Aus dem Institut für Molekularbiologie und Tumorforschung
Geschäftsführender Direktor: Prof. Dr. Alexander Brehm

des Fachbereichs Medizin der Philipps-Universität Marburg

Remodelling regulates the heterochromatin of retrotransposons in mouse embryonic stem cells

Inaugural-Dissertation

zur Erlangung des Doktorgrades

der Naturwissenschaften

(Dr. rer. nat.)

dem Fachbereich Medizin

der Philipps-Universität Marburg

vorgelegt von

Parysatis Sachs

aus Meerbusch, Deutschland

Marburg, 2021

Angenommen vom Fachbereich Medizin der Philipps-Universität Marburg am:

21.04.2021

Gedruckt mit Genehmigung des Fachbereichs Medizin

Dekanin:	Frau Prof. Dr. D. Hilfiker-Kleiner
Referentin:	Frau Dr. Jacqueline Mermoud
Korreferent:	Herr Prof. Dr. Marek Bartkuhn

Für Babl

Table of Contents

List of Abbreviations

µg	Microgram
µl	Microlitre
µM	Micromolar
5mC	5-methylcytosine
ATP	Adenosine triphosphate
cDNA	Complementary DNA
CHD	Chromodomain-helicase-DNA binding (SNFT-like protein subfamily)
ChIP	Chromatin immunoprecipitation
ChIP-seq	Chromatin Immunoprecipitation followed by sequencing
Chr	Chromosome
cm	Centimetre
CpG	cytosine and guanine separated by a phosphate
Ctrl	Control
CUE	Coupling of ubiquitin conjugation to ER degradation
d	Day
DAPI	4′,6-diamidino-2-phenylindole
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DSG	Disuccinimidyl glutarate
DTT	Dithiothreitol
E14	A male mouse ES cell line
EDTA	Ethylenediaminetetraacetic acid
EGTA	Ethylene glycol tetraacetic acid
ERV	Endogenous retroviral element
ERVK10C	ERVK family LTR retrotransposon in mouse, also known as MMERVK10C
ESC	Embryonic stem cell
FA	Formaldehyde
FCS	Foetal calf serum

FLAG	A polypeptide protein tag, sequence motif DYKDDDDK
g	Gram
GFP	Green fluorescent protein
h	Hour
HCl	Hydrochloric acid
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
IAP	Intracisternal A-type particle
ICR	Imprinting control region
IgG	Immunoglobulin
IGV	Integrative Genomics Viewer
IRES	Internal ribosome entry site
Kb	Kilobases
Kd	Knockdown
KOAc	Potassium acetate
KRAB-ZFP	Kruppel-associated box zinc finger proteins
LIF	Leukaemia Inhibitory Factor
LINE	Long interspersed element
lncRNA	Long non-coding RNA
LTR	Long terminal repeat
M	Molar
MEF	Mouse embryonic fibroblasts
MERVL/MuERVL	Murine ERV-L element
mES cells	Mouse embryonic stem cells
Min	Minute
mL	Millilitre
MLV/MuLV	Moloney murine leukaemia virus
mm	Millimetre
mM	Millimolar
$MnCl_2$	Manganese Chloride
MusD/Etn	Mus musculus type D/early transposon (Class II ERV in mice)
MYC tag	A polypeptide derived from the c-myc gene product

NaCl	Sodium chloride
$NaHCO_3$	Sodium bicarbonate
NC	Nitrocellulose
ncRNA	Non-coding RNA
ng	Nanogram
NRT	No reverse transcriptase
NTP	Nucleoside triphosphate
o/n	Overnight
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PGK12.1	A female mouse ES cell line
PHD	Plant homeodomain (A protein domain)
PMSF	Phenylmethylsulfonylfluorid
PVDF	Polyvinylidene fluoride
qPCR	Quantitative polymerase chain reaction
RBCC	RING-B box-coiled coil
rpm	Revolutions per minute
rRNA	Ribosomal RNA
RT	Room temperature
RT-qPCR	Real-time quantitative reverse transcription polymerase chain reaction
SD	Standard deviation
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SE	Standard error
shRNA	Short hairpin RNA
SINE	Short interspersed element
snoRNA	Small nucleolar RNA
TAE	TRIS acetate EDTA
TE	Transposable element
TPM	Tags per million

tRNA	Transfer RNA
UTR	Untranslated region
VL30	Virus-like 30
WT	Wild-type
ZFP	Zinc finger proteins

List of Figures

List of Tables

1. Introduction

1.1. Epigenetics

For a long time, researchers focused on DNA as the primary macromolecule storing genetic information and propagating it from mother to daughter cell, and to the next generation via the germline (Allis et al., 2015). Due to this, classical Mendelian inheritance focused on the differences found in alleles that arose by mutation of the DNA sequence. Non-Mendelian inheritance cannot be explained by allelic transmission. X-inactivation or genomic imprinting, for example manifest through the expression of only one of the two alleles even though they are both present in the same nuclear context (Allis et al., 2015). Other patterns of non-Mendelian inheritance include Barbara McClintock's "jumping genes", now more commonly known as transposable elements. (Goldberg et al., 2007). These patterns are established by the packaging of the DNA molecule. DNA exists as a complex with histone and non-histone proteins (Harr et al., 2016). Together with the DNA they comprise chromatin and distinctive forms of chromatin arise through nucleosome arrays carrying covalent and noncovalent modifications. The process of establishing these modifications encompasses a plethora of different mechanisms; such as post-translational histone modifications (Bannister and Kouzarides, 2011; Tessarz and Kouzarides, 2014), the exchange of histone variants, (Buschbeck and Hake, 2017), energy-dependent chromatin-remodelling that alters or mobilizes nucleosomes (Clapier and Cairns, 2009). In addition, DNA itself can be modified covalently, usually at the cytosine residue of CpG dinucleotides (Jones, 2012). All of these mechanisms work together and are referred to as epigenetic gene regulation or simply epigenetics. Epigenetics gives rise to distinct and dynamic chromatin landscapes, some of them stable throughout several cell divisions and important for cell identity (Clapier et al., 2017a; Du et al., 2015).

1.2. Chromatin

Chromatin is composed of repeating nucleosomal units, each of which consists of an octamer of histone proteins with 147 base pairs (bp) of DNA wrapped around them (Davey et al., 2002). The nucleosome organisation was revealed by biochemical studies (Kornberg, 1974) which were later confirmed by crystallography studies (Luger et al., 1997). Together they revealed that the nucleosome unit is made up of dimer sets of histone partners (H2A with H2B) and tetramers (H3 with H4) together forming an octamer

(Arents et al., 1991). DNA is organized on the octamer surface, wrapping around it, leading to a particle with a defined dyad axis. The crystal structures, however did not accurately reflect the histone tail domains, which are crucial for many of the functional roles of histones since they carry many post-translational modifications (PTMs). Chromatin domains can exist in several distinct functional states and this is determined by histone modifications, histone variants, nonhistone proteins, nucleosome remodelling factors, non-coding RNAs, among many other determinants (Zhou et al., 2019).

1.2.1. Euchromatin and heterochromatin

Historically chromatin has been characterized as euchromatin and heterochromatin. Euchromatin encompasses regions of "active" chromatin: coding and regulatory elements such as promoters and enhancers. The euchromatic state is generally more "open" which means decompacted and pertaining a more nuclease-sensitive configuration. The genes and regulatory elements within it are poised for transcription but not necessarily transcriptionally active even if many of them are compared to the more repressed heterochromatin. Present in euchromatic regions are the housekeeping genes of the cells which are ubiquitously active as well as developmentally or stress-regulated genes (Quina et al., 2006). Nucleosome remodelling machines, histone modifications and histone variants all play a role in regulating the euchromatic landscape and are important for the engagement of the transcription machinery (Figure 1.2.1) (Klemm et al., 2019).

When it comes to heterochromatin, a distinction is usually made between constitutive and facultative heterochromatin. Constitutive heterochromatin is highly condensed and remains so throughout the cell cycle and throughout development. These condensed regions often have structural functions such as centromeres and telomeres but constitutive silencing is also important for highly repetitive sequences which need to be tightly controlled to protect genome stability (Allis et al., 2015; Grewal and Moazed, 2003). Facultative heterochromatin on the other hand, has the "potential" for gene expression. Depending on the state of development or the cell type it can be either condensed or decondensed (Grewal and Moazed, 2003). Examples of facultative heterochromatin are X-inactivation, genomic imprinting and genes that are developmentally regulated, such as pluripotency factors which are repressed in somatic cells (Avner and Heard, 2001).

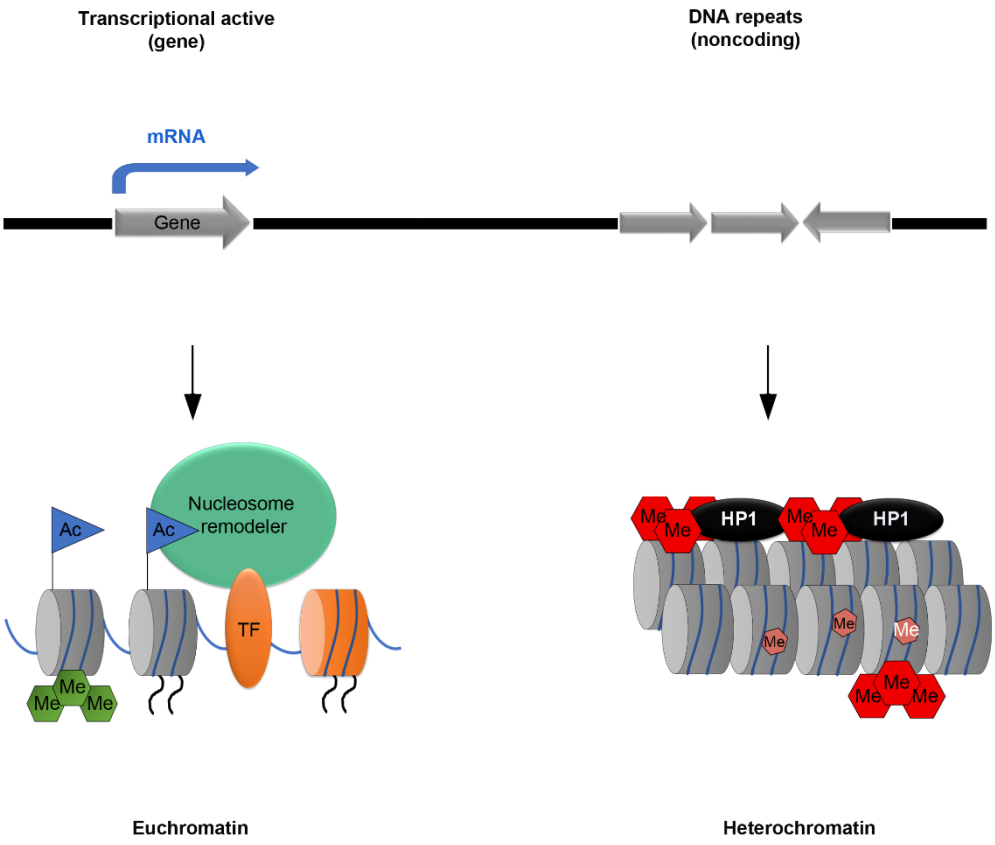

Figure 1.2.1: Characteristics of euchromatin and heterochromatin.

Euchromatin is characterized by transcription factor binding and remodeling complex recruitment associated with active transcription and open chromatin. Activating histone modifications, such as H3 acetylation and H3K4 trimethylation, are prevalent as are histone variants associated with gene transcription (e.g., H2A.Z), depicted in orange. Heterochromatin is often found in noncoding regions, such as on DNA repeats. It is characterized by repressive histone methylation, DNA methylation, and further compaction by HP1 (Histone Protein 1) recruitment and other mechanisms (adapted from (Allis et al., 2015).

Overall, there are certain features that can define these chromatin states which are present to varying degrees. The chromatin state is hardly a binary classification and likely more of a spectrum. For the sake of this introduction euchromatin and heterochromatin are defined by their conventional characterisations (Trojer and Reinberg, 2007).

1.2.2. Higher order chromatin structure

The chromatin fibre can be visualized microscopically as 11 nm "beads on a string" which corresponds to the unfolded configuration of DNA periodically wrapped around nucleosomes. Higher order chromatin conformations occur along the entire genome and shift during cell fate specification and strongly during stages of the cell cycle (Klemm et al., 2019). More compact higher-order chromatin structures of 30 nm arise through the incorporation of H1 linker histone (Robinson and Rhodes, 2006) and other associated proteins and factors such as HP1 (Histone Protein 1). The organisation into larger looped chromatin domains of 300-700 nm possibly occurs through the anchoring of the chromatin fibre to the nuclear periphery or other scaffolds via nuclear Lamins (Allis et al., 2015; Amendola and van Steensel, 2014). Furthermore, this anchoring is not random and it is emerging that apart from structural purposes it also gives rise to meaningful functional chromosome territories and is physiologically relevant (Bickmore, 2013; Bickmore and van Steensel, 2013). For example, in the clustering of silent heterochromatin particularly of pericentric foci. Finally, the most condensed chromatin structure occurs during mitosis or meiosis at the metaphase stage and it is essential for faithful chromosome segregation. This level of chromatin condensation involves a dramatic restructuring of the DNA molecules, that together are roughly 1.8m long, into discrete chromosomes of 1.5 μm in diameter (Allis et al., 2015; Hirano, 2012; Nasmyth and Haering, 2009). Another form of highly condensed chromatin is the pericentric and telomeric region which are very specialized chromosomal domains with important structural and functional roles.

