjusted to 500 μl with the GF buffer followed by the addition of 50 μl 0.15% sodium deoxycholate; tubes were vortexed and incubated at room temperature for 10 min. Protein was precipitated by the addition of 25 μl of 100% TCA (Sigma), followed by a 20-min incubation at −20 °C. Precipitated proteins were collected by centrifugation at 10,000 *g* for 10 min at 4 °C. Protein pellets were washed with acetone and air-dried. The protein pellet was solubilized in 1× sample buffer (62.5 mM Tris, pH 6.8, 0.72 M β-mercaptoethanol or 0.1 M DTT, 10% glycerol, 2% SDS and 0.05% bromophenol blue) and resolved by NuPAGE-Novex Bis-Tris 4–12% gradient gels (Invitrogen) and subjected to western blot analysis (see 'Western blotting' section for further details).

**In vitro biochemistry.** For cloning, cDNA encoding full-length human ADNP (amino acid residues 1–1102) was PCR amplified with primers NotI-ADNP-forward (5′-AAAAAAGCGGCCGCATGTTCCAACTTCCTGTCAACAA-3′) and KpnI-ADNP-reverse (5′-AAAAAAGGTACCCTAGGCCTGTTGGCTGCTC-3′) and cloned into a pFast-Bac-derived vector (Invitrogen) in frame with an N-terminal $His_6$-tag. Plasmids encoding full-length or N-terminally truncated ADNP (amino acid residues 229–1102) with a C-terminal Strep-tag II were generated by PCR amplification of *ADNP* cDNA with primers NotI-ADNP-forward and KpnI-ADNP-C-reverse (5′-AAAAAAGGTACCGGCCTGTTGGCTGCTCAGTT-3′)

or NotI-ADNP(ΔN228)-forward (5′-AAAAAAGCGGCCGCATG CCAAAGTCCTATGAAGCTTT-3′) and KpnI-ADNP-C-reverse. The amplified cDNA was cloned into a pAC8-derived vector[36]. Expression constructs encoding full-length human HP1γ (amino acid residues 1–183) were generated by amplification of cDNA using primers NotI-CBX3-forward (5′-AAAAAAGCGGCCGCATGGCCTCCAACAAAACTACATT-3′) and KpnI-CBX3-reverse (5′-AAAAAAGGTACCCTATTGAGCTTCATCTTCTGGA-3′) and cloning into pFast-Bac-derived vectors in frame with an N-terminal $His_6$-tag or Strep-tag II. cDNA for individual chromodomain (amino acid residues 11–81) or CSD (amino acid residues 109–183) domains of HP1γ was amplified with primers NotI-CBX3(11)-forward (5′-AAAAAAGCGGCCGC ATGGGAAAAAAACAGAATGGAAAG-3′) and KpnI-CBX3(81)-reverse (5′-AAAAAAGGTACCCTATTTCTGAGAGTTAAGAAACGC-3′) or NotI-CBX3(109)-forward (5′-AAAAAAGCGGCC GCGATGCTGCTGACAAACCA AGAG-3′) and KpnI-CBX3-reverse and cloned into a pAC8-derived vector in frame with an N-terminal $His_6$-tag. cDNA encoding for full-length human CHD4 (amino acid residues 1–1912) was amplified with primers NotI-CHD4-forward (5′-AAAAAAGCGGCCGCATGGCCAGCGGCCTGGGAT-3′) and KpnI-CHD4-reverse (5′-AAAAAAGGTACCCTACTGTTGCTGTGCAACCTG-3′). The resulting PCR product was cloned into a pAC8-derived vector in frame with an N-terminal $His_6$-tag.

**In vitro reconstitution of ChAHP.** Full-length and truncated versions of ChAHP subunits were subcloned into pAC8 or pFastBac-derived vectors[36]. The following constructs were generated: human ADNP (amino acid residues 1–1102 or 229–1102) with a C-terminal Strep-tag II, N-terminally $His_6$-tagged human CHD4 (isoform 1, residues 1–1912) and N-terminally $His_6$-tagged variants of HP1γ (residues 11–81 or 109–183) were cloned into pAC8-derived vectors. Full-length human ADNP (amino acid residues 1–1102) in frame with an N-terminal $His_6$-tag and full-length human HP1γ (residues 1–183) in frame with an N-terminal Strep-tag II were cloned into pFastBac-derived vectors. Baculoviruses were generated in *Spodoptera frugiperda* Sf9 cells using the Bac-to-Bac method for pFastBac-derived vectors or by cotransfection with viral DNA for pAC8-based vectors. After one round of virus amplification in Sf9 cells, *Trichoplusia ni* High5 cells were infected with the respective Baculovirus (150 μl of virus per 10 ml of High5 cells at a density of $2 \times 10^6$ cells $ml^{-1}$) and collected 48 h after infection. Cells were lysed by sonication in 50 mM Tris, pH 7.5, 300 mM NaCl, 5 mM β-mercaptoethanol, 0.1% Triton X-100, 1 mM PMSF, 1× PIC (Sigma-Aldrich). For pull-down experiments, cell lysate of a 15 ml culture was added to 30 μl of Strep-Tactin Sepharose (IBA) or 30 μl of His-tag purification resin (Roche) and incubated for 1 h at 4 °C. The beads were washed three times with lysis buffer, supplemented with 30 mM imidazole for histidine pull-down reactions. Proteins were eluted by addition of 2× sample buffer (62.5 mM Tris-HCl, pH 6.8, 2% SDS, 25% glycerol, 0.05% bromophenol blue and 5% β-mercaptoethanol) and analysed by SDS–PAGE and Coomassie staining.

For large-scale expression of the ChAHP complex, 1 l of Hi5 cells coinfected with Baculoviruses encoding for $His_6$-tagged ADNP and Strep-tagged HP1γ was combined with 2 l of Hi5 cells expressing $His_6$-tagged CHD4. Cells were lysed in lysis buffer and the cleared lysate was passed over a 50-ml Strep-Tactin Sepharose (IBA) column. The bound complex was eluted in 50 mM Tris-HCl, pH 7.5, 100 mM NaCl, 5 mM β-mercaptoethanol, 2.5 mM desthiobiotin, and bound to an anion-exchange chromatography column (Poros HQ) equilibrated in 50 mM Tris-HCl, pH 7.5, 100 mM NaCl, 5 mM β-mercaptoethanol. The bound proteins were eluted using a linear NaCl gradient, concentrated and further purified by SEC (HiLoad Superdex 200 26/600) in 50 mM HEPES-OH, pH 7.4, 150 mM NaCl and 0.5 mM TCEP. Fractions containing the ChAHP complex were concentrated and reinjected to a Superdex200 10/300 column equilibrated in the same buffer.

**Computational methods.** *RNA-seq analysis.* All sequencing reads were aligned to the December 2011 (mm10) mouse genome assembly from UCSC[37]. HP1-mutant RNA-seq data were aligned using STAR 2.5.0a with the following settings to allow reporting of one randomly chosen alignment per multi-mapping read: '–outFilterMultimapNmax 20–outMultimapperOrder Random–outSAMmultNmax 1–alignSJoverhangMin 8–alignSJDBoverhangMin 1–outFilterMismatchNmax 999–alignIntronMin 20–alignIntronMax 100000–alignMatesGapMax 100000–outSAMtype BAM SortedByCoordinate'. Aligned and sorted reads were indexed using SAMtools (version 1.2). ADNP-mutant RNA-seq data were aligned in Galaxy using Bowtie with the parameters '-m 1–best –strata'[38]. Aligned Bam files were imported in R using QuasR (1.14.0)[39]. BigWig files normalized for sequencing depth were generated using the QuasR qExportWig function. Reads were counted over exons using the qCount function and collapsed to yield one value per gene. This count table was used for differential expression calling with the EdgeR package[40]. To compare the different *Cbx* knockout cell lines with *Adnp* knockouts (Extended Data Fig. 7), all biological replicates of the parental/untreated cell lines for *Cbx3* and *Adnp* were used as control group, whereas the respective knockout replicates were considered the treatment group.

*Gene Ontology analysis.* Gene Ontology term analysis of upregulated gene sets was performed using goana from the R limma package[41]. For the analysis significantly upregulated genes (FDR $\leq 0.01$, fold change $\geq 4$) from EdgeR output were used.

*Repeat analysis.* RNA-seq libraries were mapped to the genome using STAR 2.5.0a with settings optimized for maximum repeat recovery/mappability (–outFilterType Normal–alignEndsType Local–winAnchorMultimapNmax 5000–seedPerWindowNmax 1000–alignTranscriptsPerReadNmax 100000–seedNoneLociPerWindow 100–alignWindowsPerReadNmax 20000–alignTranscriptsPerWindowNmax 1000–outFilterMultimapNmax 100000–outSAMattributes NH HI NM MD AS nM–outMultimapperOrder Random–outSAMmultNmax 1). The resulting alignment file was intersected with repeat masker coordinates for mm10 (repeat masker 2012-02-07 update, downloaded from UCSC table browser), alignments overlapping repeats were counted for all repeat classes and normalized to 1 million mapping reads per library.

**ChIP–seq read alignment.** ChIP–seq data were aligned in R using the qAlign function from the QuasR package[39] with default settings, which calls the Bowtie aligner with parameters '-m 1 —best –strata'[38]. Depth-normalized BigWig files were generated using QuasR 1.14.0. For H3K9me3, STAR 2.5.0a (–alignIntronMax 1–alignEndsType EndToEnd–outFilterType Normal–seedSearchStartLmax 30–outFilterMultimapNmax 10000–outSAMattributes NH HI NM MD AS nM–outMultimapperOrder Random–outSAMmultNmax 1–outSAMunmapped Within) was used and non-aligning and multiple mappers were filtered out using samtools. BigWig files displaying the full length for uniquely mapping reads were generated using bedtools and bedGraphToBigWig (UCSC binary utilities).

**Peak finding.** ADNP peaks were called on ChIP replicates using the corresponding inputs as background (all BAM files from QuasR alignment). MACS (version 2.1.1.20160309)[42] was run with the default parameters. Peaks detected in at least two out of three replicates were kept.

HP1γ peaks were called on both wild-type and knockout ADNP ChIP replicates individually, using the corresponding inputs as background (all BAM files from QuasR alignment). MACS was run with the following options:–nomodel–shift 100–extsize 200. Subsequently, peak lists were intersected using bedtools intersect. Peaks present in both wild-type and knockout ADNP datasets, which did not contain the top scoring ADNP motif, were defined as ADNP-independent HP1γ peaks. Note that the number of ADNP-independent HP1γ peaks is an underestimate. HP1 proteins, particularly in heterochromatic regions, often cover large domains, similar to H3K9me3. However, for consistency reasons and to compile an accurate control group for comparison with the sharp ChAHP peaks, we chose to use MACS settings that are optimized for identification of transcription-factor like, narrow peaks rather than broad peaks. Hence the peaks called here represent a stringent set of the most highly enriched loci in the absence of ADNP. Most of the broader domains, which are also ADNP-independent, were not considered, as their shape is very different compared to ChAHP peaks.

**Motif finding.** HOMER v.4.8 was used with default settings to identify DNA sequence motifs in ADNP peaks[43].

**Heat maps and meta-plots.** Heat maps and meta-plots were generated from averaged replicates using the command line version of deepTools2[44]. Peak centres were calculated based on the peak regions identified by MACS (see above). BigWig coverage files for individual replicates were generated by QuasR (see above). For averaging replicates and for calculating $\log_2$(ChIP/input) ratios, bigwigCompare from deepTools2 was used. To generate histone modification meta-plots for ChAHP-bound loci (Extended Data Fig. 6), we used the following previously published datasets: H3K4me1 (GSE27841)[45], H3K4me2 and H3K27me3 (GSE25532)[46], H3K9ac (GSE31284)[47], H3K9me2 (GSE54412)[48], H3K9me3 (GSE12241)[49].

**ATAC-seq analysis.** Paired-end reads were aligned using STAR 2.5.0a using default parameters except for–alignIntronMax 1 and–alignEndsType EndToEnd. Only uniquely mapping reads (alignment score of 255) were kept for further analysis. These uniquely mapping reads were used to generate bigwig genome coverage files similar to ChIP–seq. Meta-profiles and heat maps were generated using deeptools2. For the meta-profiles, the average fragment count per 10-bp bin was normalized to the mean fragment count in the first and last five bins. This ensures that the background signal is set to one for all experiments.

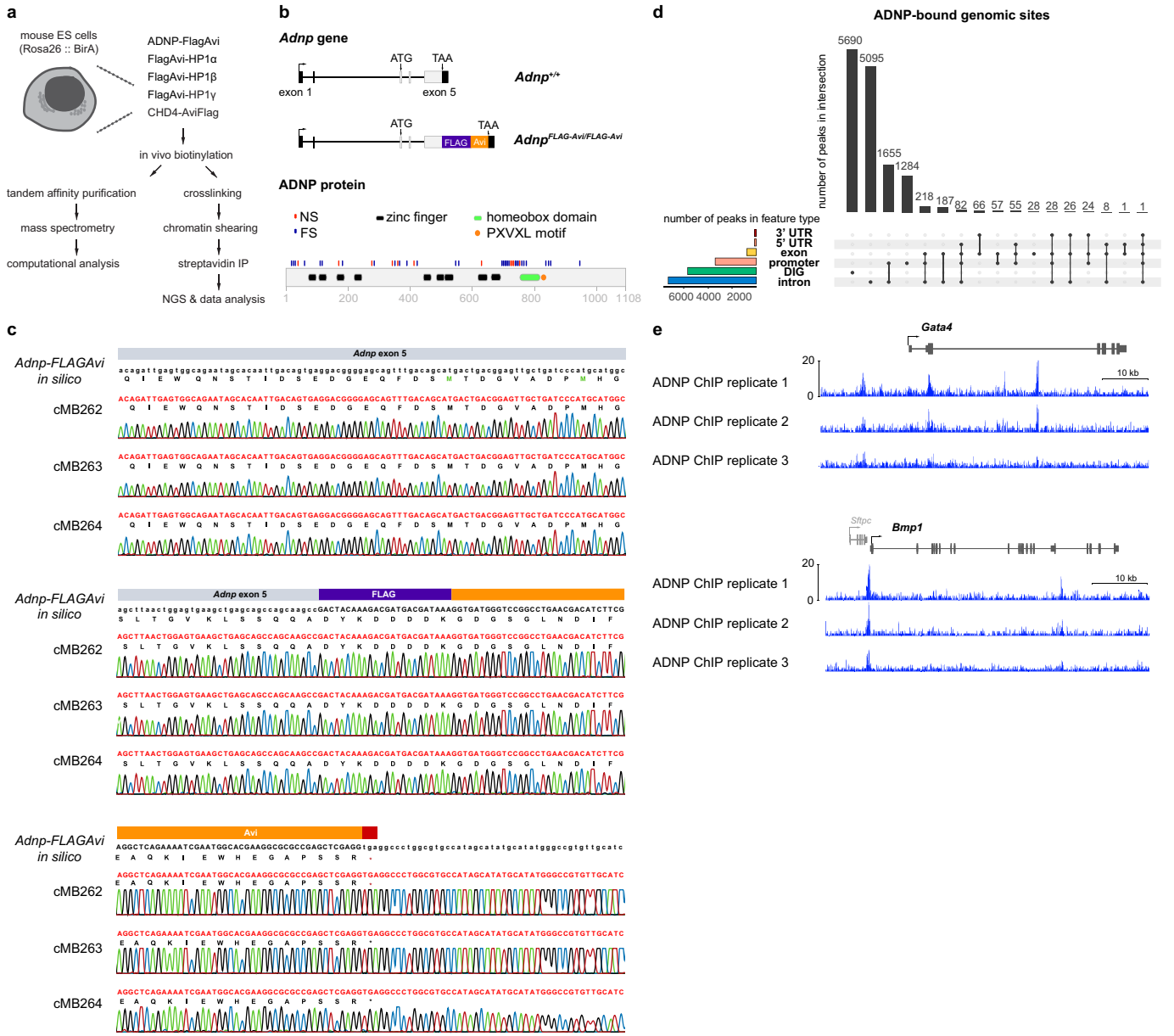
**Reporting summary.** Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

**Code availability.** All custom codes used to analyse data and generate figures are available upon reasonable request.

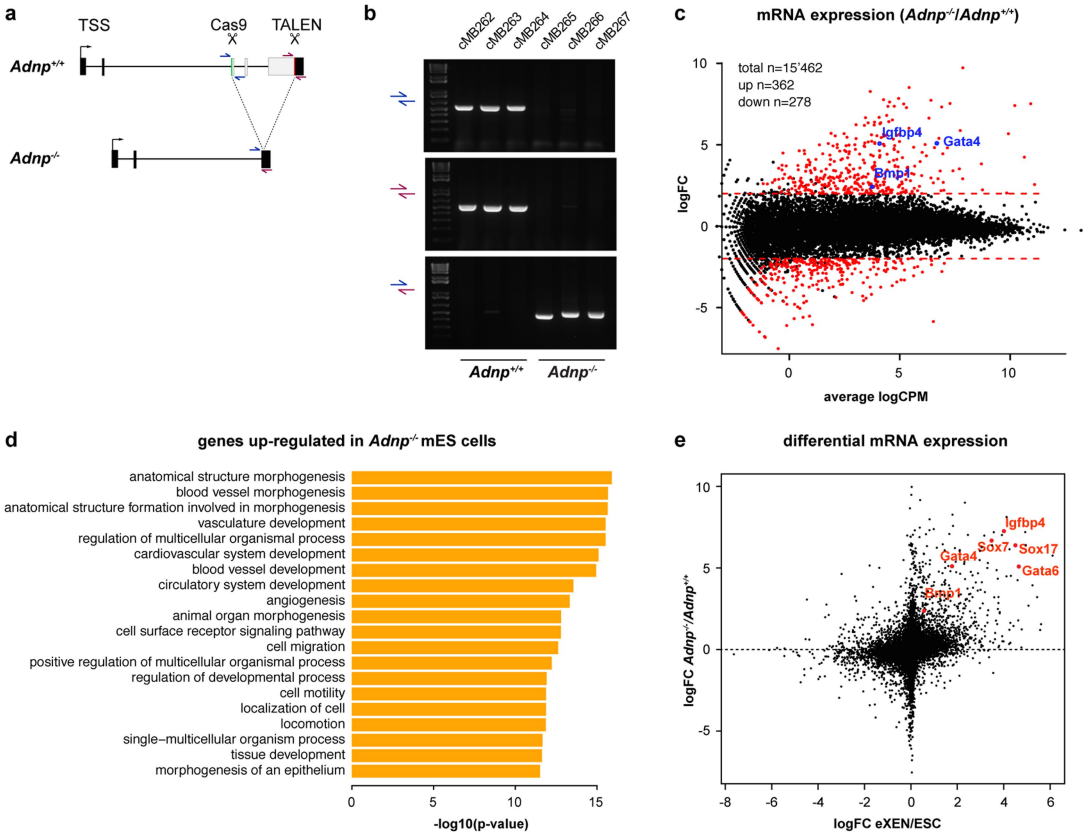
**Data availability.** Genome-wide datasets are deposited at the Gene Expression Omnibus (GEO) under the accession number GSE97945. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE[50] partner repository with the dataset identifier PXD006226.