Chromatin is very different in each cell type; a neuron has different gene regulation requirements than a leukocyte after all. This thesis will investigate chromatin remodelling in the unique "open" environment of embryonic stem cells.

1.2.3. Chromatin landscapes in embryonic stem cells

Embryonic stem cells are a synthetic cell line that is cultured *in vitro* and is derived from the inner cell mass (ICM) of the blastocyst. The development of the zygote showcases a gradual loss of totipotency in its cells. Over time the mammalian embryo gives rise to the blastocyst, and at this point the cells have acquired different characteristics; the structure of the blastocyst consists of a layer of outer cells called the trophectoderm cells which will become the placenta. The cells of the ICM of the blastocyst are pluripotent, as in they

will give rise to the foetus and ultimately to the entire new organism (Hackett and Surani, 2014). Therefore, they contain the potential to evolve into any of the ~200 specialized somatic cells. They are the pluripotent stem cells and they differ from somatic and differentiated cells in their epigenetic landscape. As development proceeds their potency becomes more restricted and the epigenetic marks more rigid. These pluripotent cells of the ICM are referred to as ES cells when they are removed from the embryo and cultured *in vitro.*

To what extent however, are epigenetic mechanisms required to maintain cells in an undifferentiated state? There are clear epigenomic differences between the differentiated state and the undifferentiated one such as a hyperdynamic plasticity of chromatin in pluripotency (Meshorer and Misteli, 2006). Another interlinked aspect is the pluripotency network of proteins. Pluripotency factors Octamer-Binding Transcription Factor 4 (Oct4), Nanog, SRY-Box Transcription Factor 2 (Sox2) are among the important factors to maintain pluripotency and they regulate many other genes involved in ES cell maintenance (Boyer et al., 2005; Loh et al., 2006). The differentiation of ES cells is defined by the reduced expression of these factors and the acquisition of other markers of somatic cells, such as a changed morphology. Epigenetic mechanisms are important for the transcriptional repression of several elements that still require silencing in the less restricted context of pluripotency due to otherwise detrimental effects on genome stability or the maintenance of pluripotency. Among these are pluripotency factors, heterogenous repetitive regions and transposable elements (Hainer et al., 2019). The repression of transposable elements in ES cells is central to this thesis and is discussed in detail in the chapter on endogenous retroviral elements, after the introduction of important factors and modifications in chromatin regulation.

1.3. Features of chromatin states

The chromatin substrate is modified by a multitude of factors and modifications which affect its regulation, architecture and access to the DNA. This thesis will introduce histone modifications, histone variants, ATP-dependent chromatin remodellers and DNA methylation, and then discuss the role they play in genomic imprinting, X-inactivation, and the regulation of ERVs.

1.3.1. Histone modifications

Histone proteins are very conserved in their amino acid sequence from yeast to humans. The core nucleosome consists of several histone proteins which are made up of a globular domain and a highly flexible and unstructured "histone tail". These histone tails are subject to extensive post-translational-modifications (PTMs), particularly of H3 and H4, giving rise to further chromatin variability and distinct functional domains. Since the first studies by Allfrey et al. (1964) and Paik and Kim (1971) that identified acetylation and methylation of core histones, many further histone modifications have been discovered including phosphorylation, sumoylation, ubiquitination, biotinylation, ADP-ribosylation, crotonylation, proline isomerization, among others (Allis et al., 2015).

How are histone modifications established and erased? These processes are carried out by a plethora of chromatin-associated enzymatic systems. The enzymes with roles in these systems often reside within large multisubunit protein complexes that catalyse the incorporation and removal of covalent modifications on their targets. They are generally highly target- and cell-type specific (Bannister and Kouzarides, 2011). The major enzymes in these group are histone acetyltransferases (HATs), their counterpart the histone deacetylases (HDACs), histone kinases, that phosphorylate serine, threonine and lysine residues, the phosphatases, that in turn remove these groups, the histone lysine methyltransferases, and protein arginine methyltransferases (Bannister and Kouzarides, 2011; Kouzarides, 2007). In the context of this study the focus will be on histone lysine methyltransferases and their action is explored in more detail in the chapter dealing with the control of endogenous retroviral elements. The enzymes that counteract the methyltransferases are the histone demethylases and arginine deaminases (Bannister and Kouzarides, 2011; Kouzarides, 2007).

Histone modifications habitually affect chromatin structure. For example; histone tail acetylation has traditionally been linked to active chromatin domains whereas methylation designates both repressed and transcribed areas. Histone modifications work in concert with many other epigenetic factors, which they can often provide a binding platform for. Examples include; transcription factors, histone chaperones, chromatin remodellers, and DNA or histone modifying enzymes (Allis and Jenuwein, 2016; Kouzarides, 2007). The misregulation of histone modifications can have severe effects and has been implicated in developmental defects and the pathogenesis of cancer,

highlighting the biological importance of the regulation of histone modifications (Hyun et al., 2017; Wang and Allis, 2009)

The following study has a particular focus on histone methylation. Methylation can occur on lysine or arginine residues of histone tails and the first lysine histone methyltransferase discovered was Suppressor of Variegation 3-9 Homolog 1 (SuV39H1). It contains a catalytic SET (Su(var)3-9, Enhancer of Zeste, and Trithorax) domain (Rea et al., 2000). Until the first demethylase was discovered histone lysine methylation was expected to have a slow turnover. The discovery of the H3K4 demethylase LSD1 and later JmjC (Shi et al., 2004; Tsukada et al., 2006) demonstrated that these residues can be actively removed in the cell.

There are three distinct lysine methylation states, mono-, di- and trimethylation (me1, me2, and me3). None of them result in changes of the electric charge of the amino acid side-chain, which suggests that they play a role in chromatin dynamics by attracting other effector molecules that contain methyl-lysine binding motifs, such as a PHD (plant homeodomain), chromo or tudor domain (Black et al., 2012). These motifs are very specific since the location of the lysine and its methylation state, results in vastly different outputs, and can lead to activation or repression. Histone lysine methylation on H3K4 is generally associated with activation and is found on active promoters. H3K36 is also associated with activation and found on actively transcribed gene bodies (Black et al., 2012). Marks associated with repression are H3K27, H3K9, and H4K20 methylation. H3K27me3, for example, is a hallmark of transcriptional repression. It is deposited by the histone methyltransferase Enhancer Of Zeste 2 Polycomb Repressive Complex 2 Subunit (EZH2) contained within the polycomb complex PRC2 (Cao et al., 2002; Hyun et al., 2017; Muller et al., 2002; Pirrotta, 1998). Two specific histone lysine methylation marks will become important during this study: H4K20me3 and particularly H3K9me3.

H3K9 methylation

H3K9 methylation is a conserved modification linked to silencing and heterochromatin (Barski et al., 2007). Most notably it is found in the constitutive heterochromatin of the telomeres and centromeres and it is crucial for the repression of many repetitive DNA sequences. Moreover, it is known for its role in chromosomal architecture and is required

for the correct segregation of the chromosomes (Peters et al., 2001; Sarraf and Stancheva, 2004).

In mammalian cells the histone methyltransferases SuV39h1, SuV39h2, SET Domain Bifurcated Histone Lysine Methyltransferase 1 (SETDB1), and G9a are responsible for H3K9 methylation, and they possess vastly different activities and targets. SuV39h1/h2 catalyse H3K9 di-, and trimethylation in constitutive heterochromatin including on pericentric regions (Collins et al., 2005; Lachner et al., 2001; Rea et al., 2000; Rice et al., 2003). The histone methyltransferase G9a consists of a heterodimer of G9a and GLP (G9a-like Protein), and it plays a role in euchromatic gene repression via H3K9me1 and H3K9me2 (Tachibana et al., 2002). SETDB1 catalyses H3K9 monomethylation on pericentric regions where it also provides a substrate for SuV39h1/h2 to produce H3K9me3 (Loyola et al., 2009). On retroviral elements SETDB1 is required for H3K9me3 together with several necessary co-factors, and many endogenous retroviral elements (ERVs) are upregulated upon the depletion of SETDB1 concurrent with the loss of H3K9me3 (Bulut-Karslioglu et al., 2014; Karimi et al., 2011; Liu et al., 2014; Matsui et al., 2010).

There are many examples of crosstalk between H3K9 methylation and DNA methylation. SuV39h1/h2 interact with the de novo DNA methyltransferases-3a/b (DNMT3a/b) (Fuks et al., 2003; Lehnertz et al., 2003), as does G9a-GLP. This illustrates the link of the two mechanisms in the establishment of facultative heterochromatin (Epsztejn-Litman et al., 2008; Feldman et al., 2006). Additionally, components of the DNA methylation machinery such as G9a, and the maintenance DNA methyltransferase-1 (DNMT1) have been found to co-localise at the replication fork with a conserved replication protein called Proliferating Cell Nuclear Antigen (PCNA) (Rothbart et al., 2012), placing H3K9 methylation conjointly with the maintenance of DNA methylation during DNA replication. This is achieved by the protein Ubiquitin Like With PHD And Ring Finger Domains 1 (UHRF1/NP95), which brings DNMT1 to the replication fork by interacting with methylated H3K9 and hemimethylated CpGs, bridging H3K9 methylation and DNA methylation (Hyun et al., 2017; Rothbart et al., 2012).

H3K9 methylation also plays an important role in the establishment of pericentric heterochromatin (Allis et al., 2015). The histone methyltransferases SETDB1 and SuV39h1/h2 are recruited to pericentric heterochromatin and catalyse H3K9 methylation (Dodge et al., 2004; Loyola et al., 2009; Rea et al., 2000). HP1 then binds H3K9me3

through its chromodomains forming multimers that in turn can interact with SuV39h1/h2. HP1 further contributes by recruiting other silencing factors and histone deacetylases HDAC1/2 (Hall et al., 2002; Lechner et al., 2005; Yamada et al., 2005).

H4K20 trimethylation

The methyltransferases that catalyse H4K20me3 are restricted to specific methylation states (Hyun et al., 2017). One example is the enzyme SET Domain-Containing Protein 8 (SET8) which is in charge of monomethylation (Fang et al., 2002; Nishioka et al., 2002). H4K20me1 is further modified to me2 and me3 by Suppressor Of Variegation 4-20 Homolog 1 (Suv4-20H1) and Suppressor Of Variegation 4-20 Homolog 2 (Suv4-20H2), often together with other factors (Schotta et al., 2004; Schotta et al., 2008; Yang et al., 2008). Additionally, H4K20 methylation can be regulated by a histone trans-tail mechanism where the C-terminal region of Suv4-20H2 interacts with HP1 which in turn recognizes H3K9 methylation (Schotta et al., 2004). Hence, the deletion of HP1 or a reduction of H3K9 methylation results in decreased heterochromatin targeting of Suv4-20H2 and in a decrease of H4K20 methylation (Schotta et al., 2004; Yang et al., 2008).

1.3.2. Histone variants

Histone proteins are very conserved, but some evolved into specialized variants of the core histones to allow for another mechanism to introduce variation into the epigenome. This in turn results in different possible chromatin structures and nucleosomal dynamics. All core histone except H4 have evolved histone variants (Maze et al., 2014).

For example, the core histone H3 possesses several histone variants. H3.1 and H3.2 are considered the "canonical" variants. They are deposited during DNA replication in the S-phase of the cell by the CAF1 complex and are therefore replication-dependent. Less widespread variants of H3 are H3.3 and the centromere-specific isoform Centromere Protein A (CENP-A) (Buschbeck and Hake, 2017). H3.3 differs from the canonical versions only by a few amino acids but its distribution in the genome and functional studies reveal that these differences matter greatly. The canonical H3.1 and H3.2 are exchanged for H3.3 at transcriptionally active genes in a transcription-coupled manner, replication-independent (Ahmad and Henikoff, 2002). This incorporation into active chromatin is performed by the histone regulator A (HIRA) complex. Interestingly,

H3.3 is also enriched at repressed chromatin regions, such as telomeric, pericentric and repetitive regions. Incorporation into these regions is however, performed by a different chaperone complex, the Death Domain Associated Protein/Alpha Thalassemia/Mental Retardation Syndrome X-Linked (DAXX-ATRX) complex (Buschbeck and Hake, 2017; Goldberg et al., 2010). Hence, even a single histone variant can exhibit very different genomic localization and biological function.

1.3.3. ATP-dependent chromatin remodellers

Remodellers are enzymes that use the energy generated by the hydrolysis of ATP to perform a series of processes on chromatin. They can move, destabilize, eject, or restructure nucleosomes thereby regulating the accessibility of the DNA polymer. By doing this they can transiently expose individual sequences to interacting proteins (Figure 1.3.1). Often these are regulatory DNA elements (enhancers, promoters, replication origins) which in turn are important for a plethora of processes, such as gene transcription, DNA replication, DNA repair, and DNA recombination. During replication the DNA polymerase must have unhindered access to the DNA and after replication, nucleosomes must be deposited once more and spaced properly genome-wide, while still maintaining cell identity (Clapier and Cairns, 2009; Durand-Dubief et al., 2012; Muller and Leutz, 2001). When nucleosomes are tightly packaged and the DNA is wrapped around the nucleosome surface, regulatory factors and transcription factors might not be able to bind important cis regulatory elements (Figure 1.3.1). In addition, during transcription the RNA polymerase must not be impeded by nucleosomes (Clapier and Cairns, 2009). For DNA repair it is necessary that all base pairs are exposed to the repair machinery, and for DNA recombination it is necessary that long sequences are exposed in a synchronized manner for recombination to occur. Remodellers playing a role in this variety of processes does not mean they are solely there to enable them. Many remodellers have important roles in packaging chromatin more tightly, promoting a dense packaging of histones, in order to hinder the access of transcription factors or delay replication (Clapier and Cairns, 2009; Durand-Dubief et al., 2012; Muller and Leutz, 2001).