25. Mohn, F. et al. Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Mol. Cell* **30**, 755–766 (2008).

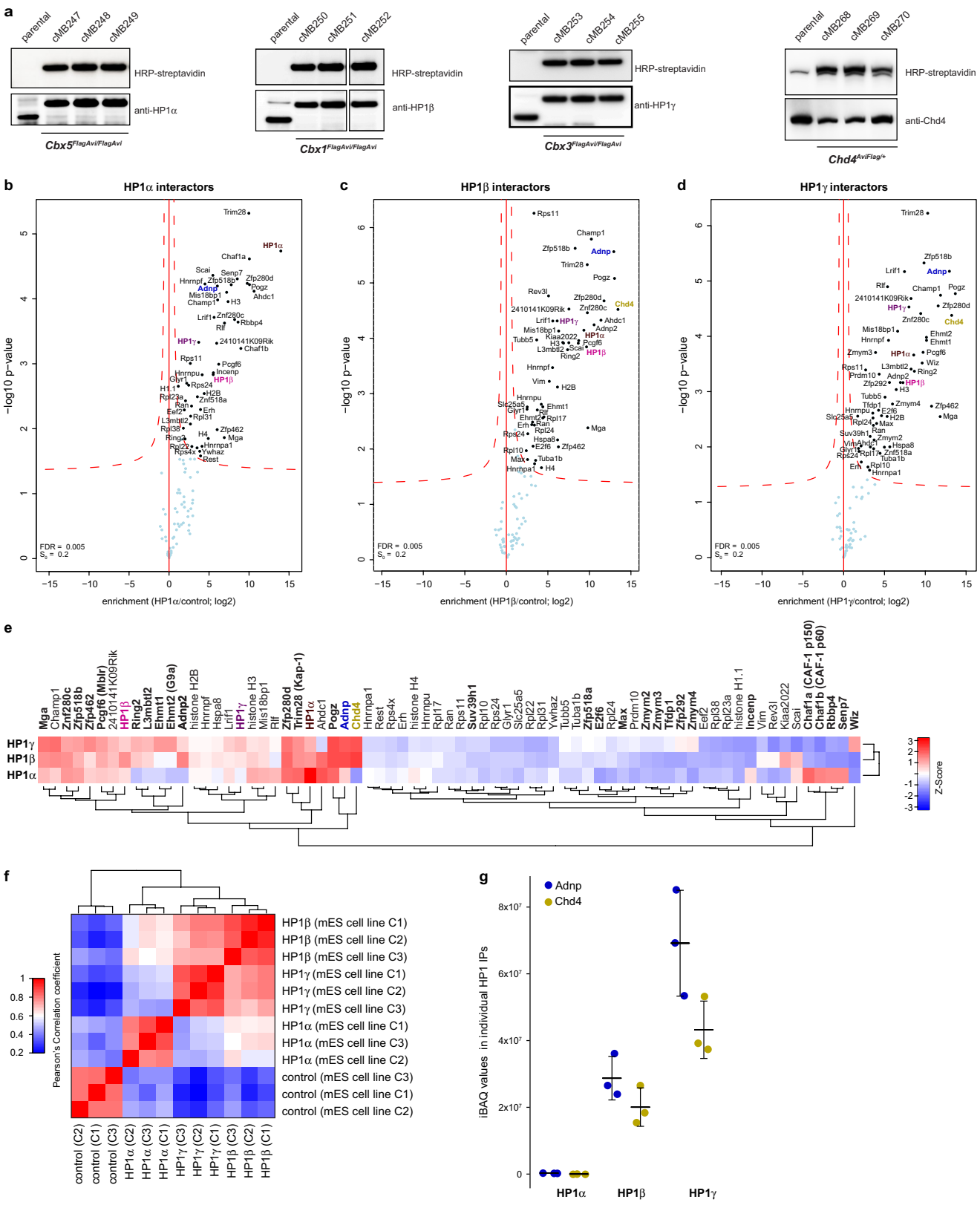
26. Knuckles, P. et al. RNA fate determination through cotranscriptional adenosine methylation and microprocessor binding. *Nat. Struct. Mol. Biol.* **24**, 561–569 (2017).
27. Jacobs, S. A. & Khorasanizadeh, S. Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. *Science* **295**, 2080–2083 (2002).
28. Bibel, M., Richter, J., Lacroix, E. & Barde, Y.-A. Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells. *Nat. Protocols* **2**, 1034–1043 (2007).
29. Cox, J. et al. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10**, 1794–1805 (2011).
30. Cox, J. et al. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics* **13**, 2513–2526 (2014).
31. Hubner, N. C. et al. Quantitative proteomics combined with BAC TransgeneOmics reveals in vivo protein interactions. *J. Cell Biol.* **189**, 739–754 (2010).
32. Tyanova, S. et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**, 731–740 (2016).
33. Tusher, V. G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl Acad. Sci. USA* **98**, 5116–5121 (2001).
34. Schwanhäusser, B. et al. Global quantification of mammalian gene expression control. *Nature* **473**, 337–342 (2011).
35. MacLean, B. et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**, 966–968 (2010).
36. Abdulrahman, W. et al. A set of baculovirus transfer vectors for screening of affinity tags and parallel expression strategies. *Anal. Biochem.* **385**, 383–385 (2009).
37. Rosenbloom, K. R. et al. The UCSC Genome Browser database: 2015 update. *Nucleic Acids Res.* **43**, D670–D681 (2015).
38. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
39. Gaidatzis, D., Lerch, A., Hahne, F. & Stadler, M. B. QuasR: quantification and annotation of short reads in R. *Bioinformatics* **31**, 1130–1132 (2015).
40. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
41. Young, M. D., Wakefield, M. J., Smyth, G. K. & Oshlack, A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* **11**, R14 (2010).
42. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
43. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
44. Ramírez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44** (W1), W160–W165 (2016).
45. Whyte, W. A. et al. Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. *Nature* **482**, 221–225 (2012).
46. Tiwari, V. K. et al. A chromatin-modifying function of JNK during stem cell differentiation. *Nat. Genet.* **44**, 94–100 (2011).
47. Karmodiya, K., Krebs, A. R., Oulad-Abdelghani, M., Kimura, H. & Tora, L. H3K9 and H3K14 acetylation co-occur at many gene regulatory elements, while H3K14ac marks a subset of inactive inducible promoters in mouse embryonic stem cells. *BMC Genomics* **13**, 424 (2012).
48. Liu, N. et al. Recognition of H3K9 methylation by GLP is required for efficient establishment of H3K9 methylation, rapid target gene repression, and mouse viability. *Genes Dev.* **29**, 379–393 (2015).
49. Mikkelsen, T. S. et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553–560 (2007).
50. Vizcaíno, J. A. et al. 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* **44** (D1), D447–D456 (2016).
51. Kwon, S. H. & Workman, J. L. The heterochromatin protein 1 (HP1) family: put away a bias toward HP1. *Mol. Cells* **26**, 217–227 (2008).
52. Murzina, N., Verreault, A., Laue, E. & Stillman, B. Heterochromatin dynamics in mouse cells: interaction between chromatin assembly factor 1 and HP1 proteins. *Mol. Cell* **4**, 529–540 (1999).
53. Maison, C. et al. The SUMO protease SENP7 is a critical component to ensure HP1 enrichment at pericentric heterochromatin. *Nat. Struct. Mol. Biol.* **19**, 458–460 (2012).
54. Smothers, J. F. & Henikoff, S. The HP1 chromo shadow domain binds a consensus peptide pentamer. *Curr. Biol.* **10**, 27–30 (2000).




**Extended Data Fig. 1 | Generation of isogenic mouse ES cell lines to interrogate protein–protein and protein–chromatin interactions.** **a**, Mouse ES cells expressing the BirA biotin ligase from the *Rosa26* locus were used as a parental cell line for endogenous gene tagging with the Flag-AviTag[12]. For a full list of mouse ES cell lines used in this study (cMB#), see Supplementary Information. **b**, Top, scheme depicting Flag-AviTag (not drawn to scale) insertion at the endogenous *Adnp* locus. Arrow indicates transcription start site. Boxes represent exons. Bottom, scheme depicting ADNP protein. Protein domains as predicted by InterPro. Nonsense (NS) and frameshift (FS) mutations found in children with Helsmoortel–Van der Aa syndrome are indicated (https://www.adnpkids.com, dated March 2017). Numbers denote amino acids. **c**, Sanger sequencing of $Adnp^{Flag\text{-}AviTag/Flag\text{-}AviTag}$ cell lines. **d**, Distribution of ADNP-bound genomic sites with respect to protein-coding genes. Peaks were called from ChIP–seq data acquired from three independent biological replicates (that is, three independent $Adnp^{Flag\text{-}AviTag/Flag\text{-}AviTag}$ mouse ES cell lines). Horizontal bars represent peaks annotated to individual categories, and vertical bars represent peaks annotated jointly to specified combinations of categories. DIG, distal intergenic; UTR, untranslated region. **e**, ChIP–seq profiles at two lineage-specifying gene loci that were generated from three independent $Adnp^{Flag\text{-}AviTag/Flag\text{-}AviTag}$ ES lines.

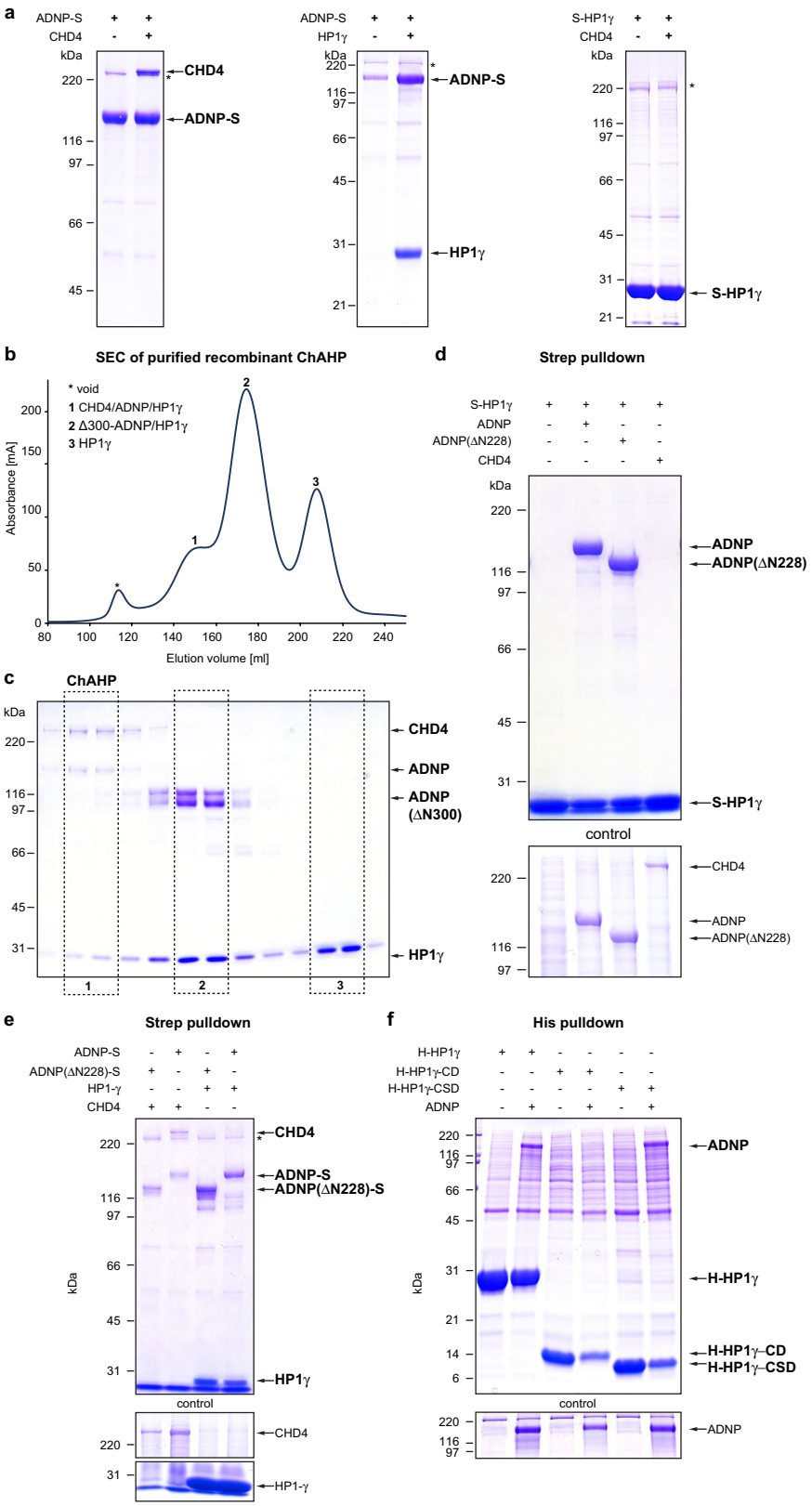



**Extended Data Fig. 2 | Generation and analysis of isogenic *Adnp* knockout mouse ES cell lines. a**, Scheme depicting CRISPR–Cas9 and TALEN-induced double-stranded DNA breaks to delete the *Adnp* open-reading frame. TSS, transcription start site. **b**, PCR genotyping confirming homozygous deletion of the *Adnp* open-reading frame in three different mouse ES cell lines used in this study. The experiment was performed twice. **c**, MA plot comparing fold change (FC) in gene expression for $Adnp^{-/-}$ versus $Adnp^{+/+}$ cells (*y* axis) with mean mRNA abundance (*x* axis). Representative endoderm-specific genes are highlighted in red. Dashed red lines indicate fourfold up- or downregulation. CPM, counts per million. **d**, Gene Ontology enrichment analysis of genes upregulated in $Adnp^{-/-}$ cells. $n = 3$ independent biological replicates. **e**, Scatterplot comparing gene expression fold change upon *Adnp* knockout (*y* axis) with expression changes between extraembryonic endoderm (eXEN) and ES cells (*x* axis). Known key lineage markers are indicated in blue.



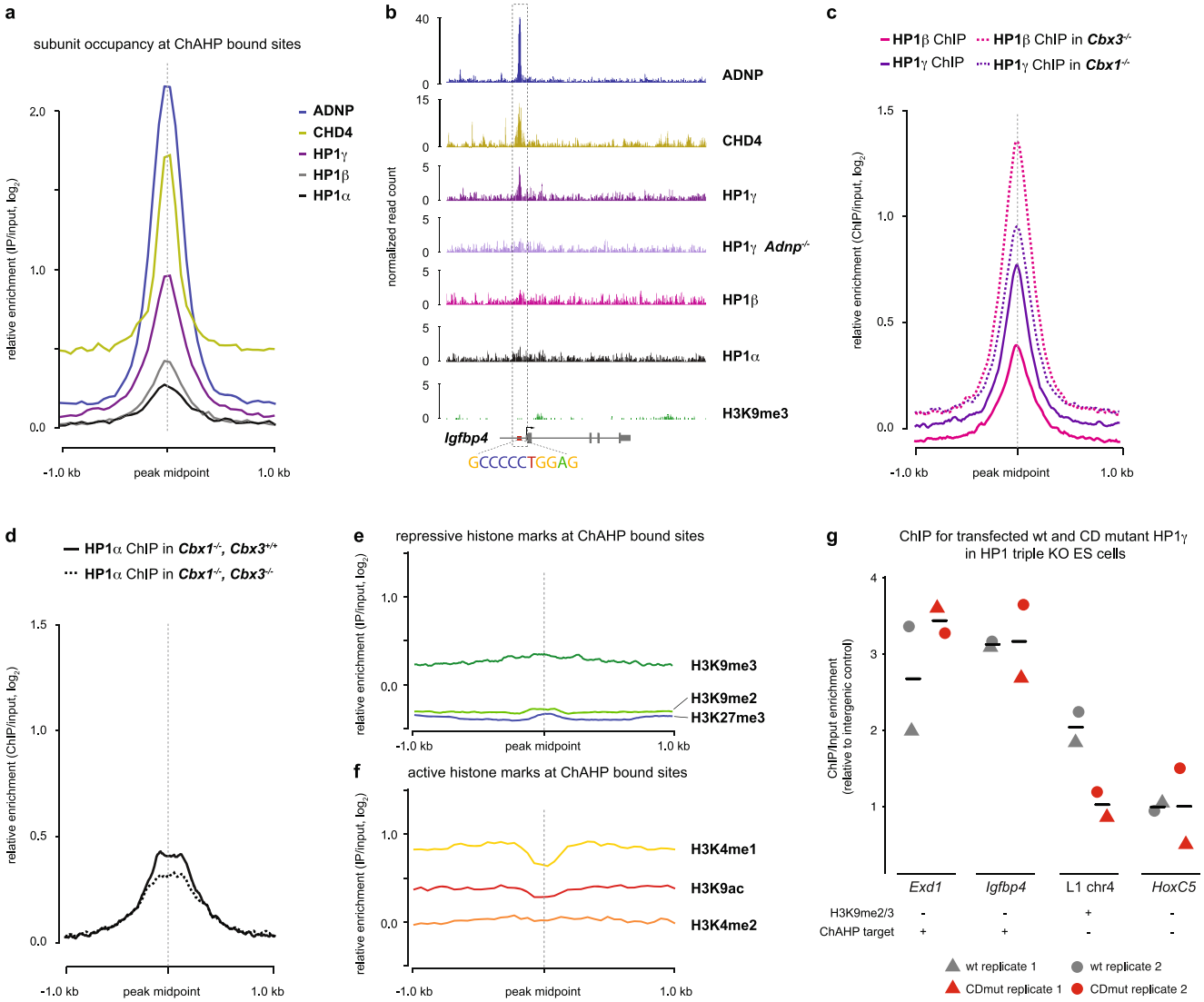

**Extended Data Fig. 3** | See next page for caption.

**Extended Data Fig. 3 | The HP1 interactome of mouse ES cells.** **a**, Isogenic mouse ES cell lines expressing endogenously tagged CHD4 and HP1 proteins. Western blot demonstrating expression of FlagAviTag-tagged proteins. The high molecular mass of CHD4 (218 kDa) does not allow discernable separation of tagged from non-tagged protein. See Supplementary Information for detailed genotype descriptions of the individual ES cell lines. For gel source data, see Supplementary Fig. 1. Experiments were performed twice. **b**, TAP–LC–MS/MS of endogenously FlagAviTag-tagged HP1α. Protein purification was performed in the presence of 350 mM NaCl. Parental ES cell line serves as background control. $n = 3$ independent biological replicates (that is, three independent *Cbx5*$^{Flag\text{-}AviTag/Flag\text{-}AviTag}$ mouse ES cell lines). **c**, TAP–LC–MS/MS of endogenously Flag-AviTag-tagged HP1β. Protein purification was performed in the presence of 350 mM NaCl. Parental ES cell line serves as background control. $n = 3$ independent biological replicates (that is, three independent *Cbx1*$^{Flag\text{-}AviTag/Flag\text{-}AviTag}$ mouse ES cell lines). **d**, TAP–LC–MS/MS of endogenously Flag-AviTag-tagged HP1γ. Protein purification was performed in the presence of 350 mM NaCl. Parental mouse ES cell line serves as background control. $n = 3$ independent biological replicates (that is, three independent *Cbx1*$^{Flag\text{-}AviTag/Flag\text{-}AviTag}$ mouse ES cell lines) **e**, Heat map showing the variation in co-purifying (*Z*-score) proteins across HP1 isoform-specific TAP–LC–MS/MS experiments. Proteins that were significantly enriched in at least one experiment (**b**–**d**) were included in the analysis. **a**–**e**, Validating our approach, all three HP1 isoforms co-precipitated a large number of proteins. Many of these were common to all three HP1 proteins and have previously been described[51]. We also observed several proteins that interacted uniquely with specific isoforms, such as the previously identified CAF-1 or SENP7 interactions with HP1α[52,53]. **f**, Heat map visualization of Pearson's correlation coefficients for the individual HP1 isoform-specific TAP–LC–MS/MS experiments. Three independent biological replicates for HP1α and HP1γ, two biological and one technical replicate for HP1β, and three technical replicates for the parental cell line. **g**, iBAQ values of ADNP and CHD4 in HP1 isoform-specific TAP–LC–MS/MS experiments. Three independent biological replicates for HP1α and HP1γ, two biological and one technical replicate for HP1β. Centre value denotes the mean; error bars denote s.d. **b**–**g**, Statistical analysis was performed using Perseus (see Methods). Mass spectrometry raw data are deposited with ProteomeXchange.