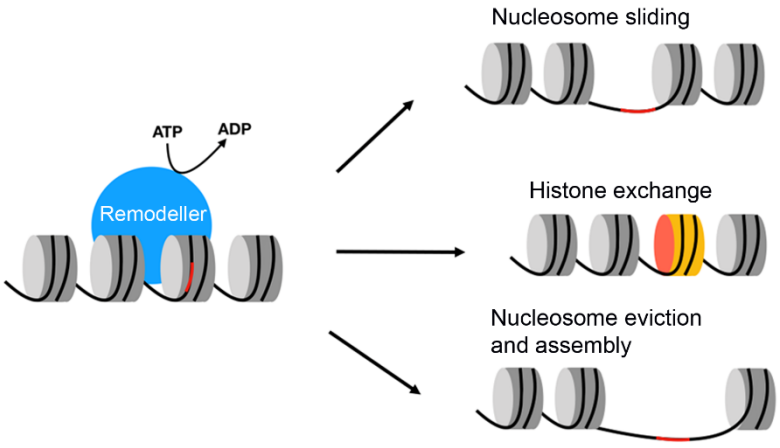

Figure 1.3.1: Processes of ATP-dependent chromatin remodelling.
The energy generated by remodellers from the hydrolysis of ATP to ADP is used to produce changes in chromatin compaction and DNA accessibility. This is achieved in a variety of functions; nucleosome deposition and spacing, eviction and nucleosome sliding, and the exchange of histone variants.

The role of remodellers in activation and repression of transcription affects the cell cycle progression (Muchardt and Yaniv, 2001), cell differentiation (Muller and Leutz, 2001), and therefore the development of multicellular organisms. Mutations or the deletion of many remodellers often has a severe effect on development. For example, homozygous null mutation of Imitation SWitch-1 (ISW1) in flies is lethal as a result of impaired cell viability (Deuring et al., 2000). Remodellers generally have a binding affinity for the nucleosome, apart from DNA-binding potential, they often have domains that recognize histone modifications, they possess a DNA-dependent ATPase domain that serves as a DNA-translocator, other domains or proteins that regulate the ATPase domain, and domains or proteins that regulate the interaction with other factors, such as transcription factors. They can be grouped into four different families according to sequence features outside of their ATPase domain and they often possess unique associated subunits depending on their specialization (Flaus et al., 2006). Altogether, this allows for a striking diversification of their functions, an example of which is the Nucleosome Remodeling Deacetylase (NuRD)-complex. NuRD contains several other subunits in addition to the ATPase Mi-2, such as a module consisting of the histone deacetylases HDAC1/2 and the H4 interacting proteins Retinoblastoma protein associated proteins-46/48 (RbAP46/48)

are also present (Kikyo et al., 2000; Knoepfler and Eisenman, 1999), so this complex combines ATP-dependent remodelling with the ability to covalently modify histone tails (deacetylation) on chromatin. Additionally, the methyl-cytosine binding protein MBD2 (Wade et al., 1999; Zhang et al., 1999) can associate with NuRD to form the MeCP1 complex. By this NuRD is targeted, remodels and deacetylates nucleosomes containing methylated DNA (Feng and Zhang, 2001). In addition, the two chromodomains that reside in Mi-2 appear to recognize DNA rather than methylated histone tails (Brehm et al., 2004). Thus, the remodeller integrates many processes of chromatin regulation.

For a long time, it was thought that remodellers scan along the chromatin fibre not unlike the repair machinery. This was because many remodellers themselves have no DNA binding capacity. It soon emerged, however, that this function is often provided by other proteins they interact with that do possess sequence-specific DNA-binding (Becker and Horz, 2002). For example, the corepressor KAP1 uses a bromodomain-PHD domain to target the NuRD complex to promoters for repression (Becker and Horz, 2002).

1.3.4. DNA methylation

DNA methylation is one of the main epigenetic mechanisms of repression. It involves converting cytosine residues by addition of a methyl group to 5'methylcytosine (5mC) on the DNA and this modification can be transmitted through cell division (Allis et al., 2015; Greenberg and Bourc'his, 2019; Smith and Meissner, 2013). It is present in most multicellular organisms, mainly at CpG dinucleotides or CpG "islands". DNA methylation is enriched at noncoding regions, such as centromeric heterochromatin, and on interspersed repeats, such as retrotransposons. On the other hand, it is low in 5' regulatory regions of genes (Bird, 1986) but is present in most exons and introns at high levels.

The effectors of DNA methylation are the DNA methyltransferases. They are often grouped into DNMTs for de novo and for maintenance DNA methylation. Maintenance methylation is performed by DNMT1 following replication (Song et al., 2011). DNMT1 is bound to unmethylated CpGs at replication via its CXXC-motif. When the enzyme encounters hemimethylation it is allosterically activated and adds methyl groups on the unmethylated sister strand. DNMT1 interacts with UHRF1/NP95, and this interaction enhances its stability on chromatin. Furthermore, NP95 interacts with H3K9me3 histone

modifications which gives rise to a functional connection between DNA methylation and repressive histone methylation (Du et al., 2015).

De novo DNA methylation is performed by the de novo DNA methylases DNMT3A and DNMT3B, which associate with the catalytically inactive DNMT3L. There is interplay between DNA methylation and histone modifications (Allis et al., 2015). For example, the active histone mark H3K4me3 inhibits DNMTL3 binding protecting CpG islands from DNA methylation (Cedar and Bergman, 2009). DNA methylation plays a variety of roles in epigenetic repression. In genomic imprinting, for example, the cause of allelic difference is DNA methylation on the maternal or paternal allele, including on the inactive X. It is also found in repetitive regions in the mammalian genome where DNA methylation is important to ensure genome stability, since many repeats have mutagenic potential if activated, which results in chromosomal abnormalities and is often seen in cancer progression.

1.3.5. Genomic imprinting

Mammals are diploid which means they have matched sets of two chromosomes in their cells, one from each parent. Therefore, they have two copies of every gene. Genomic imprinting is the phenomenon wherein one of these copies is not expressed. In humans only a few hundred of the approximately 25,000 genes in our genome show imprinting and this variation in expression is a consequence of parental inheritance (Reik and Walter, 2001). Genes that are imprinted, largely code for factors regulating embryonic and neonatal growth. Genomic imprinting is a cis-acting mechanism of gene repression and the imprints are acquired during gamete formation; thus, no type of recognition system is required to differentiate between the two alleles. Imprinted genes are usually arranged in clusters (Barlow and Bartolomei, 2014; Reik and Walter, 2001). In the mouse there are ~150 known imprinted genes on 17 chromosomes and roughly 80% of them are clustered into 16 genomic regions that contain at least two imprinted genes (Barlow and Bartolomei, 2014). Furthermore, imprints can act on multiple genes at once and these are mainly protein-coding genes, albeit lncRNAs are also represented in the clustered regions of imprints. These clusters are controlled by imprinting control regions (ICRs), and these regions are gametic differentially methylated, which means that they are established in one gamete and maintained on only one parental chromosome in the diploid cells of the embryo (Barlow and Bartolomei, 2014). These elements control the expression of the

cluster and DNA methylation is the initial gametic imprint that establishes the cluster and the whole imprinted region. Once formed in the gametes, the imprints need to be maintained and additionally, possess the capacity to escape genome-wide reprogramming after fertilization, which includes DNA demethylation in the preimplantation embryo (Barlow and Bartolomei, 2014; Reik and Walter, 2001).

1.3.6. X-inactivation

Non-coding RNAs (ncRNAs) play a crucial role in triggering epigenetic changes in chromatin dynamics. One example is X-inactivation in mammalian females. It involves the directed chromosome-wide silencing of one X chromosome during the development in female cells. It occurs to achieve dosage compensation between the sex chromosomes and the autosomes, so genes on the autosomes can function using the same dosage as they would in male cells. The inactive X can be seen localized to the nuclear periphery of the cell and is referred to as a Barr body (Allis et al., 2015).

In mice the large ncRNA Xist (~17 kb) is the primary trigger (Brockdorff et al., 1992; Brown et al., 1991) of the chromosome-wide repression on the X chromosome. Its antisense transcript *Tsix* is also involved in the initiation, it is however expressed only before the onset of X-inactivation and it appears that its role is not through RNAi-dependent mechanisms (Allis et al., 2015; Lee and Lu, 1999). The X-inactivation centre (XIC) and DNA docking sites are sites where the Xist RNA associates and functions as a scaffolding molecule decorating the entire inactive X chromosome in cis (Allis et al., 2015). The DNA docking sites are thought to be specialized repetitive DNA elements that are enriched on the X chromosome (Chow et al., 2010). The XIC region contains *Xist* and *Tsix*, in addition to several other ncRNAs that play both activating and repressive roles in X-inactivation. One example is the ncRNA activator *Rnf12* (Jonkers et al., 2009), another example is the *Jpx* ncRNA (Tian et al., 2010).

Chromatin modifications, PcG complex binding, subsequent incorporation of histone variant macroH2A and extensive DNA methylation all contribute to the formation of facultative heterochromatin along the entire inactive X (Allis et al., 2015). Once the repressive chromatin environment is established the Xist ncRNA is no longer required for maintenance (Gendrel and Heard, 2014; Morey and Avner, 2011).

1.4. Endogenous retroviral elements

Endogenous retroviral elements (ERVs) are one of the most fascinating types of transposable elements (TEs) that can shape the genome of the host they invade. TEs are DNA sequences that can change their position within the genome and due to their long evolutionary history in genomes and the fact that they continue to diversify, they come in a large variety of forms and shapes (Bourque et al., 2018). These mobile elements were first discovered in maize by Barbara McClintock in the 1950s. She had the remarkable foresight to recognize them as "controlling elements" and it was one of the first discoveries of non-mendelian inheritance (Schlesinger and Goff, 2015).

There are two major classes of TEs based on their mode of transposition, DNA transposons and retrotransposons, and each of these classes can be further divided into subclasses based on the mechanism of chromosomal integration (Figure 1.4.1) (Bourque et al., 2018). DNA transposons use a "cut-and-paste" mechanism of genomic integration via a DNA intermediate whereas retrotransposons employ a mechanism of "copy-and-paste" where an RNA intermediate is reverse transcribed into cDNA and integrated elsewhere in the genome (Boeke et al., 1985). Retrotransposons are then further divided into long terminal repeat (LTR) containing elements and non-LTR elements (Figure 1.4.1). TE subclasses are further divided into groups which can be found across different organisms. These elements share a common genetic organisation and a monophyletic origin (Jern et al., 2005; Malik and Eickbush, 2001). Finally, they are grouped into families of closely related elements which are all descendants of a single ancestral unit and share a consensus sequence which defines their characterisation and is representative of the entire family (Jurka and Smith, 1988; Smit, 1999). Focusing in more detail on specific elements, there are LINE (long interspersed elements) and SINE (short interspersed elements) repeats within the non-LTR containing retrotransposons and they make up around 35% of human and mice genomes (Figure 1.4.1). There are roughly 4800 full-length LINE copies in the mouse and 3000 are predicted to be active (Grabundzija et al., 2016). SINEs arose from accidental retrotransposition of Pol III transcripts and they rely on the machinery of LINEs for trans-acting transposition. In humans the Alu family is an active SINE family. In mouse several different families exist that are derived from tRNA and 7SL genes (Stocking and Kozak, 2008).

LTR-containing retrotransposons include endogenous retroviral elements (ERVs) which make up almost 10% of human and mouse genomes (Figure 1.4.1). They are however

much more active in the mouse where 10% of spontaneous mutations in inbred mice are due to the action of ERVs (Rowe and Trono, 2011). ERVs strongly resemble the exogenous retrovirus they evolved from. They generally encode a *gag* (group specific antigen) a *pro* (protease) and a *pol* (polymerase). Some few ERVs still encode envelope proteins (*env*) (Figure 1.4.1). ERVs have often accumulated several mutations and/or become truncated since their first integration into the hosts' genome.