**Extended Data Fig. 4** | See next page for caption.

**Extended Data Fig. 4 | In vitro characterization of ChAHP complex composition. a**, Strep-tag pull-down assays with recombinant human proteins overexpressed in Hi5 insect cells, revealing that ADNP binds to both CHD4 and HP1γ, whereas CHD4 and HP1γ do not interact directly. **b**, SEC of the recombinant ChAHP complex. ChAHP was reconstituted from Hi5 insect cells and further purified by separation according to its molecular mass on a HiLoad Superdex 300 column. Largest fractions eluting first contain ChAHP (1), followed by ADNP–HP1γ (2) and HP1γ alone (3). **c**, Fractions from **b** were separated on SDS–PAGE and visualized by Coomassie staining. **d**, Pull-down analysis of Strep-tagged HP1γ (S-HP1γ) with full-length or N-terminally truncated ADNP (ΔN228) or CHD4. **e**, Pull-down analysis of Strep-tagged full-length or N-terminally truncated ADNP. **b**–**e**, Note that N-terminally truncated ADNP does not co-elute with CHD4 on SEC (**b**, **c**). This is confirmed by pull-down experiments (**d**, **e**), which show that ADNP lacking the first 228 amino acids is only able to bind to HP1γ but no longer to CHD4. Thus, ADNP contacts CHD4 through its N terminus. **f**, Pull-down analysis of His-tagged (H) full-length HP1γ, and isolated chromodomain (CD) and chromoshadow domain (CSD). Similar to other proteins containing the conserved PXVXL pentapeptide[54], ADNP directly interacts with the CSD of HP1γ. This is consistent with the previously reported interaction of ADNP with HP1α[17]. The chromodomain of HP1γ does not bind to ADNP. Experiments in **a**–**f** were performed at least twice. S denotes the streptavidin tag added to the respective protein; asterisks denote a common contaminating protein in streptavidin pull-down assays. For gel source data, see Supplementary Fig. 1.



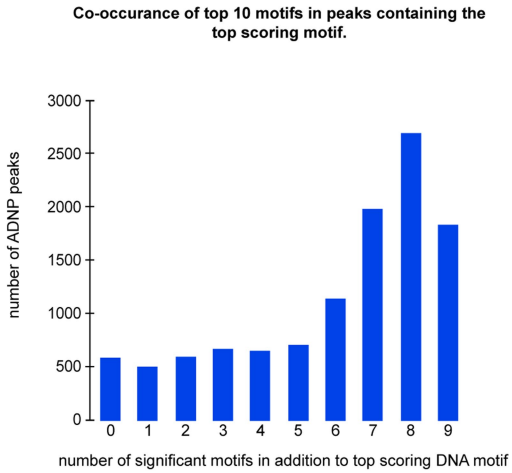

**Extended Data Fig. 5 | HP1 occupancy at ChAHP-binding sites.** **a**, Subunit occupancy at ChAHP-bound sites displayed as meta-profile integrating signal of all peaks. **b**, Genome browser screen shot of the *Igfbp4* locus. ChIP–seq tracks represent depth-normalized read counts of averaged replicate experiments. The predicted ADNP DNA-binding motif upstream of the *Igfbp4* transcription start site is shown. **c**, Average HP1β and HP1γ ChIP–seq enrichment on ChAHP-bound sites in wild-type cells, and average HP1β and HP1γ ChIP–seq enrichment on ChAHP-bound sites in $Cbx3^{-/-}$ and $Cbx1^{-/-}$ mouse ES cell lines, respectively. $n = 2$ biological replicates (that is, independently tagged mouse ES cell lines). **d**, Average HP1α ChIP–seq enrichment on ChAHP-bound sites in wild-type and $Cbx1^{-/-}Cbx3^{-/-}$ double-knockout ES cell lines. $n = 2$ biological replicates. **e**, Histone modifications associated with heterochromatin are absent at ChAHP-bound sites. **f**, Histone modifications associated with active transcription are absent at ChAHP-bound sites. **e**, **f**, Histone modification profiles are displayed as meta-profile integrating signal over all peaks. **g**, Binding of wild-type and chromodomain mutant HP1γ to ChAHP targets (*Igfbp4* and *Exd1*), an H3K9me3-modified region next to an L1 repeat (L1 chr4) and an inactive promoter of an unrelated gene (*Hoxc5*), quantified by ChIP–qPCR. Fold enrichment was normalized to an intergenic control region devoid of HP1γ and H3K9me3. Wild-type (grey) and mutant (red) HP1γ constructs were transiently transfected into HP1 triple-knockout ES cells in biological duplicates. Note the decrease of HP1γ binding at the H3K9me3-modified region (L1 chr4), whereas ChAHP targets remain unaffected in the chromodomain mutant (CDmut) that can no longer bind to H3K9me3. Black lines indicate average enrichments.

**a**

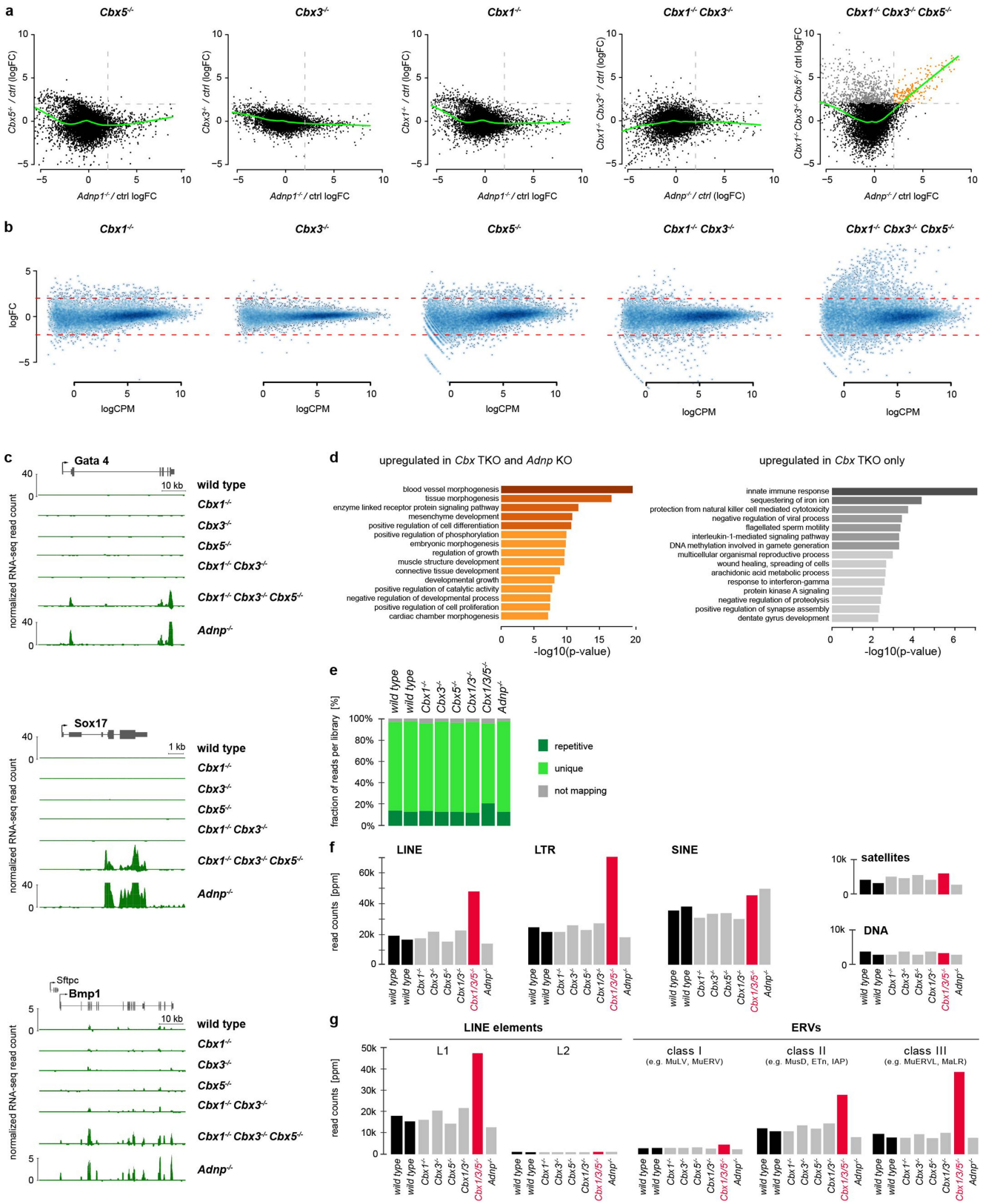
**Table of significant motifs in ADNP peaks**

| Rank | Motif | P-value | % of Targets | % of Background |
|---|---|---|---|---|
| 1 | | 1e-10538 | 63.22% | 2.80% |
| 2 | | 1e-6714 | 55.28% | 4.66% |
| 3 | | 1e-6478 | 34.95% | 0.94% |
| 4 | | 1e-6104 | 55.26% | 5.56% |
| 5 | | 1e-5759 | 56.24% | 6.49% |
| 6 | | 1e-5416 | 49.14% | 4.75% |
| 7 | | 1e-5379 | 52.92% | 6.03% |
| 8 | | 1e-5226 | 29.68% | 0.89% |
| 9 | | 1e-4742 | 42.68% | 3.89% |
| 10 | | 1e-4374 | 36.99% | 2.90% |
| 11 | | 1e-4305 | 33.13% | 2.10% |
| 12 | | 1e-3839 | 37.99% | 3.96% |
| 13 | | 1e-3770 | 46.29% | 7.04% |
| 14 | | 1e-3626 | 29.57% | 2.06% |
| 15 | | 1e-2282 | 37.57% | 7.74% |
| 16 | | 1e-1994 | 23.86% | 3.07% |
| 17 | | 1e-1804 | 30.53% | 6.25% |
| 18 | | 1e-1560 | 26.67% | 5.41% |
| 19 | | 1e-1341 | 26.54% | 6.24% |
| 20 | | 1e-853 | 19.09% | 4.85% |
| 21 | | 1e-471 | 6.50% | 0.94% |
| 22 | | 1e-414 | 9.58% | 2.41% |
| 23 | | 1e-98 | 7.18% | 3.57% |

**b**

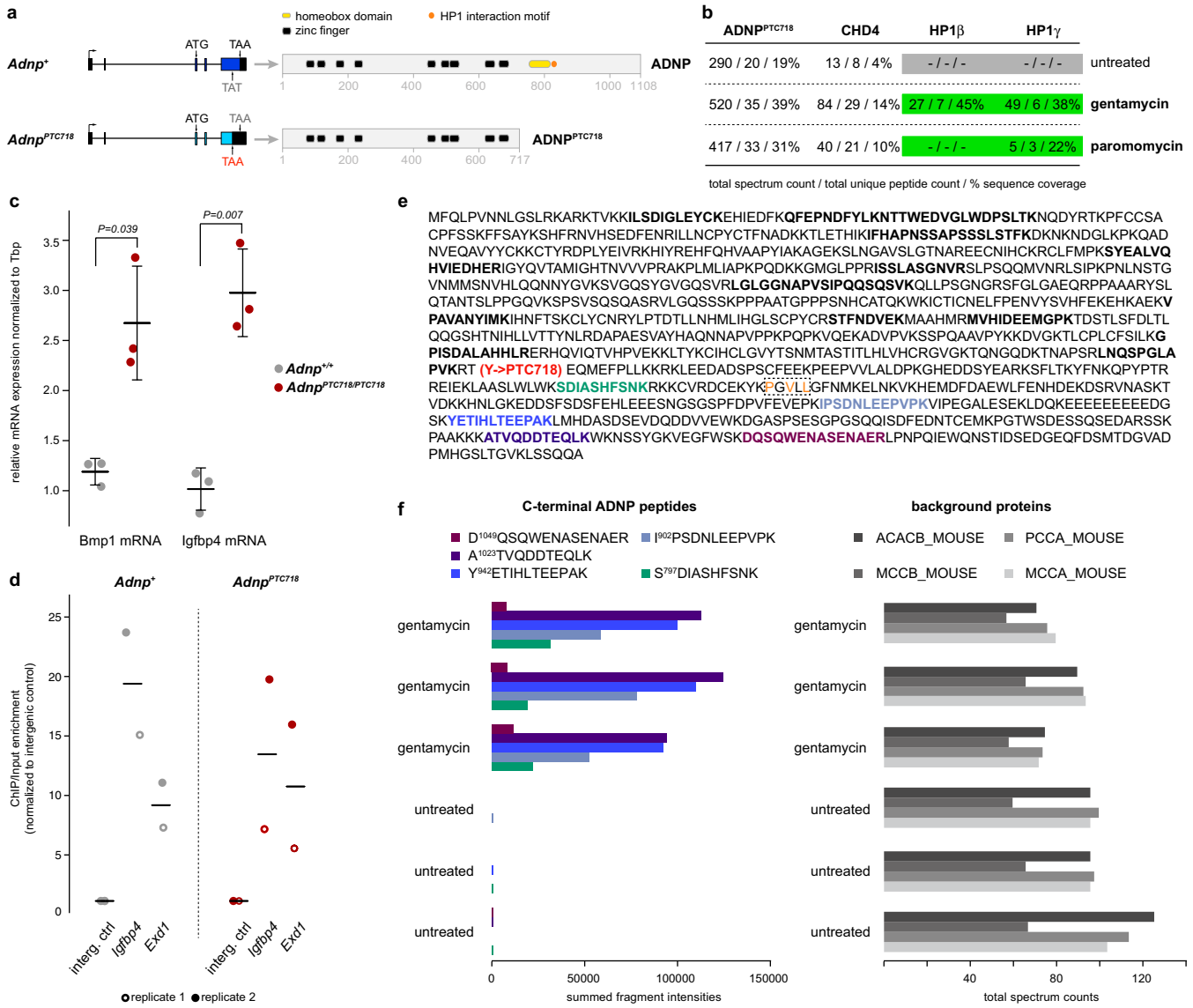



**Extended Data Fig. 6 | Motif analysis of ADNP-bound loci. a**, ADNP DNA-binding motifs predicted by HOMER. Frequency of occurrence and *P* values for motif enrichment compared to genomic background are indicated. $n = 3$ independent cell lines. **b**, Analysis of co-occurrence of the top-ten scoring ADNP DNA motifs. The bar graph shows the frequency of peaks containing the top-scoring ADNP motif and up to nine additional motifs as indicated on the *x* axis. Note that most peaks besides the GCCCCCTGGAG motif also contain more than five other sequence motifs out of the top-ten list.

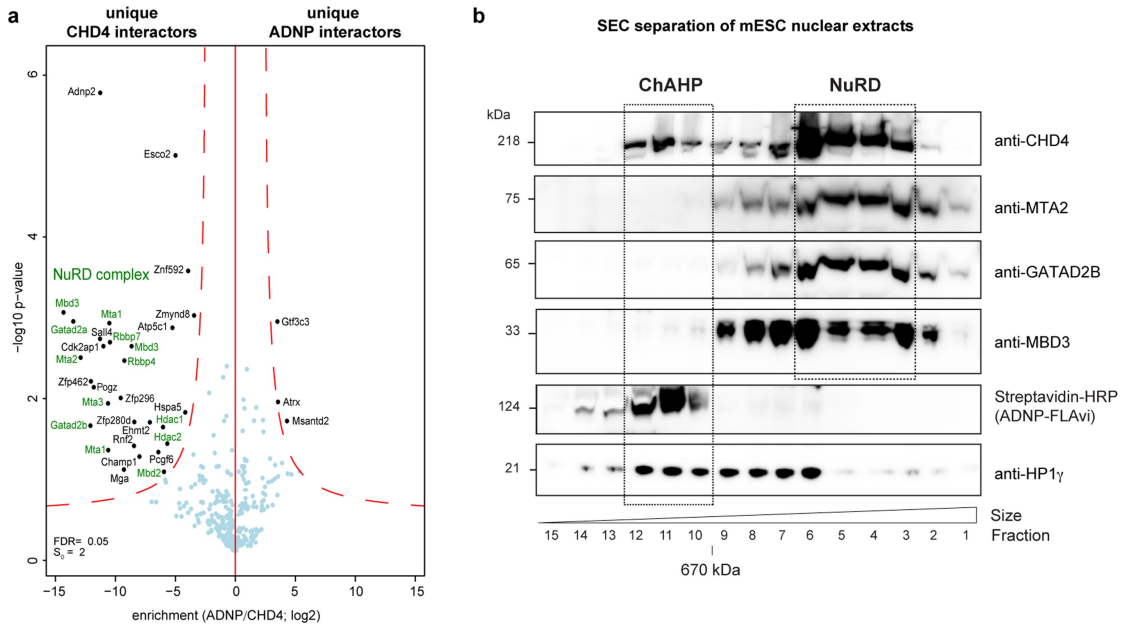



**Extended Data Fig. 7** | See next page for caption.

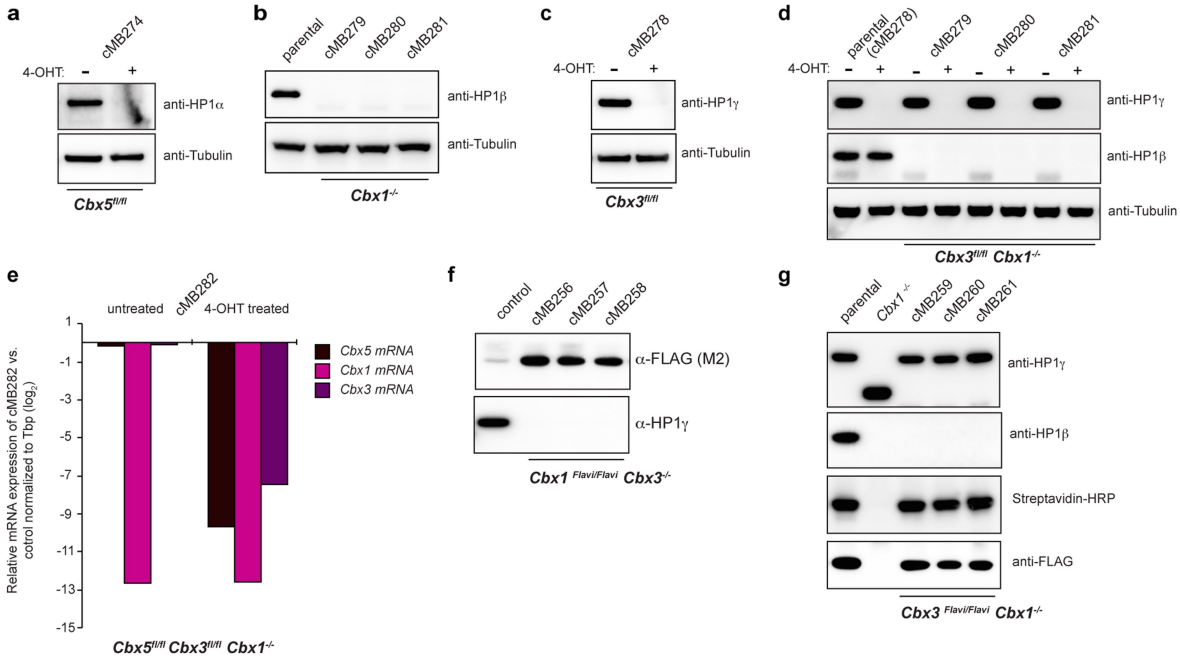
**Extended Data Fig. 7 | Different HP1 isoforms can functionally substitute each other. a**, Scatterplots comparing mRNA expression changes after deletion of *Adnp* versus single, double or triple deletions of *Cbx* genes measured by RNA-seq. Green trend lines indicate a loess (locally weighted scatterplot smoothing) regression. $n = 3$ biological replicates (that is, three independent $Adnp^{-/-}$, $Cbx1^{-/-}$, $Cbx3^{-/-}$, $Cbx5^{-/-}$, $Cbx1^{-/-}Cbx3^{-/-}$ double knockout, or $Cbx1^{-/-}Cbx3^{-/-}Cbx5^{-/-}$ triple knockout mouse ES cell lines). **b**, MA plot displaying fold changes in gene expression for individual *Cbx* knockout cell lines versus wild type. *x* axis denotes the mean mRNA abundance, $\log_2$(counts per million); *y* axis denotes the $\log_2$(fold change) between knockout and wild type. Dashed red lines indicate fourfold up- or downregulation. **c**, UCSC genome browser shots of three lineage-specifying genes. RNA-seq profiles normalized by library size of representative wild-type and mutant ES cell lines are shown. Experiments were performed three times. **d**, Gene Ontology enrichment analysis of the genes that are upregulated in $Cbx1^{-/-}Cbx3^{-/-}Cbx5^{-/-}$ triple-knockout (TKO) and $Adnp^{-/-}$ knockout (KO) cells (orange group of genes in **a**), and of the genes that are upregulated in $Cbx1^{-/-}Cbx3^{-/-}Cbx5^{-/-}$ triple-knockout but not $Adnp^{-/-}$ knockout cells (grey group of genes in **a**). See also Supplementary Table 5. $n = 3$ independent cell lines. **e**, RNA-seq library statistics showing fraction of uniquely, multi- and non-mapping reads. Note the increase in multi-mappers in the HP1 triple-knockout cells. **f**, Quantification of reads mapping to the major repeat classes in counts per million mappable reads. **g**, Quantification of reads mapping to the different LINE and LTR elements in counts per million mappable reads. All mutant cell lines were derived from the same parental mouse ES cell line through direct genome editing and are therefore isogenic.