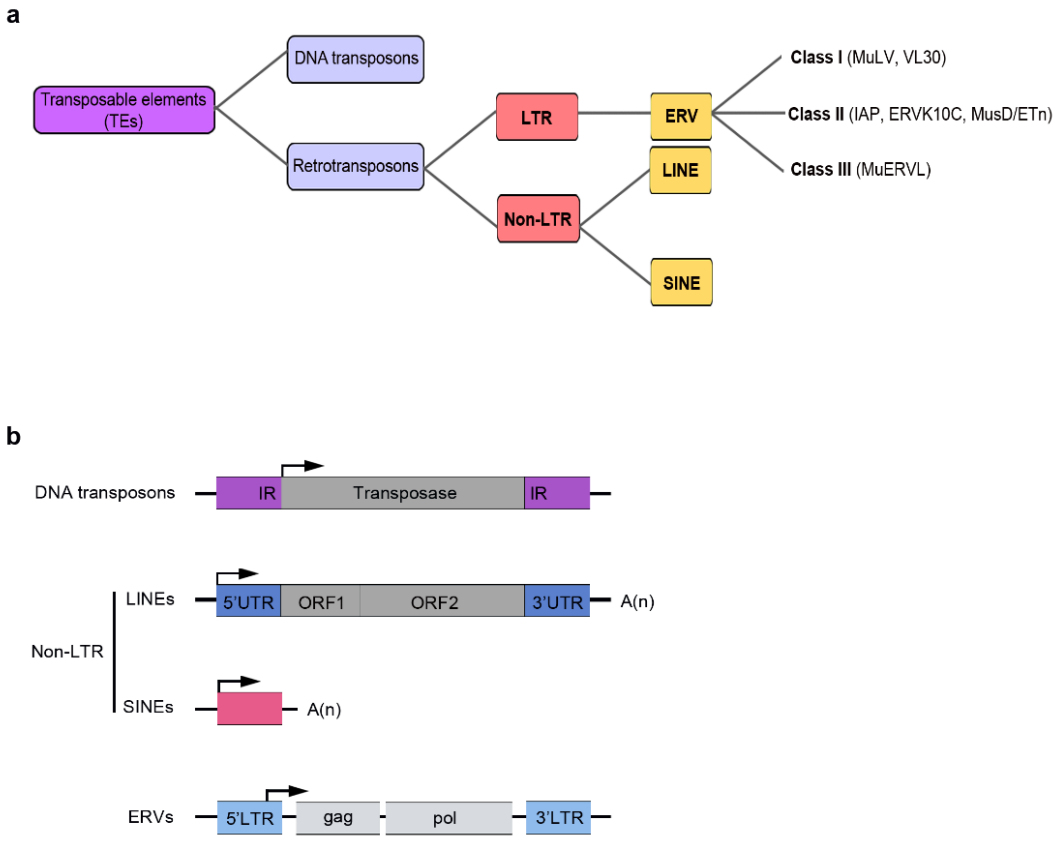

Figure 1.4.1: Classification of transposable elements in the mammalian genome.
(a) Classification of transposable elements depicting the major classes of elements endogenous to mammalian genomes. (b) Depiction of the main structural differences in transposable elements (DNA transposons; ERVs, endogenous retrovirus; LINEs, long interspersed elements; SINEs, short interspersed elements) in mammalian genomes. IR, inverted repeat; UTR, untranslated region; ORF, open reading frame; EN, endonuclease; RT, reverse transcriptase; LTR, long terminal repeat; ERV, endogenous retrovirus. This illustration is modified from (Martens et al., 2005a)

ERVs are generally grouped into classes according to their similarity to modern exogenous retroviruses, which have been classified into seven genera: alpha-, beta-, gamma-, delta, and epsilonretrovirus, lentivirus, and spumavirus. Using this system ERVs clustering with gamma- and epsilonretrovirus are termed class I, those that cluster with lentivirus, alpha-, beta- and deltaviruses are termed class II, and those that cluster with spumaviruses are termed class III. Intermediates between the classes have been identified suggesting an evolutionary continuum (Stocking and Kozak, 2008). Class III ERVs make up 5.4% of the mouse genome (Mouse Genome Sequencing et al., 2002). They are likely the most ancient ERVs, predating the human-mouse speciation (Benit et al., 1999; Mouse Genome Sequencing et al., 2002). Two types of transposon elements constitute the class III ERVs: murine ERV-L elements (MuERVL or MERVL) and the non-autonomous MaLRs. Despite its age this class has maintained some MuERVL elements in an active state in the mouse (Benit et al., 1999; Costas, 2003). Class II ERVs make up 3.2% of the mouse genome, which is a 10-fold higher proportion than in humans (Mouse Genome Sequencing et al., 2002). A large family within this class are MusD elements, with their deleted variants denoted ETn (early transposon (Brulet et al., 1983; Mager and Freeman, 2000). MusD elements lack an *env* gene. They are flanked by LTRs but contain mainly non-retroviral, non-coding sequences of unknown origin. Another well-characterised clade within the class II ERVs are intracisternal A-type particles (IAPs), which are present in mice at approximately 1000 copies per cell (Kuff and Lueders, 1988). Some of them still encode an *env* gene (Reuss, 1992; Ribet et al., 2008). They span several kb in length, generally up to 10kb, and display a number of subclasses and are actively transposing in the mouse. Intronic insertions are predominant for IAP elements and are denoted with the -int suffix (IAP-int, IAPEz-int). Solitary truncated IAP-LTRs are similarly common. The class I ERVs in the mouse include VL30 (virus-like 30) elements, which are 5-6 kb long DNA sequences which contain no open reading frames and no evidence for an *env* retroviral gene (Markopoulos et al., 2016). Thus, it is likely that they are non-autonomous in their retrotransposition.

1.4.1. Control of ERVs

ERVs are motors of evolution but at the same time they are threats to the stability of the genome. (Rowe and Trono, 2011). If ERVs are left unchecked within the genome they can cause harm by insertional mutagenesis and by erroneous non-allelic homologous

recombination if many copies of the element are present in the genome. This in turn leads to genome instability (Feschotte and Gilbert, 2012). Therefore, they need to be controlled tightly. The cell has many available pathways to silence ERVs in order to prevent harm. Among them DNA methylation, histone modifications and RNA-based mechanisms (Groh and Schotta, 2017), and these are all employed at different stages of development and in varying combinations in distinct cell types. For example, DNA methylation is the common mechanism in differentiated cells (Hutnick et al., 2010). In the dynamic chromatin environment of the developing embryo, it is important that repression withstands the demethylation wave in the cleavage embryo and ES cells are famously hypomethylated. In these contexts, histone modifications, particularly H3K9 methylation become very important for the repression of ERVS (Bulut-Karslioglu et al., 2014; Karimi et al., 2011; Rowe et al., 2010).

Mammalian genomes contain hundreds of ERVs and the transcriptional regulator KRAB-associated protein-1 (KAP1/TRIM28) is crucial for the repression of many different ERV types in both pluripotent and somatic cells (Ecco et al., 2016; Rowe et al., 2010; Tie et al., 2018). KAP1 is a major transcriptional regulator which also has non-transcriptional functions. KAP1 is ubiquitously expressed throughout development (Cammas et al., 2000) and it is a critical regulator of normal development and differentiation: mice deficient in KAP1 die before gastrulation (Cammas et al., 2000). KAP1 is also involved in maintaining pluripotency (Hu et al., 2009) and is required for the terminal differentiation of mES cells (Cammas et al., 2004; Cammas et al., 2002). The conditional genetic deletion of KAP1 in mES cells leads to overexpression of class I and II ERVs and class III to a lesser extent. Particularly IAP class II ERVs, are deregulated and the repressive histone marks H3K9me3 and H4K20me3 are lost (Rowe et al., 2010; Rowe et al., 2013). Additionally, IAPs are also derepressed in the epiblast embryo upon the loss of KAP1 (Rowe et al., 2010).

The best studied example of KAP1 repression is the Moloney murine leukaemia virus (MuLV) that has become endogenous in mice. In ES cells this viral sequence is silenced and this is dependent on the conserved sequence of the primer binding site (PBS) (Barklis et al., 1986). This is an 18-nucleotide sequence which is complementary to the 3’-end of the proline tRNA, which is the tRNA primer employed my MuLV for reverse transcription (Harada et al., 1979; Petersen et al., 1991). The PBS of MuLV and VL30, and many other class I ERVs, are bound by a specific Kruppel-associated box zinc finger

protein (KRAB-ZFP), the ZFP809 (Figure 1.4.2) (Wolf and Goff, 2009; Wolf et al., 2015). These are sequence-specific transcription factors that have C-terminal arrays of C2H2 (Cys2-His2) zinc fingers that bind specific DNA sequences (Urrutia, 2003). There are more than 400 human KRAB-ZFP genes that encode for more than 700 different proteins (Huntley et al., 2006) postulated to regulate all kind of different processes from embryonic development to cancer progression (Urrutia, 2003).

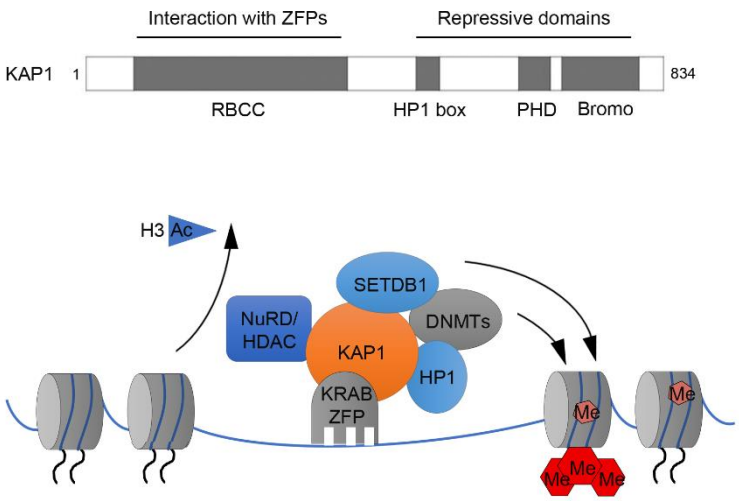

Figure 1.4.2: Recruitment of the KAP1-silencing complex.

ZFPs bind to the PBS of ERVs and recruit KAP1 which serves as a binding platform for repressive factors such as the histone methyltransferase SETDB1, DNA methyltransferases (DNMTs), histone deacetylases (HDACs) and HP1. SETDB1 and the DNMTs gives rise to H3K9me3 and DNA methylation whereas the HDACs remove histone acetylation. HP1 recruitment results in further compaction of chromatin resulting in ERV repression.

KRAB-ZFPs differ significantly between species and many are primate-specific and are thought to be involved in the regulation of the immune and nervous systems (Nowick et al., 2010). A key question is how KAP1 and other silencing proteins recognize and differentiate between different repetitive elements and mediate their finetuned expression control. One of the best studied strategies is performed by proteins that recognize specific DNA sequences and thus mediate specific targeting. Notably, KAP1 itself cannot bind DNA directly but different KRAB-ZFPs have evolved to recognize different repetitive elements. KAP1 is recruited to the chromatin by the interaction of its RING-B box-coiled

coil (RBCC) domain with KRAB-ZFPs. KAP1 has a HP1-binding domain by which HP1 is recruited, promoting heterochromatisation (Figure 1.4.2). It also has a PHD and bromodomain that interacts and recruits the histone methyltransferase SETDB1 which leads to H3K9me3 and, subsequently H4K20me3. HP1 in turn can bind H3K9me3, which reinforces the heterochromatin environment (Figure 1.4.2) (Iyengar and Farnham, 2011). KAP1 and the downstream silencing machinery discussed here is mainly employed in order to control class I and II ERVs, such as IAPs and VL30 elements.

LINE1 elements (non-LTR retrotransposons) similarly encompass many actively transposing copies in the mouse (Goke and Ng, 2016; Jacobs et al., 2014; Robbez-Masson and Rowe, 2015). The most ancient LINE1 families are neither KAP1-bound nor DNA methylated in mES cells, likely they have accumulated mutations and are no longer active. The new families (3.8-7.3 million years old), however, are KAP1-bound, repressed and highly methylated, depicting an evolutionary dynamic L1 regulation (Castro-Diaz et al., 2014; Robbez-Masson and Rowe, 2015). L1MdF and L1MdF2 are two examples of new L1 families in the mouse which are bound by KAP1 and de-repressed upon its removal (Castro-Diaz et al., 2014)

Class III displays a different expression pattern. These elements are highly upregulated in early stages after fertilisation in the oocyte and downregulated again at the morula and blastula stages. An example of this class is the MERVL element which is derepressed in cells lacking the histone demethylase LSD1 (Macfarlan et al., 2011a), Rex1 (Schoorlemmer et al., 2014) and the histone methyltransferase G9a (Leung et al., 2011), revealing a distinct mechanism than seen in class I and II ERVs. In the case of ERVs, the loss of KAP1 causes increased expression of the element in mES cells, but not in mouse fibroblasts (Cheng et al., 2014b; Iyengar and Farnham, 2011). This suggests that in some cell types, KAP1 may not be required for ERV repression or be redundant with other mechanisms.

Other mechanisms play a role in the silencing of ERVs. As mentioned, DNA methylation plays a major role particularly in differentiated cells. However, there are ATP-dependent remodellers that have been implicated in the repression of repetitive elements in mES cells. The depletion of ATRX has been linked to the upregulation of class II and III ERVs (He et al., 2015), however its role remains debated, as is explained in detail in the discussion. Chromodomain Helicase DNA Binding Protein 5 (CHD5) is another example (Hayashi et al., 2016) and the enzymatic activity of Helicase, Lymphoid Specific

(HELLS), a Snf2-like factor, is required for *de novo* DNA methylation of IAPs (Ren et al., 2015). However, this was shown during differentiation and not in pluripotent ES cells. A role of ATP-dependent remodelling has not been reported in the control of ERVs in mES cells.

1.4.1. The role of ERVs in gene regulation

Retrotransposons can negatively affect genome stability and disrupt DNA sequences due to transposition. Therefore, it is important that they are maintained in a repressed state. Many ERVs are however no longer active but even their strict regulation is required, not to limit retrotransposition, but due the role of ERVs in gene regulation. Retroviral elements have frequently been co-opted by the cell to control genes. Transcriptome profiling of mammalian stem cells has shown this in pluripotency maintenance and embryogenesis (Fort et al., 2014).

The exact mechanism by which ERVs, or truncated versions of them, are converted into regulatory sequences, remains obscure. However, they are uniquely suitably as cis-regulatory elements (Thompson et al., 2016). Their LTRs harbour the regulatory regions required for proviral transcription, including combinations of transcription factor binding sites (Thompson et al., 2016). And indeed, ERVs evolve faster than other TEs: ERVs in humans and chimpanzees have acquired signatures of directional evolution since the species split ~5 million years ago (Gemmell et al., 2015). There are now several examples known of ERVs as regulators of cellular genes. ERV promoters can drive the expression of tissue-specific genes, such as MERVL LTRs which serve as primary or alternative promoters of nearby genes (Macfarlan et al., 2011b), or an ERV that sits adjacent to the transcriptional start site of the mouse Sex-limited protein (*Slp*)-gene and dictates androgen-dependent regulation of this gene (Stavenhagen and Robins, 1988). Many of these ERVs functioning as tissue-specific promoters are unmethylated, bound by TFs and marked by H3K4me3, H3K27ac and H3K9ac (Bourque et al., 2008; Marino-Ramirez and Jordan, 2006). ERVs have further been shown to function as enhancers and they can even be integrated directly into existing genes increasing isoform diversity and influencing posttranslational regulation (Goke and Ng, 2016; Sorek et al., 2002; Zemojtel et al., 2007), but the adaptive consequences of this process for specific physiological functions remain largely unexplored.

1.5. KAP1 in mouse ES cells

Among the many functions KAP1 has in ES cells, apart from its role in ERV silencing, is the role it plays in both constitutive and facultative heterochromatin. Together with the chromatin remodeller SMARCAD1 (SWI/SNF-Related, Matrix-Associated Actin-Dependent Regulator of Chromatin, Subfamily A, Containing DEAD/H Box 1) it is involved in the maintenance of pericentric heterochromatin. In somatic cells SMARCAD1 and KAP1 bind to pericentric foci and are crucial to establish heterochromatin after replication (Rowbotham et al., 2011).

On *Zfp* genes KAP1 is required for heterochromatisation. ZFP274 recruits KAP1 to the 3' end of a *Zfp* gene (Frietze et al., 2010). Surprisingly, there is no correlation between the level of KAP1 at *Zfp* genes and the expression of that gene (Blahnik et al., 2011), thus it may not be transcription that is the target of KAP1-induced heterochromatisation. *Zfp* genes are highly homologous, having arisen from genomic duplications (Emerson and Thomas, 2009) and their 3'-coding exons encode tandemly arranged highly repetitive zinc finger domains. It has been suggested that KAP1 is involved in suppressing recombination at these sites, and interestingly binding of KAP1 positively correlates with the number of repeated zinc fingers within the *Zfp* 3'-exons (Blahnik et al., 2011).

KAP1 also plays a central role in genomic imprinting in the mouse, together with ZFP57. ZFP57 null mice show embryonic lethality and loss of imprinting at many, but not all loci. ZFP57 binds to co-factor KAP1 which recruits major repressive epigenetic regulators (SETDB1, HP1) (Li et al., 2008). Furthermore, its binding is sequence- and DNA methylation-specific and thus allows for allelic binding of KAP1 and further heterochromatisation potentially safeguards imprints from DNA demethylation waves (Quenneville et al., 2011; Strogantsev et al., 2015).

1.6. Chromatin remodelling by SMARCAD1 and roles in repression

The SWI/SNF-Related, Matrix-Associated Actin-Dependent Regulator of Chromatin, Subfamily A, Containing DEAD/H Box 1 (SMARCAD1) family of remodellers is one of the most conserved families. It includes Function Unknown Now-30 (Fun30) in *Saccharomyces cerevisiae*, Fission yeast Fun Thirty-3 (Fft3) in *Schizosaccharomyces pombe*, and SMARCAD1/Etl1 in mammals. The yeast protein has been biochemically well characterised and was shown to bind both DNA and chromatin and its ATPase

activity is similarly stimulated by both DNA and chromatin (Awad et al., 2010). Budding yeast Fun30 was shown to possess ATP-dependent chromatin remodelling activity in assays *in vitro* where its ability for histone dimer exchange was slightly higher than nucleosome repositioning (Awad et al., 2010). Based on sequence homology the Fun30 protein is most closely related to the Swr1 and Ino80 proteins (Flaus and Owen-Hughes, 2011) both of which show activity in histone exchange (Mizuguchi et al., 2004; Papamichos-Chronakis et al., 2006). Additionally, it was shown that Fun30 represses transcription by sliding promoter-proximal nucleosomes in an ATP-dependent manner (Byeon et al., 2013). And yet another study revealed a role of Fft3 in the disassembly of nucleosomes at transcribing regions to facilitate transcription elongation (Lee et al., 2017). Hence, the SMARCAD1 homologs in yeast have been implicated in several mechanisms of remodelling, ranging from histone dimer exchange, nucleosome repositioning to nucleosome disassembly.

When it comes to their function, SMARCAD1 homologs are important in heterochromatin maintenance. In *S. cerevisiae* Fun30 is required for the silencing of reporter genes embedded within transcriptionally repressed domains (Neves-Costa et al., 2009). Interestingly, its ATPase function is critical for repression since mutating an essential lysine residue in the ATPase domain abolishes this function (Neves-Costa et al., 2009). The SMARCAD1 homolog Fft3 in fission yeast is localized to centromeric insulators at the subtelomeric transition zone where it is similarly essential to maintain chromatin structure by inhibiting euchromatin assembly in silent domains (Stralfors et al., 2011). In fission yeast there is a well-known insulator element; two inverted repeats called IRCs that flank the border of the centromeres 1 and 3 (Cam et al., 2005). Fft3 binds these insulator elements and in its absence euchromatin marks, such as histone acetylation and the histone variant H2A.Z, invade centromeres and subtelomeres which results in misregulated gene expression and chromosome segregation defects (Stralfors et al., 2011). Fft3 is also localized to the LTR elements at the borders of subtelomeres in fission yeast (Steglich et al., 2015). LTR elements can function as insulators, as has been shown in the mouse (Carabana et al., 2011). When Fft3 is removed the nucleosome occupancy over the LTR elements is reduced at the subtelomeric border (Steglich et al., 2015). This suggests that Fft3 maintains chromatin structure, specifically of heterochromatic domains, by binding and remodelling nucleosomes at defining regulatory elements. Interestingly, this heterochromatic role of SMARCAD1 homologs is not restricted to Fft3. Fun30 shows a similar role in budding yeast even though the centromere structure

is vastly different. Budding yeast has defined point centromeres that are not embedded in heterochromatin. Similarly, to fission yeast, Fun30 is enriched at centromeres where us suppresses transcriptional noise, and it is required for normal nucleosome positioning and occupancy surrounding the centromeric nucleosome. Fun30 is involved in supporting faithful chromosome segregation through its role in determining centromeric and pericentromeric chromatin. And similar to Fft3, Fun30 is found binding over LTRs and telomeric repeats, among others (Durand-Dubief et al., 2012).

In human somatic cells SMARCAD1 acts together with the co-factor KAP1 to establish and maintain heterochromatin, specifically histone hypoacetylation and H3K9 methylation, during replication-coupled chromatin assembly (Rowbotham et al., 2011). SMARCAD1 is recruited to sites of DNA replication, interacts with HDAC1/2, G9a, PCNA and HP1, is required for KAP1 and HDAC1 binding to pericentric heterochromatin, and upon its loss there is a global reduction of H3K9me3 and HP1, and an increase of H3ac. These changes cannot be rescued by a SMARCAD1 ATPase mutant and they are linked to the progression through S-phase (Rowbotham et al., 2011). This illustrates that in somatic cells SMARCAD1 is required for the perpetuation of silenced loci since upon its loss the cells show mitotic defects and chromosome segregation is perturbed (Rowbotham et al., 2011). The cooperation of SMARCAD1 with KAP1 is not restricted to somatic cells. In both mouse ES cells and human somatic cells KAP1 is a stoichiometric partner of SMARCAD1 (Ding et al., 2018; Rowbotham et al., 2011). The interaction was mapped to the RBCC domain of KAP1 and the Coupling of Ubiquitin to ER degradation (CUE1) domain of SMARCAD1 and this interaction is required for the nuclear retention of SMARCAD1 (Ding et al., 2018). When it comes to the role of SMARCAD1 in ES cells it should be of note that SMARCAD1 levels are very high in the inner cell mass of the embryo at blastocyst stage (Schoor et al., 1993) and also in ES cells in culture when compared to somatic cells (Sachs et al., 2019; Xiao et al., 2017). Upon the loss of SMARCAD1 ES cells lose their typical morphology and show defects in the exit from self-renewal (Sachs et al., 2019; Xiao et al., 2017). In accordance with this observation the pluripotency marker Nanog and Oct4 are reduced upon the loss of SMARCAD1 as is the stem cell marker alkaline phosphatase (Sachs et al., 2019).

SMARCAD1 and its homologs thus play conserved and crucial functions in heterochromatin maintenance and assembly. Additionally, both Fun30 and SMARCAD1 promote long-range DNA end resection during repair of DNA double-strand breaks

(DSBs) (Chen et al., 2012; Costelloe et al., 2012; Eapen et al., 2012). DNA end resection is the initiation of the homologous recombination pathway upon DNA damage.

1.7. Objectives

The chromatin remodeller SMARCAD1 has emerged as an interesting factor in the context of pluripotency. The diverse functions of its orthologs point towards diverse potential roles of SMARCAD1 in the chromatin of ES cells. The objective of this study is to elucidate these functions with a focus on the genome-wide binding of SMARCAD1.

The first aim of this study is to establish a robust protocol for ChIP-seq with the endogenous protein SMARCAD1 as well as with stable FLAG-SMARCAD1 clones. The FLAG and the endogenous ChIP-seq were used to investigate the categories of SMARCAD1 binding sites genome-wide in pluripotent cells. In order to understand the function of SMARCAD1 in ES cells its binding sites were determined. Histone marks and other chromatin modifiers were investigated for enrichment on SMARCAD1 binding sites. The recruitment of SMARCAD1 to chromatin was another aim of this study, a KAP1-interaction mutant was employed to determine whether SMARCAD1 recruitment to different categories of its binding sites is dependent on its interaction with KAP1.

A loss-of-function approach was employed on subcategories of SMARCAD1 binding sites to determine the effect SMARCAD1 has on other co-factors involved in repression. The main focus was on the role of SMARCAD1 on repetitive regions such as endogenous retroviral elements. The function of SMARCAD1 was further elucidated on endogenous retroviruses by employing and ATPase mutant of the protein to determine whether its function on these sites is dependent on chromatin remodelling by observing the effect of the mutant on the binding of several repressive co-factors.

SMARCAD1 binding was investigated in differentiated ES cells since its protein interaction partners are very conserved in ES and somatic cells, as is its function in heterochromatin establishment and maintenance.

2. Materials and Methods

2.1. Tissue culture techniques

2.1.1. Cell lines and media

PGK12.1 female mouse ES cells were used to make the stable cell lines expressing FLAG-SMARCAD1-WT and FLAG-SMARCAD1-ATPase mutant and their respective empty vector control for ChIP-seq. A stable SMARCAD1 knockdown of this cell line was also used for ChIP-seq and for ES cell differentiation.

E14 male mouse ES cells containing various SMARCAD1 constructs were used for a 2-day and 4-day inducible knockdown of SMARCAD1. E14 cells were also used for a 3-day transient knockdown of KAP1. A stable knockdown of SMARCAD1 in this cell line was used in ChIP experiments.

All three mouse ES cell lines and controls were routinely cultured in feeder-free conditions on gelatin-coated cell culture flasks, in the LIF+serum medium. Cells were routinely tested using MycoAlert^{TM} Mycoplasma Detection Kit (Lonza Bioscience, cat. no. LT07-418) to ensure they were free of mycoplasma.

ES cell growth medium: Leukaemia Inhibitory Factor (LIF) + serum

Reagent	Volume
DMEM (Gibco 41965-039)	500 mL
Fetal Bovine Serum (Batch tested for mouse ES cell in the laboratory)	75 mL
L-Glutamine 200 mM (Gibco 25030-024)	5 mL
MEM Non-Essential Amino Acids (100×) (Gibco 11140-035)	5 mL
β-Mercaptoethanol (50 mM) (Gibco 31350-010)	570 µL
LIF, mouse recombinant (10^7 units/mL) (Cell GS GFM200)	57.5 µL
Penicillin/streptomycin (pen/strep) (Gibco 15140-122) (optional)	5 mL

LIF serum-free medium

The serum-free medium was used in transfection experiments. All the reagents besides serum were mixed as making the LIF+serum medium. An aliquot (50 or 100 mL) from the mix was taken as the serum-free medium.

EC10 medium (for ES cell differentiation)

Reagent	Volume
DMEM (Gibco 41965-039)	500 mL
Fetal Bovine Serum (Batch tested for mouse ES cell in the laboratory)	75 mL
L-Glutamine 200 mM (Gibco 25030-024)	5 mL
MEM Non-Essential Amino Acids (100×) (Gibco 11140-035)	5 mL
β-Mercaptoethanol (50 mM) (Gibco 31350-010)	570 µL
Pen/strep (Gibco 15140-122) (optional)	5 mL

Medium for mouse embryonic fibroblasts (MEFs)

Reagent	Volume
DMEM high glucose (4.5g/L) (Gibco 11965-092)	250 mL
Ham's F-10 (Gibco Thermo Fisher Scientific)	250 mL
Fetal Bovine Serum (Batch tested for mouse ES cell in the laboratory)	100 mL
Pen/strep (Gibco 15140-122) (optional)	5 mL

2.1.2. Gelatin coating

ES cells were grown feeder-free in cell culture vessels. This was achieved by coating the vessels with 0.1% gelatin (Sigma, cat. no. G1890-100G) in Phosphate Buffered Saline (PBS). The gelatin was applied for a minimum of 5 minutes at room temperature and the remaining liquid removed shortly before use via aspiration.