**Extended Data Fig. 8 | A patient-specific nonsense mutation in *Adnp* impairs the interaction with HP1 but not with DNA. a**, Scheme depicting the wild-type and mutant *Adnp* alleles, which code for Tyr (blue) and a patient-specific premature termination codon (red) at amino acid position 718, respectively. Full-length and truncated protein products are shown on the right. Arrow indicates transcription start site. Boxes represent exons. Numbers denote amino acids. **b**, N-terminally Flag-AviTag-tagged ADNP$^{PTC718}$ was streptavidin-purified from cells with and without aminoglycoside treatment (gentamycin or paromomycin) and subjected to LC–MS/MS analysis. ADNP$^{PTC718}$-expressing cells were treated with 2 mg ml$^{-1}$ gentamycin (2.9 mM) or paromomycin (3.2 mM) for 24 h. The table depicts total spectral counts, unique peptides and percentage sequence coverage (derived from Scaffold) for all ChAHP components from the different treatments. **c**, qRT–PCR measurement of *Bmp1* and *Igfbp4* mRNA levels in ES cells expressing full-length *Adnp* (*Adnp*$^{+/+}$) or C-terminally truncated *Adnp* that interacts with CHD4 but not with HP1 (*Adnp*$^{PTC718/PTC718}$). $n = 3$ biological replicates (that is, three independent RNA isolations). *P* values were calculated using two-tailed unpaired unequal variances *t*-tests. Centre value denotes the mean; error bars denote s.d. **d**, ChIP–qPCR enrichments for transiently transfected Flag-AviTag-tagged wild-type ADNP and ADNP$^{PTC718}$ constructs on two ADNP targets, normalized to an intergenic control. Black lines indicate means. **e**, C-terminally Flag-AviTag-tagged ADNP$^{PTC718}$ was streptavidin-purified from cells with or without gentamycin treatment (2.9 mM) and subjected to LC–MS/MS analysis. Bold letters indicate unique peptides further quantified by parallel reaction monitoring (PRM). C-terminal peptides encoded downstream of PTC718 are shown in colour. Dashed box denotes the HP1 interaction motif. **f**, Summed fragment intensities of five C-terminal ADNP peptides that are encoded downstream of PTC718 are shown on the left. Background proteins shown on the right serve as loading controls. Intensities were measured by PRM. Total spectrum counts were derived from Scaffold. $n = 3$ biological replicates.




**Extended Data Fig. 9 | ChAHP and NuRD are distinct protein complexes. a**, Single-step purification followed by LC–MS/MS of endogenously Flag-AviTag-tagged CHD4 and ADNP. Protein purification was performed in the presence of 350 mM NaCl. Proteins that interact predominantly with CHD4 or ADNP are indicated by UniProt names. NuRD complex components are labelled in green. $n = 3$ biological replicates (that is, three independent $Chd4^{Flag\text{-}AviTag/Flag\text{-}AviTag}$ and $Adnp^{Flag\text{-}AviTag/Flag\text{-}AviTag}$ ES cell lines). Statistical analysis was done with Perseus (see Methods for details). Mass spectrometry raw data are deposited with ProteomeXchange. **b**, SEC of nuclear protein extracts from $Adnp^{Flag\text{-}AviTag/Flag\text{-}AviTag}$ ES cells. Each fraction (indicated at the bottom) was resolved by SDS–PAGE and immunoblotted with the indicated antibodies. Molecular mass of individual proteins is indicated on the left. For gel source data, see Supplementary Fig. 1. Experiment was performed twice.




**Extended Data Fig. 10 | Isogenic *Cbx* knockout ES cell lines. a**, Western blot demonstrating depletion of HP1α protein in $Cbx5^{fl/fl}$ mouse ES cell line after treatment with 4-hydroxytamoxifen (4-OHT). **b**, Western blot demonstrating depletion of HP1β protein in three independent $Cbx1^{-/-}$ ES cell lines. **c**, Western blot demonstrating depletion of HP1γ protein in $Cbx3^{fl/fl}$ ES cell line after treatment with 4-OHT. **d**, Western blot demonstrating depletion of HP1β and HP1γ proteins in three independent $Cbx1^{-/-}Cbx3^{fl/fl}$ double-knockout cell lines after treatment with 4-OHT. $n = 3$ independent 4-OHT treatments. **e**, qRT–PCR demonstrating depletion of *Cbx5*, *Cbx1* and *Cbx3* mRNAs in the $Cbx1^{-/-}Cbx3^{fl/fl}Cbx5^{fl/fl}$ triple-knockout cell line upon treatment with 4-OHT. **f**, Western blot demonstrating depletion of HP1γ protein in three independent $Cbx1^{Flag\text{-}AviTag/Flag\text{-}AviTag}$ cell lines. **g**, Western blot demonstrating depletion of HP1β protein in three independent $Cbx3^{Flag\text{-}AviTag/Flag\text{-}AviTag}$ cell lines. For gel source data shown, see Supplementary Fig. 1. Experiments were performed twice.

nature research

Corresponding author(s): Marc Bühler

# Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

## Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

| n/a | Confirmed | |
|---|---|---|
| ☐ | ☒ | The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ | An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ | The statistical test(s) used AND whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☒ | ☐ | A description of all covariates tested |
| ☐ | ☒ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ | A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ | For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted<br>*Give P values as exact values whenever suitable.* |
| ☒ | ☐ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☒ | ☐ | Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated |
| ☐ | ☒ | Clearly defined error bars<br>*State explicitly what error bars represent (e.g. SD, SE, CI)* |

*Our web collection on statistics for biologists may be useful.*

## Software and code

Policy information about availability of computer code

| | |
|---|---|
| Data collection | For data collection and conversion to fastq format RTA 1.18.64 (HiSeq2500), RTA 2.4.11 (NextSeq500) and bcl2fastq2 v2.17 were used. |
| Data analysis | For data analysis, the following publicly available tools have been used: STAR 2.5.0a<br>SAMtools 1.2<br>bedtools 2.26.0<br>MACS 2.1.1.20160309<br>HOMER 4.8<br>deepTools 2.4.2<br>Perseus 1.5.2.6<br>MaxQuant 1.5.3.8<br>Skyline 4.1<br>MASCOT 2.5<br>R version 3.4.3<br>edgeR 3.20.5<br>limma 3.34.5<br>QuasR 1.14.0 |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

## Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Genome-wide data sets are deposited at GEO under the accession number GSE97945. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD006226.

# Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

# Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|---|---|
| Sample size | No statistical methods were used to predetermine sample size. All experiments were conducted with at least two independent biological replicate cell lines. For full list of cell lines used in this study see supplementary information. |
| Data exclusions | no data was excluded |
| Replication | Experimental data was reliably reproduced. Each experiment was performed at least twice. |
| Randomization | samples were allocated to groups according to genotype |
| Blinding | No blinding. Group allocation according to genotype was done before data collection. |

# Reporting for specific materials, systems and methods

### Materials & experimental systems

| n/a | Involved in the study |
|---|---|
| ☐ | ☒ Unique biological materials |
| ☐ | ☒ Antibodies |
| ☐ | ☒ Eukaryotic cell lines |
| ☒ | ☐ Palaeontology |
| ☒ | ☐ Animals and other organisms |
| ☒ | ☐ Human research participants |

### Methods

| n/a | Involved in the study |
|---|---|
| ☐ | ☒ ChIP-seq |
| ☒ | ☐ Flow cytometry |
| ☒ | ☐ MRI-based neuroimaging |

## Unique biological materials

Policy information about availability of materials

| | |
|---|---|
| Obtaining unique materials | Published research reagents from the FMI are shared with the academic community under a Material Transfer Agreement (MTA) having terms and conditions corresponding to those of the UBMTA (Uniform Biological Material Transfer Agreement). |

## Antibodies

| | |
|---|---|
| Antibodies used | ChIP: mouse-anti-Chd4 (10 µg/IP, Abcam, ab70469), Dynabeads™ M-280 Strepavidin (400µg/IP, Thermofisher)<br>Western blot: mouse anti-FLAG (1:1000, Sigma, clone M2), goat-anti-HP1alpha (1:1000, Abcam, ab77256), mouse-anti-HP1alpha (1:1000, Millipore, mab3446), rat-anti-HP1beta (1:500, Serotec, MCA1946), mouse-anti-HP1gamma (1:2000, Cell Signaling Technology; 2619), mouse-anti-Chd4 (1:1000, Abcam, ab70469), rabbit-anti-Mta2 (1:1000, Bethyl, A300-395A-T), rabbit-anti- |

| | |
|---|---|
| | Gatad2b (1:1000, Bethyl, A301-283A-T), rabbit-anti-Mbd3 (1:1000, Bethyl, A302-528A-T) and rat-anti-tubulin (1:5000, Abcam clone YL1/2)<br>secondary antibodies: HRP-conjugated affinipure goat anti-mouse IgG (Jackson immunology), HRP-conjugated affinipure goat anti-rabbit IgG (Jackson immunology), HRP-conjugated affinity isolated rabbit anti-goat IgG (Sigma A5420), HRP-conjugated rabbit polyclonal anti.rat IgG-H&L (Abcam, ab6734, lot: 514318) |
| Validation | Western blot - correct size of the detected bands based on the protein marker; antibodies used in ChIP were first tested by WB |

## Eukaryotic cell lines

Policy information about cell lines

| | |
|---|---|
| Cell line source(s) | mES cells (hybrid 129xC57Bl6 ) were kindly provided by Dirk Schübeler (FMI, Basel) and served as parental cell line to generate all the cell lines described in Methods and listed in supplementary information. Hi5 insect cells were purchased from Invitrogen (now ThermoFisher Scientific). |
| Authentication | Genotyping followed by Sanger sequencing, followed by qPCR or Western blot, whole genome sequencing |
| Mycoplasma contamination | cell lines were routinely tested and found negative for mycoplasma infection. |
| Commonly misidentified lines<br>(See ICLAC register) | no commonly misidentified cell lines were used |