2.1.3. Growing and splitting cells

ES cells were fed with fresh medium every 24 hours and routinely passaged at confluence 90%. Usually, ES cells were split 1:4 to 1:6 every 48 hours. For passaging, medium was aspirated off, cells washed in 1xPBS and detached using 0.05% trypsin-EDTA (Gibco, cat. no. 25300-054). The cells were dissociated at toom temperature for 3 minutes. Trypsin was inactivated with fresh medium containing serum and the cells were transferred into new containers.

2.1.4. Freezing cells

Cells were typsinised as described for splitting and the trypsin inactivated with freezing medium (FCS with 10% DMSO). The suspension was carefully mixed by minimal pipetting and divided into cryogenic tubes. For this, 1.5mL were transferred into 1.8mL cryotubes and transferred to a freezing container (Mr. Frosty™) filled with isopropanol. Usually, mouse ES cells from a cell culture vessel with 80% confluence were frozen down into two vials, each of which could be recovered in a vessel of the original size. The containers were immediately placed at -80 °C and let to cool down overnight until the cells in the cryogenic tubes were frozen. For long-term preservation, they were then placed in liquid nitrogen containers.

2.1.5. Thawing cells

Cells frozen in cryogenic tubes were thawed in a warm water bath until defrosted. They were then gently transferred into 15 mL falcon tubes containing 10 mL of fresh growth medium. The cells were pelleted by centrifugation (300 g x 3 minutes) and the medium was removed by aspiration. The cells were then resuspended in a small amount of medium and plated into pre-gelatin coated cell culture vessels containing the appropriate amount of medium

2.1.6. Stable cell lines

Cells	Constructs	Treatments
PGK12.1	FLAG-empty vector control	1µg/mL puromycin
	FLAG-SMARCAD1-WT	1µg/mL puromycin
	FLAG-SMARCAD1-ATPase mt	1µg/mL puromycin
	Stable SMARCAD1 kd	1µg/mL puromycin
	Control kd	1µg/mL puromycin
E14	Stable SMARCAD1 kd	1µg/mL puromycin
	Control kd	1µg/mL puromycin

2.1.7. Eccentricity assay with IncuCyte real live imaging

For observing cell growth and stemness in real time cells were plated in triplicates at dilutions ranging from 10 000 cells to 312 cells per well on 96 well tissue culture plates and incubated at 37 °C. Photomicrographs were captured every three hours using an IncuCyte cell live imager (Essen BioScience) and the IncuCyte ZOOM's Confluence Processing analysis tool was used to calculate cell confluence metrics in real-time, specifically the eccentricity of the cells defined as the average of how round or compact the objects are ranging from 0 to 1 with a perfect circle having a value of 0.

2.1.8. ESC differentiation

PGK12. ES cells were plated in non-gelatinised 10 cm^2 cell culture vessels, 10-15 plates per cell line. Cells were trypsinised and the reaction stopped with EC10 medium. They were counted with the Casy Model TTC (Roche) and $8x10^5$ cells were plated on each vessel in 10 mL EC10 medium. The cells were left to aggregate for three days and the EC10 medium changed daily. On the third day they were blown off using a pipette boy and 10 mL pipette tips. Without gelatin they are easily detachable and were collected in 50 mL falcons. The cells from 5 plates were collected in one 50 mL falcon. The cells were left to settle without centrifuging and the medium gently aspirated. They were transferred to sterile 10 cm^2 bacterial plates in 10 mL EC10 medium and left to grow in suspension where they form embryoid bodies. The EC10 medium was changed daily by transferring the cells in suspension to 50 mL falcons, aspirating the medium off after the cells had settled and transferring them into fresh medium. Importantly, the cells were not centrifuged since this tends to break apart the embryoid bodies and differentiation is not successful. After five days of differentiation the cells were collected and used in a ChIP experiment in suspension.

2.1.9. Cell pellets

Cells were trypsinised as described and neutralised in medium containing serum. They were transferred into tubes and kept on ice. The cells were pelleted by centrifugation (2000 x g, 3 min, 4 °C) and the medium was removed. The cell pellet was resuspended in cold PBS supplemented with 0.1 mM PMSF and 0.1 mM benzamidine and pelleted again

by centrifugation. The PBS was removed and the washing step repeated. Afterwards, the cell pellet was snap-frozen in liquid nitrogen and stored at -80 °C.

2.1.10. Transfection and stable clones

Stable FLAG-SMARCAD1 and control cell lines were generated encoding 3xFLAG-SMARCAD1, 3xFLAG-ATPase mutant and a non-target control containing only 3xFLAG.

PGK12.1 ES cells were transfected using Lipofectamine 2000 (Life Technologies) following manufacturer's instructions. A GFP-vector was used as a control for the transfection (see plasmid list/plasmid maps). $1x10^6$ cells were plated per well in a 6-well tissue culture dish on the day of the transfection. The transfection was executed in suspension, per transfection 4µg of purified plasmid DNA diluted in 250µl serum-free DMEM was mixed with 8µl Lipofectamine 2000 diluted in 250µl serum-free DMEM. The mixed solution was incubated at room temperature for 20 minutes and then added to the ES cells and the cells were placed in a humidified CO_2 incubator at 37 °C. After 5 hours, the medium was replaced with fresh ES cell medium containing serum. Twenty-four hours after transfection the cells were split in different dilutions into four 100 mm dishes. Antibiotic selection with 1.7 µg/ml puromycin to generate stable cell lines was started 48 hours after transfection and the cells were kept under selection for 10-13 days. Untransfected cells were treated the same way and were used as a selection control. After selection, the cells were picked as colonies or grown as pool. The obtained cell lines were expanded and analyzed by Western blot and indirect immunofluorescence.

2.1.11. Colony picking

Each 100 mm dish with colonies to be picked was washed with PBS and 3ml PBS were added to the plate to stop the cells from becoming too dry. Individual colonies were picked and transferred to 25µl trypsin diluted ½ in 1xPBS in a well of a 96 well plate. Once the entire plate was picked it was incubated for 10-15 min in a CO_2 incubator at 37 °C. Cells were resuspended and transferred to a pre-prepared 96 well plate with fresh medium. The plate was incubated at 37 °C and the media changed every day.

2.1.12. Cloning out using Li-Cor In-Cell Western™

In order to increase the level of expression of the FLAG-SMARCAD1-ATPase mutant construct the stable cell line was cloned out using 96-well plates and In-Cell Western by Li-Cor for fast processing. Initially, a T75 of clone 8 was diluted to $5x10^4$ cells/mL and 200 µL of this dilution transferred to each 96-well. The empty vector cell line as well as the FLAG-SMARCAD1-WT cell line were used as negative and positive controls, respectively. The levels of construct expression were determined via FLAG detection according to the manufacturer's protocol. Briefly, the medium was aspirated off and the cells in the wells were fixed and permeabilized as described for the FLAG antibody in the section on immunofluorescence. A multichannel pipette was used for all steps and the cells are not left to dry out at any point in the protocol until imaging. 150 µL of Odyssey™ blocking buffer is used after fixation and permeabilization for 1.5 hours at room temperature. Afterwards, the primary antibody is diluted in blocking buffer as described (2.9 key resource tables) and 50 µL are applied for two hours at room temperature on a shaking platform. Wells with no primary antibody are used in every experiment as negative control. The wells are washed thrice with PBS and 50 µL of the secondary antibody are applied (2.9 key resource tables) for one hour at room temperature on a shaking platform while the plate is kept protected from light. The cells are washed again thrice with PBS and the last wash is removed completely. The FLAG expression was determined with Li-Cor In-Cell Western imaging as described by the manufacturer with the wavelength setting appropriate for the secondary antibody that was employed. The highest expressing well was selected and expanded to a T75 and the procedure was repeated until a robust expression of the FLAG-construct was achieved (in comparison to the WT construct)

2.1.13. Transient knockdowns (KAP1 and SETDB1)

Transient knockdown of KAP1 and SETDB1 was performed in order to carry out ChIP experiments. E14 WT ES cells were employed for both transient transfections. The cells of the transient KAP1 knockdown were collected three days after the initial transfection. The cells in the transient SETDB1 knockdown were collected four days after transfection and a re-transfection was performed at day two. The transfection protocol was upscaled from the protocol described in the generation of stable FLAG-SMARCAD1 clones. Briefly, the transfection is carried out in 15 cm^2 cell culture dishes with $20x10^6$ cells per

dish in 12 mL medium. The shRNA is upscaled to 64 µg per dish and prepared in 4 mL serum-free medium as are the 128 µL of lipofectamine. After complex formation the 8 mL serum-free medium containing the complexes are added drop by drop onto the cells and left for five hours as described in the transfection subheading. Puromycin selection is started the next day at 1.4 µg/mL. For ChIP the cells of both transfections were plated 24 hours before crosslinking as described in the protocol for ChIP with adherent cells.

2.2. Chromatin Immunoprecipitation

2.2.1. ChIP in adherent cells

Chromatin immunoprecipitation was used to identify the genome-wide binding sites of chromatin factors (ChIP-seq) and to validate the ChIP-seq results (ChIP-qPCR). ChIP assays were performed using the One Day ChIP kit (Diagenode). $25x10^6$ embryonic stem cells were plated in 150 mm dishes 24 hours prior to crosslinking. This resulted in ~50 million cells per plate for crosslinking. Preceding the fixation step, cells were washed with three times with 1xPBS. Ten ml of 1xPBS was added per dish and Disuccinimidylglutarate (DSG) freshly resuspended in DMSO was added to a final concentration of 2 mM. The cells were fixed for 45 minutes at room temperature on a shaking platform. Following the fixation, the cells were washed three times with 1xPBS and a second crosslinking step was performed with 1% Formaldehyde in PBS for 10 minutes at room temperature. The crosslinking was stopped by the addition of glycine to a final concentration of 0.125 M for 5 minutes at room temperature. The cells were washed twice with ice-cold 1xPBS, scraped off in wash buffer B (10 mM HEPES/KOH pH 6.5, 10 mM EDTA, 0.5 mM EGTA, 0.25% Triton X, protease inhibitors PMSF and Benzamidine) and collected in 15 ml conical Falcon tubes. From here on everything was performed on ice. The cells were then pelleted by centrifugation at 2000 rpm at 4 ºC, resuspended in wash buffer C (10 mM HEPES/KOH pH 6.5, 10 mM EDTA, 0.5 mM EGTA, 200 mM NaCl and protease inhibitors PMSF and Benzamidine) and incubated 10 minutes on ice. They were again pelleted by centrifugation at 2000 rpm at 4 ºC and resuspended in lysis buffer D (50 mM Tris/HCl pH 8.0, 10 mM EDTA, 1% SDS, Protease inhibitor cocktail 200x from Diagenode)

The samples were incubated for 20 minutes on ice before sonication. Sonication was performed using a Bioruptor sonicator (Diagenode) to generate fragments between 200-600 basepairs (bps). Typically, 20-25 sonication cycles of 30 seconds on and 30 seconds off were performed for ES cells. The sonicated chromatin was centrifuged for 10 minutes at 13000 rpm at 4 ºC in 1.5ml Eppendorfs. The supernatant was transferred into a new tube and the chromatin concentration was determined using a Nanodrop.

Fragment sizes were analysed by taking a 6 µl aliquot, incubating it for 30 minutes with RNAse at room-temperature, after which 8µl 5M NaCl were added o/n at 65 °C. Afterwards, Proteinase K treatment was performed for 1h and 50 minutes at 45 °C. The samples were then Phe/CHCl3 extracted and the upper phase was run on a 1.2 % agarose gel. If fragment sizes were lower than 1 kb the procedure was continued. The sheared chromatin was stored at -80 °C.

43 µl protein A beads (OneDay ChIP kit Diagenode) were used per ChIP, corresponding to ~50 million cells. They were washed twice in ChIP buffer (OneDay ChIP kit) before immunoprecipitation. Chromatin corresponding to 100 µg of DNA was used for each ChIP. The sonicated chromatin was diluted ~1:5 in ChIP buffer. Samples were incubated with 3 µg of antibody for 2-3 hours on a rotating wheel at 4 °C. A 10-minute centrifugation step at 13,000 rpm was performed and the antibody-bound chromatin was collected and added to the protein A beads. Samples were incubated for 1 hour on a rotating wheel at 4 °C and subsequently washed twice with ChIP buffer from the OneDay ChIP kit. To elute the bound DNA the beads were resuspended in DNA purifying slurry (OneDay ChIP kit) and boiled for 10 min. The samples were incubated for 30 min with 1 µl Proteinase K on a ThermoMix at 550 rpm and subsequently boiled at 95 °C for 10 minutes to inactivate the proteinase K. The samples were centrifuged at 13,000 rpm and the DNA collected in 200 µl PCR-clean water.

2.2.2. ChIP in solution

Cells were plated 24 hours before crosslinking as described in the ChIP protocol for adherent cells. They were trypsinised immediately before crosslinking and collected in a 50 mL falcon. For ES cell differentiations the cells were collected directly from the bacterial dishes without trypsinization. The DSG was prepared as described in the ChIP

protocol on adherent cells, and the cells washed thrice with PBS. 10 mL of DSG-PBS was added for each dish (~40-50x10^6 cells) to the 50 mL falcon. For ESC differentiation, 10 mL DSG-PBS was added for four bacterial dishes. The falcons were rolled for 45 minutes at room temperature, the DSG-PBS was removed after brief centrifugation at 1,000 rpm, and the cells washed thrice in PBS. 10 mL PBS was added for each dish of cells collected initially (e.g., 30 mL for three dishes). Formaldehyde was added to a final concentration of 1% in PBS and the falcons were rolled for 10 minutes at room temperature. The reaction was stopped with glycine at a final concentration of 125 mM for 5 minutes at room temperature. The crosslinked cells were washed twice in cold PBS and resuspended in cold buffer C from the OneDay ChIP kit (Diagenode) and the experiment was carried out as described for ChIP in adherent cells.

2.2.3. Re-ChIP

Sequential ChIP was performed using SMARCAD1 (3 µg) or KAP1 (3 µg) antibody in the first round as described in the ChIP protocol, precipitated material was eluted twice from the beads with elution buffer containing 10 mM DTT, 100 mM NaCL, and 1% SDS at 37 °C for 30 min. The eluted DNA was diluted 40-fold with ChIP buffer, and a second round of immunoprecipitations was performed in accordance with the OneDay ChIP kit (Diagenode) manual, performing an overnight antibody (6 µg) incubation at 4 °C of SMARCAD1, KAP1 and IgG antibody.

2.2.4. Real-time qPCR after ChIP

qPCR analysis was performed on DNA obtained from ChIP experiments. 4µl DNA was added together with 1x iQ SYBR Green supermix (Bio-Rad) and 0.5µM forward and reverse primers. A sample with water instead of DNA was used to rule out primer dimer formation and other contamination. Primers are listed in the key resources table.

qPCR conditions were as follows: 10 minutes at 95 °C followed by 40 cycles at 95 °C for 15 seconds and 60 °C for 30 seconds, followed by a plate read after each cycle. Melting curve test was performed at the end of each experiment (from 55 °C to 95 °C, read plate every 0.5 °C) to ensure the specificity of amplification. Each primer was executed in triplicate.

qPCR was executed on a CFX96 Connect (Bio-Rad) or an Agilent MX3000P. Enrichment values for specific factors were normalized to input and plotted either as % input using the standard error of technical replicates or as fold change over H3 (for histone modifications), using error propagation:

$$Error\ propagation\ of\ SE =$$

$$(Fold\ change\ over\ H3 \times \sqrt{(SE \div percent\ input)^2 + (SE\ of\ H3 \div percent\ input\ H3)^2}).$$

2.2.5. Primer Design and testing

Primers were designed with PrimerBlast using mouse genome sequences from the UCSC Genome browser and *in silico* PCR was used to test initial specificity. Primers were designed to be 180-210 bp, with melting temperatures of 58-62 C°, 20-28 bp long, with GC content of 20-80 %, and with no more than 3 out of 5 bases at 3' end with G or C. For primers amplifying repetitive sequences the automatic repeat filter was removed. qPCR primers were tested on genomic DNA and resolved by gel electrophoresis. Each primer pair was tested for efficiency and quantification with serial 10-fold dilutions of genomic DNA. Only primers with a defined melting curve and good efficiency (80-120%) were used in subsequent experiments. Primers that yielded unspecific bands on the agarose gel in the no-template-control were excluded, as well as primers that gave products of different sizes.

2.3. ChIP-seq

For each ChIP-seq, up to seven ChIP samples were pooled during DNA purification in order to have enough material for library preparation and amplification. Otherwise, the ChIP was carried out as described. During DNA purification the chromatin was eluted from the antibody binding beads (OneDay ChIP kit Diagenode) using 230 µL elution buffer (100 mM $NaHCO_3$,1% SDS, freshly prepared and sterile filtered). The beads were incubated for 30 minutes at room temperature while shaking on a ThermoMix (Eppendorf). They were centrifuged at 13,000 rpm for 1 minute and 200 µL of the supernatant was collected. The beads were discarded. The chromatin was de-crosslinked overnight with 8 µL NaCl at 65 °C. 8 µL 1M Tris/HCl pH 6.5, 4 µL 0,5 M EDTA and 2 µL Proteinase K (10 µg/µL) were added for 1 hour at 45 °C to digest the proteins. For

DNA purification the QIAquick PCR Purification Kit (Qiagen) was used according to the manufacturer's instructions and all the samples pertaining to the same antibody were pooled via the same DNA binding column and eluted from the column with 40 µl sterile 2 mM Tris/Cl, pH8.5. Of each sample 1 µL was measured using the Quant IT kit (Invitrogen) on the FluoroNanoDrop and the samples stored at -20 °C.

2.3.1. Library preparation and sequencing

Library preparation was performed using the MicroPlex Library Preparation Kit v2 (Diagenode) according to manufacturer's instructions and all reagents used were made freshly and sterile filtered. The bench and all instruments used were cleaned with bleach. Briefly: 4-8 ng DNA was used for each sample. The MicroPlex Library Preparation kit provides adapters to sequence and identify different samples. This adapter-ligated DNA was amplified on a thermal cycler with up to 9 cycles of PCR amplification before size selection and DNA purification with AMPure XP beads (Agencourt). Size and concentration of the DNA fragments was assessed on a Bioanalyzer (Agilent Technologies) by the sequencing facility. Sequencing was performed in the Genomics Core Facility of the Philipps-Universität Marburg on the Illumina 1500 platform with on-board cluster generation using the HiSeq Rapid SR Cluster Kit v2 (Illumina) and single read 50 nucleotide sequencing on a HiSeq Rapid SR Flow Cell v2 (Illumina).

2.3.2. Data analysis

Data analysis was performed by Boris Lamp. Reads from the sequencing were aligned to the Mus Musculus genome version mm10 retrieved from Ensembl revision 83 using Bowtie 2.0.0-beta7 (Langmead and Salzberg, 2012), and with the default parameter settings. After all reads were aligned to sequences in the genome the laned were de-duplicated to the expected number of duplicate reads for a given region based on binomial distribution. Only these reads were kept for further analysis. Peak calling was performed for each ChIP-seq separately and for each antibody an IgG antibody or input or both was used as control. The MACS program v1.4.0rc2 (Zhang et al., 2008) was used for all samples, except for the histone modification mark H3K9me3. H3K9me3 ChIP peaks were called using SICER 1.1 (Zang et al., 2009) with these parameters: windows size 200, gap size 200, fragment size 51, mappability percentage 0.78. Only those peaks were

kept that were not present in the backgrounds (IgG or input, or if present, both). Peak filtering was used in order to reduce false positive peaks. This was achieved by identifying only those peaks that show a strong enrichment over their backgrounds. SMARCAD1 and KAP1 peaks were only retained if they had a minimum of 30 effective foreground reads, not more than 50 effective reads in either of their background(s), and showing at least a 2.5-fold increase in the normalized read counts (TPM) compared to their background(s). For H3K9me3 only peaks with a minimum increase of three-fold in the normalized read counts compared to either of their backgrounds were kept. To permit comparison between the samples, tag counts were calculated and normalized to one million mapped reads (TPM, tag per million). The foreground-background ratio employed to filter reported peaks was determined on basis of TPMs in foreground versus those in the background. If more than one background value was available the ratio was calculated using the maximum background TPM value from either input or IgG antibody. To determine the genomic distribution of SMARCAD1 in the murine genome of stem cells the overlapping peaks from both the endogenous and the FLAG-ChIP-seq were taken into consideration if their left most position overlaps with defined genomic elements. The percentage of overlap was calculated and plotted. The annotations were taken from the Ensembl Genome database (Mus musculus, Rev. 83, mm10). Heatmaps were generated to compare the enrichment of the signal of the most pronounced peaks from SMARCAD1-FLAG and from the endogenous ChIP-seq. To each other but also to other chromatin factors. Regions spanning 1kb around the 2380 peaks (overlap of the two ChIP-seqs) were centred to the summit of the SMARCAD1-WT signal (TPM). For the different heatmaps to be comparable, all signals were normalized to TPM with the 98th percentile of the SMARCAD1-wt signal set as the maximum value. Darker colors indicate higher signal intensities. The expression heatmap was generated in Excel (Microsoft) using the function of Conditional Formatting on the log2 fold changes. Publicly available external ChIP-Seq data was obtained from GEO using the following accession numbers: GSM1555120 (KAP1), GSM1429923 (KAP1 input), GSM307622 (H4K20me3), GSM1033638 (H3K27me3), GSM594578 (H3K27ac), GSM1033636 (H3K4me3), GSM1555116 (H3.3), GSM459273 (SETDB1), GSM1215219 (G9a), GSM1375157 (SUV39H1), GSM1375158 (SUV39H2). For all screenshots depicted was well as for the manual inspection of the sequenced reads and peak distribution the Integrative Genomics Viewer (IGV) (Robinson et al., 2011) was used by myself and other members in the lab. Area proportional Venn Diagrams displaying overlap between peaks

from different datasets were generated by Philipp Bergmaier using eulerAPE (Micallef and Rodgers, 2014).

2.3.3. Analysis of repetitive elements

For the analysis of SMARCAD1 peaks in repetitive elements the method from (Elsasser et al., 2015) was employed by Boris Lamp. Briefly, for the alignment the annotated repetitive regions from Repeatmasker were downloaded from UCSC Table Browser (http:\\genome.ucsc.edu) for mm10 on April 17th, 2017. The mm10 genome build was filtered to these regions to generate repetitive element sequences. All repeat sequences were expanded at their beginning and their end by one average read length (51 bp) in order to compensate for repeat elements that might be too short for correct read alignment. A bowtie index was created using bowtie-build and ChIP-seq reads were aligned against these repeat sequences using Bowtie (parameters: -k 1 and --best) (Langmead et al., 2009). For the analysis of these aligned reads normalized read counts (TPM) were calculated for each repetitive element. Log2 fold-change over input was calculated for all repeat classes to identify elements enriched for SMARCAD1. The boxplot depicts the log2 fold-changes over input for all regions belonging to each of the custom repeat categories and the repeat class with the highest (averaged) fold-change. The box- and jitterplot depicts the log2 fold-changes over the corresponding input for all repeat regions belonging to each of the custom repeat categories.

2.4. Protein work

2.4.1. Total protein extraction

Frozen cell pellets were thawed quickly and immediately re-suspended in 3-4x pellet volume of the lysis/digestion mix (20 mM HEPES [pH 7.3], 10 mM KOAc, 5 mM NaOAc, 2 mM MgOAc, 1 mM EGTA, 2 mM DTT, 0.1% NP-40, 10 mM $MnCl_2$, 20 µg/mL DNase I). They were subsequently incubated in the water bath for 30 minutes at 37 °C and the tubes were regularly flicked during incubation to disperse precipitates. Protein concentrations were determined by Bradford assay (BioRad, cat. no. 500-0006) following the manufacturer's instructions.

2.4.2. Whole-cell lysate

Protein extracts were lysed in 6x laemmli buffer (50 mM Tris-HCl [pH 6.8], 10% Glycerol, 1% SDS, 100 mM DTT) and boiled at 95 °C for 5 minutes. A sample of each protein extract was prepared at 1µg/µl with 1x laemmli for immediate use and the corresponding sample with a higher concentration was stored at -20 °C. The whole-cell lysates were used for Western blot.

2.4.3. Western blot

Western blot was performed following standard procedures. Total protein extract or whole-cell lysate was prepared in 1x laemmli buffer, separated by SDS-PAGE and transferred to either nitrocellulose (NC) or polyvinylidene difluoride (PVDF) membranes. The blots were blocked in 5% milk in PBS for upwards of two hours and subsequently probed with primary antibodies (see primary antibody list), washed in PBS, followed by an incubation with a secondary antibody. The bound antibodies were detected using Immobilon Western chemiluminescent HRP substrate (Millipore, cat. no. WBKLS0500) using X-ray films or ChemiDoc (Bio-Rad) for visualization. Quantification of the signal was performed with Image Lab (BioRad) or with Image J for scanned films. Lamin B1 was used as a loading control and for normalization of the detected signal.

2.4.4. Stripping of Western blots

This study used two methods of stripping. Generally, a technique of “gentle” stripping was employed using the Restore Plus Western Blot Stripping Buffer (Thermo Scientific, cat. no. 46430) following the manufacturer’s instructions. The success of this method was then tested by incubation with the respective secondary antibody and if the removal of the primary antibody was not successful with the commercial buffer, the membrane was incubated in 50 mL of “harsh” stripping buffer at 50 °C for 30 minutes

Harsh stripping buffer (50 mL)

Reagent	Volume
Tris HCl (pH 6.8, 1 M)	3.1 mL
SDS 10%	10 mL
MilliQ water	36.5 mL
β-mercaptoethanol	450 µL

2.5. RNA work

2.5.1. Total RNA extraction

Fresh cells growing in culture were lysed in an appropriate volume of TRIzol reagent (Invitrogen, cat. no. 15596018), for example a 6-well dish of both mouse ES cells and after 5 days of differentiation were lysed in 800-1000 µl TRIzol. Following the manufacturer's instructions, total RNA was isolated from the TRIzol lysate. Chloroform was added to the lysate (0.2mL per 1 mL of TRIzol) and incubated at room temperature for 5 minutes and mixed thoroughly beforehand. The sample was subsequently separated by centrifugation (12,000 g, 15 min, 4 °C) into a lower red phenol-chloroform phase, an interphase, and a colourless upper aqueous phase. The aqueous phase contains the RNA and was cautiously transferred to a new tube. The rest was discarded. Glycogen (Thermo Scientific, cat. no. R0551) was employed for precipitation (0.5µl) and the RNA was precipitated by adding an equal volume of isopropanol to the aqueous phase, followed by 10 minutes incubation at room temperature. The mix was centrifuged to pellet the precipitated RNA (12,000 g, 10 min, 4 °C). The supernatant was carefully decanted and the RNA pellet was washed twice in 75% ethanol. The pellet was left to air-dry and was then dissolved in nuclease-free water.