## ChIP-seq

### Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

| | |
|---|---|
| Data access links<br>*May remain private before publication.* | Genome-wide data sets are deposited at GEO under the accession number GSE97945. |
| Files in database submission | Adnp_wt_Input_r1.bw, Adnp_wt_Input_r2.bw, Adnp_wt_Input_r3.bw, Adnp_wt_ChIP_r1.bw, Adnp_wt_ChIP_r2.bw, Adnp_wt_ChIP_r3.bw, HP1a_wt_Input_r1.bw, HP1a_wt_Input_r2.bw, HP1a_wt_Input_r3.bw, HP1a_wt_ChIP_r1.bw, HP1a_wt_ChIP_r2.bw, HP1a_wt_ChIP_r3.bw, HP1b_wt_Input_r1.bw, HP1b_wt_Input_r2.bw, HP1b_wt_Input_r3.bw, HP1b_wt_ChIP_r1.bw, HP1b_wt_ChIP_r2.bw, HP1b_wt_ChIP_r3.bw, HP1g_wt_Input_r1.bw, HP1g_wt_Input_r2.bw, HP1g_wt_Input_r3.bw, HP1g_wt_ChIP_r1.bw, HP1g_wt_ChIP_r2.bw, HP1g_wt_ChIP_r3.bw, HP1b_Cbx3KO_Input_r1.bw, HP1b_Cbx3KO_Input_r2.bw, HP1b_Cbx3KO_ChIP_r1.bw, HP1b_Cbx3KO_ChIP_r2.bw, HP1g_Cbx1KO_Input_r1.bw, HP1g_Cbx1KO_Input_r2.bw, HP1g_Cbx1KO_ChIP_r1.bw, HP1g_Cbx1KO_ChIP_r2.bw, HP1g_AdnpKO_Input_r1.bw, HP1g_AdnpKO_Input_r2.bw, HP1g_AdnpKO_Input_r3.bw, HP1g_AdnpKO_ChIP_r1.bw, HP1g_AdnpKO_ChIP_r2.bw, HP1g_AdnpKO_ChIP_r3.bw, H3K9me3_wt_ChIP.bw, H3K9me3_wt_Input.bw, H3K9me3_wt_ChIP_r2.bw, H3K9me3_wt_Input_r2.bw, H3K9me3_wt_ChIP_r3.bw, H3K9me3_wt_Input_r3.bw, HP1a_Cbx1KO_Cbx3KO_r1_input.bw, HP1a_Cbx1KO_Cbx3KO_r1_IP.bw, HP1a_Cbx1KO_Cbx3KO_r2_input.bw, HP1a_Cbx1KO_Cbx3KO_r2_IP.bw, HP1a_Cbx1KO_r1_input.bw, HP1a_Cbx1KO_r1_IP.bw, HP1a_Cbx1KO_r2_input.bw, HP1a_Cbx1KO_r2_IP.bw, Adnp_wt_Input_r1.fastq.gz, Adnp_wt_Input_r2.fastq.gz, Adnp_wt_Input_r3.fastq.gz, Adnp_wt_ChIP_r1.fastq.gz, Adnp_wt_ChIP_r2.fastq.gz, Adnp_wt_ChIP_r3.fastq.gz, HP1a_wt_Input_r1.fastq.gz, HP1a_wt_Input_r2.fastq.gz, HP1a_wt_Input_r3.fastq.gz, HP1a_wt_ChIP_r1.fastq.gz, HP1a_wt_ChIP_r2.fastq.gz, HP1a_wt_ChIP_r3.fastq.gz, HP1b_wt_Input_r1.fastq.gz, HP1b_wt_Input_r2.fastq.gz, HP1b_wt_Input_r3.fastq.gz, HP1b_wt_ChIP_r1.fastq.gz, HP1b_wt_ChIP_r2.fastq.gz, HP1b_wt_ChIP_r3.fastq.gz, HP1g_wt_Input_r1.fastq.gz, HP1g_wt_Input_r2.fastq.gz, HP1g_wt_Input_r3.fastq.gz, HP1g_wt_ChIP_r1.fastq.gz, HP1g_wt_ChIP_r2.fastq.gz, HP1g_wt_ChIP_r3.fastq.gz, HP1b_Cbx3KO_Input_r1.fastq.gz, HP1b_Cbx3KO_Input_r2.fastq.gz, HP1b_Cbx3KO_ChIP_r1.fastq.gz, HP1b_Cbx3KO_ChIP_r2.fastq.gz, HP1g_Cbx1KO_Input_r1.fastq.gz, HP1g_Cbx1KO_Input_r2.fastq.gz, HP1g_Cbx1KO_ChIP_r1.fastq.gz, HP1g_Cbx1KO_ChIP_r2.fastq.gz, HP1g_AdnpKO_Input_r1.fastq.gz, HP1g_AdnpKO_Input_r2.fastq.gz, HP1g_AdnpKO_Input_r3.fastq.gz, HP1g_AdnpKO_ChIP_r1.fastq.gz, HP1g_AdnpKO_ChIP_r2.fastq.gz, HP1g_AdnpKO_ChIP_r3.fastq.gz, H3K9me3_wt_ChIP.fastq.gz, H3K9me3_wt_Input.fastq.gz, H3K9me3_wt_ChIP_r2a.fastq.gz, H3K9me3_wt_ChIP_r2b.fastq.gz, H3K9me3_wt_ChIP_r3a.fastq.gz, H3K9me3_wt_ChIP_r3b.fastq.gz, H3K9me3_wt_Input_r2a.fastq.gz, H3K9me3_wt_Input_r2b.fastq.gz, H3K9me3_wt_Input_r3a.fastq.gz, H3K9me3_wt_Input_r3b.fastq.gz, HP1a_Cbx1KO_Cbx3KO_r1_input.fastq.gz, HP1a_Cbx1KO_Cbx3KO_r1_IP.fastq.gz, HP1a_Cbx1KO_Cbx3KO_r2_input.fastq.gz, HP1a_Cbx1KO_Cbx3KO_r2_IP.fastq.gz, HP1a_Cbx1KO_r1_input.fastq.gz, HP1a_Cbx1KO_r1_IP.fastq.gz, HP1a_Cbx1KO_r2_input.fastq.gz, HP1a_Cbx1KO_r2_IP.fastq.gz |
| Genome browser session<br>(e.g. UCSC) | no longer applicable |

### Methodology

| | |
|---|---|
| Replicates | All data was generated with 2-3 independently tagged mES cell lines. |
| Sequencing depth | All samples were sequenced single end, reads are 50nt long.<br><br>sample total_reads uniquelg_mapped_reads |

Adnp_r1_input 45531949 33260007
Adnp_r2_input 41040125 29946158
Adnp_r3_input 41636557 30484917
Adnp_r1_IP 41661386 29041883
Adnp_r2_IP 38085206 26541174
Adnp_r3_IP 48530469 34809813
HP1g_Adnp-KO_r1_input 45678814 33870008
HP1g_Adnp-KO_r2_input 42467108 31258583
HP1g_Adnp-KO_r3_input 43699943 32476799
HP1g_Adnp-KO_r1_IP 32842004 22321526
HP1g_Adnp-KO_r2_IP 44889420 29045598
HP1g_Adnp-KO_r3_IP 35109090 21150753
HP1a_r1_IP 47043976 31235404
HP1a_r1_input 52437768 38293017
HP1a_r2_IP 54535998 37032216
HP1a_r2_input 51620387 37230624
HP1a_r3_IP 51671472 33674034
HP1a_r3_input 51577728 37279754
HP1b_r1_IP 53002042 34312064
HP1b_r1_input 68044805 49040528
HP1b_r2_IP 44585356 29809808
HP1b_r2_input 55793864 41560084
HP1b_r3_IP 41795773 28668149
HP1b_r3_input 64711564 47574471
HP1b_Cbx3KO_r1_IP 41422135 26755179
HP1b_Cbx3KO_r1_input 48649519 36300614
HP1b_Cbx3KO_r2_IP 55593574 34780831
HP1b_Cbx3KO_r2_input 53412383 39198060
HP1g_r1_IP 52607556 33568161
HP1g_r1_input 67240921 48812998
HP1g_r2_IP 40100424 23783981
HP1g_r2_input 68715330 49723788
HP1g_r3_IP 45046950 31042455
HP1g_r3_input 46085583 34090529
HP1g_Cbx1KO_r1_IP 60061302 41378985
HP1g_Cbx1KO_r1_input 66773350 48687798
HP1g_Cbx1KO_r2_IP 46458292 30007558
HP1g_Cbx1KO_r2_input 53364441 39299621
H3K9me3_r1_IP 38739286 20156760
H3K9me3_r1_input 49519631 41152713
H3K9me3_r2_IP 86730433 33122220
H3K9me3_r2_input 98010497 79343833
H3K9me3_r3_IP 84764823 33502277
H3K9me3_r3_input 85158611 69629951
HP1a_Cbx1KO_r1_IP 32771860 19538747
HP1a_Cbx1KO_r1_input 25703051 20436216
HP1a_Cbx1KO_r2_IP 36358775 22847351
HP1a_Cbx1KO_r2_input 27751050 21811944
HP1a_Cbx1KO_Cbx3KO_r1_IP 35262081 20846058
HP1a_Cbx1KO_Cbx3KO_r1_input 28576000 22816994
HP1a_Cbx1KO_Cbx3KO_r2_IP 37494503 22379899
HP1a_Cbx1KO_Cbx3KO_r2_input 29814815 23576762

Antibodies

Dynabeads™ M-280 Streptavidin (11205D, Thermofisher)
Abcam anti-H3K9me3 (ab8898)

Peak calling parameters

MACS2 version 2.1.1.20160309 was used for peak-calling with the parameters " -f BAM -g mm --keep-dup auto". Peaks were called for each IP sample vs. its matched input control sample.
For HP1gamma peaks, the following non-default settings were added: --nomodel --shift 100 --extsize 200.

Data quality

A q-value cutoff of 0.05 was applied to all Adnp ChIP replicates separately.
Plotting -log10 q-values for each replicate against each other showed a good correlation, indicating reproducibility. For the final peak set, all peaks present in at least 2 out of 3 replicates were retained.

Software

Basic data processing and analysis of enrichments was performed using the R package QuasR. The deepTools suite was used to create heatmaps and meta-profiles of binding around peaks.