2.5.2. RNA electrophoresis

A bleach agarose gel was used to determine the integrity of the total RNA precipitated (Aranda et al., 2012). Ordinary household bleach was added to the standard 1% agarose suspension in 1x TAE (Tris-acetate-EDTA) buffer to a final concentration of 0.06%

sodium hypochlorite. The agarose-bleach suspension was incubated at room temperature for 5 minutes with occasional swirling before it was dissolved in a microwave oven. The RNA samples were diluted in standard DNA loading buffer and directly loaded unto the gel without pre-heating.

2.5.3. DNase treatment

DNase treatment was performed on RNA samples in order to remove contaminating DNA from the samples using the TURBO DNA-free kit (Ambion, cat. no. AM1907) and following the manufacturer's instructions. RNA sample were usually diluted to 200 ng/µL in 50 µL of 1x TURBO DNase buffer. 1 µL of the DNase was added to the RNA sample and the tube mixed gently. The digestion was incubated at 37 °C for 25 minutes in a PCR cycler. Afterwards, the enzyme was deactivated with DNase Inactivation Reagent (5 µL per 1 µL of DNase). The sample was inverted gently to mix and incubated at room temperature for 5 minutes, with occasional inverting. The DNase Inactivation Reagent was pelleted by centrifugation (1 minute on a mini centrifuge), and the supernatant containing RNA was transferred to a fresh tube. The obtained DNase-treated RNA was subsequently analysed by electrophoresis and NanoDrop.

2.5.4. cDNA library synthesis

Total RNA obtained from cells was reverse transcribed using SuperScript II Reverse Transcriptase (Invitrogen, cat. no. 18064014). Generally, a 20 µL reaction was performed for 800 ng of DNase-treated RNA. The reaction was set up in duplicate for each RNA sample, one for cDNA synthesis and one for the no reverse transcriptase control (NRT), which checks for DNA contamination. The RNA was diluted in nuclease-free water to a final volume of 8 µL. 1 µL of random hexamers (50 µM, Invitrogen, cat. no. N8080127), 2 µL of dNTPs (5 mM each), 1 µL of nuclease-free water were added and mixed. The sample was heated to 65 °C for 5 minute and chilled on ice immediately after. After a brief centrifugation on the mini centrifuge, 4 µL of First-Strand Buffer (5×), 2 µL of DTT (0.1 M) and 1 µL of RNase Inhibitor (40 U/µL, Thermo Scientific, cat. no. EO0382) were added, mixed and incubated at 25 °C for 2 min. 1 µL (200 units) of the SuperScript II Reverse Transcriptase was added to each RNA sample (but not to their respective NRT controls). All sample were then incubated at 25 °C for 10 minutes, 42 °C for 50 minutes, and finally at 70 °C for 15 minutes. All samples and their NRT controls were diluted in 380 µL MilliQ water and subsequently used as templates for real-time PCR.

2.5.5. Real-time qPCR for cDNA expression levels

Real-time qPCR analysis for expression studies was performed using iTaq Univeral SYBR Green Supermix (BioRad cat. no. 1725124) in 10 µL reaction volume following the manufacturer's instructions. The reaction mix contains 2 µL of the diluted cDNA (or NRT), 2 µL of primer mix (2.5 µM of each), 1 µL of water and 5 µL of the supermix (2×). The thermal cycling was consistently performed on a CFX96 Touch (Bio-rad). Initial denaturation and polymerase activation for 2 min at 95 °C, followed by 40 cycles of 5 sec at 95 °C and 30 sec at 60 °C. Real-time PCR primers used in this study are listed in Table 2.6.3. For primer sets that amplify repetitive targets, negative controls of NRT were included in parallel. The results were analysed using the CFX Manager Software (Bio-rad).

2.6. Microscopic analysis

2.6.1. Indirect immunofluorescence

Mouse ES cells and mouse embryonic fibroblasts were grown with their respective culture medium on microscope slides (Thermo Scientific, cat. no. J1800AMNT) coated with 0.1% gelatine for upwards of 1.5 hours, except for MEFs, where the slides were used uncoated. The slides were placed in round tissue culture dishes in a tissue culture CO_2 incubator at 37 °C. After the cells attached to the slides the medium was removed by pouring PBS into the tissue culture dishes containing the slides. The slides were then transferred into Coplin jars and washed once in PBS. For immunostaining with anti-FLAG antibody, cells were fixed in 4% formaldehyde in PBS at room temperature for 5 minutes, followed by a brief wash in PBS. Cells were then permeabilized with 0.4% triton in PBS at room temperature for 15 minutes, followed by another wash in PBS. For staining with anti-SMARCAD1 antibody, cells were fixed in 4% formaldehyde in PBS for 10 minutes, permeabilized with 0.1% triton for 15 minutes, followed by another wash in PBS. Slides were then blocked in 10% filtered serum in PBS at 4 °C overnight. The primary antibody was diluted in the 10% serum blocking buffer and applied to the slides in a humidified chamber for one hour at room temperature. The slides were then washed three times in a Coplin jar with PBS for 5 minutes per wash in a shaker. The secondary

fluorochrome-coupled antibody was similarly diluted in 10% serum blocking buffer and applied to the cells for one hour at room temperature as indicated for the primary antibody. The three washing steps in PBS were repeated and the slides then mounted using Vectashield mounting medium containing DAPI (Vector Laboratories, cat. no. h-1200). Slides were viewed using a Leica DMR fluorescence microscope, and images were collected using a QuantiFire XI digital microscope camera.

2.6.2. GFP visualisation

In order to visualise and monitor the GFP expression of transient transfections with GFP vectors or in inducible *Smarcad1* knockdown cells for the induction of GFP-shRNA transcription, ES cells were grown on slides as described for immunofluorescence stainings. Cells were fixed with 2% formaldehyde for 5 minutes, permeabilized with 0.5% triton for 5 minutes, and mounted using Vectashield mounting medium containing DAPI. Slides were viewed as indicated for immunofluorescence.

2.7. Key resource tables

Antibody	Antigen	Source of Antibody	Method		
			IF	ChIP	Western
SMARCAD1 HPA016737	Rabbit	Sigma	1/80	3µg	1/1,000
SMARCAD1 PAB15737	Rabbit	Abnova		3µg	1/2,500
SMARCAD1 A301-593A	Rabbit	Bethyl		3µg	1/1,000
SMARCAD1 A301-592A	Rabbit	Bethyl		3µg	
SMARCAD1 ab67548	Rabbit	Abcam		3µg	
SMARCAD1 anti-CUE	Rabbit	Rowbotham et al. 2011	1/400	3µL	1/5,000
Lamin B1 ab16048	Rabbit	Abcam			1/10,000

KAP1 ab22553	Mouse	Abcam		3µg	1/10,000
IgG kCH-504 C15410206	Rabbit	Diagenode		3µg	
IgG I5381	Mouse	Sigma		3µg	
FLAG F3165	Mouse	Sigma			1/5,000
FLAG F1804	Mouse	Sigma	1/400	3µg	1/4,000
V5 R960-25	Mouse	Thermo Fisher Scientific			1/5,000
H3 ab1791	Rabbit	Abcam		3µg	1/10,000
H3K9me3 ab8898	Rabbit	Abcam		3µg	
H3K9me3 07-523	Rabbit	Milipore			1/5,000
H4K20me3 9053	Rabbit	Abcam		3µg	
SETDB1 11231-1-AP	Rabbit	Proteintech		3µg	
HDAC1 05-100	Mouse	Millipore			1/2,500
OCT4 sc-5279	Mouse	Santa Cruz			1/1,000
NANOG A300-397A	Rabbit	Bethyl			1/10,000
G9a PP-A8620A-00	Mouse	R&D			1/2,000
TFIIIC A301-242A	Rabbit	Bethyl		3µg	
SUV39H1 8729	Rabbit	Cell signaling			1/1,000
SETDB1 sc-66884	Rabbit	Santa Cruz			1/1,000
H3.3 09-838	Rabbit	Milipore		3µg	1/5,000

Table 2.7.1. List of Antibodies

Primer List

Target	Sequence (5'-3')	Applicati-on	Genomic copy number in silico PCR	Amplicon size, bp	Refere-nce
MLV_F	AAACCCCGGAAGAAAGAGAG	ChIP	61	80	M. Branco
MLV_R	CTGCTCATCCTCTGCCCTAC	ChIP			
VL30 pro (PBS-UTR)_F	TGGGGGCTCGTCCGGGAT	ChIP/RT	15	~ 78	(Wolf et al., 2015)
VL30 pro (PBS-UTR)_R	ATTACCAAGCGACAGAACTTACC	ChIP/RT			
IAPc* (5'UTR)_F	CGGGTCGCGGTAATAAAGGT	ChIP	1038	91	(Rowe et al., 2010)
IAPc* (5'UTR)_R	ACTCTCGTTCCCCAGCTGAA	ChIP			
IAPs* (MIER3)_F	GAACCTCTTGCCTCTTCCCC	ChIP	1	196	this study
IAPs* (MIER3)_R	GCTTGGGAGATGGATCCCAG	ChIP			
IAPs* (Bglap3)_F	AGGTGTTGCAGAGGTTTTGG	ChIP/RT	1	89	(Ecco et al., 2016)
IAPs* (Bglap3)_R	AATATCGGACACAGGGCAAG	ChIP/RT			
IAPs* (ZFP575)_F	TGAGCCTCTGTGTGGGTCCTA	ChIP	1	67	(Rowe et al., 2013)
IAPs* (ZFP575)_R	TGATTAAGAGCACTTGTTGCTTAGC	ChIP			
IAPez_F	GCTCCTGAAGATGTAAGCAATAAAG	ChIP	688	~143	(Liu et al., 2014)
IAPez_R	CTTCCTTGCGCCAGTCCCGAG	ChIP			

IAPez gag_F	CACGCTCCGGTAGAATACTTACAAAT	ChIP/RT	844	96	(Sharif et al., 2016)
IAPez gag_R	CCTGTCTAACTGCACCAAGGTAAAAT	ChIP/RT			
IAPU3_F	CGAGGGTGGTTCTCTACTCCAT	ChIP	190	88	(Rowe et al., 2010)
IAPU3_R	GACGTGTCACTCCCTGATTGG	ChIP			
IAPpol_F	CTTGCCCTTAAAGGTCTAAAAGCA	ChIP	490	77	(Rowe et al., 2010)
IAPpol_R	GCGGTATAAGGTACAATTAAAAGATATGG	ChIP			
IAP1pol_F	TGGCCATACCCCAAAGATAA	ChIP	36	119	(Faschin g et al., 2015)
IAP1pol_R	CCAGTTTACTGGGGCTGGTA	ChIP			
IAPgag_F	AATCTCAGAACCGCTCCATGA	ChIP/RT	447	77	(Rowe et al., 2010)
IAPgag_R	TTTCTTAAAATGCCCAGGCTTT	ChIP/RT			
Etn/Musd_F	GATTGGTGGAAGTTTAGCTAGCAT	ChIP	129	149	(Faschin g et al., 2015)
Etn/Mud_R	TAGCATTCTCATAAGCCAATTGCAT	ChIP			
MMERKV10C_F	TTCGCCTCTGCAATCAAGCTCTC	ChIP	79	~137	(Liu et al., 2014)

MMERKV10C_R	TCGCTCRTGCCTGAAGATGTTTC	ChIP			
MERVL_F	CTTCCATTCACAGCTGCGACTG	ChIP	512	155	(Liu et al., 2014)
MERVL_R	CTAGAACCACTCCTGGTACCAAC	ChIP			
L1_F	TTTGGGACACAATGAAAGCA	ChIP	1426	155	(Fadloun et al., 2013)
L1_R	CTGCCGTCTACTCCTCTTGG	ChIP			
L1MdF_F	GCATCTCTGGGGTGAGCTAG	ChIP	1	148	(Castro-Diaz et al., 2014)
L1MdF_R	AAAAGGGTGCTGCCTCAGAA	ChIP			
Intergenic_F	CAGCATTCCAGGAGGTTAGC	ChIP	1	147	(Karnowski et al., 2008)
Intergenic_R	GTGCCTCATGTGCAGTCAGT	ChIP			
EzrP1_F	GGCCCCGTAACTGCTCTTTA	ChIP	1	199	(Ding et al., 2018)
EzrP1_R	AGTATAAGACGCTGCGGCAA	ChIP			
EzrP2_F	TCCCTCCTGCACGTGGTAAT	ChIP	1	157	This study
EzrP2_R	AAGGGTCCTCACTTGACCAG	ChIP			
EzrP3_F	ATAGGGGCGAGTAGCCATTG	ChIP	1	71	This study
EzrP3_R	GTGAAGGCGTTTGTTGGACTC	ChIP			

Srrm2P1_F	CCTCAGAAAGTTGCCCACCA	ChIP	1	147	This study
Srrm2P1_R	TGCCATGCGAGCGATTCTTA	ChIP			
Srrm2P2_F	AGACCAGCAGTTCTCGCTCA	ChIP	1	82	This study
Srrm2P2_R	CTCCAACCAACTTACCGTCCT	ChIP			
Peg13_F	ACACTGGATGGAGACTTGCAG	ChIP	1	83	(Mikkelsen et al., 2007)
Peg13_R	CGCGGTGTACCGAGTATTTG	ChIP			
Peg3_F	CACCGCCACTGCGGCAAAAC	ChIP	1	88	(Voon et al., 2015)
Peg3_R	GGCTGGCAGGGTCTTCGCAA	ChIP			
Fkbp6_F	ATCTTGCCGCACAACTGTCT	ChIP	1	95	(Sachs et al., 2013)
Fkbp6_R	GATCACGTGCCGTTTCTCTC	ChIP			
ZFP629_F	TTGGCTCCTGGTGGCGGAGT	ChIP	1	74	(Voon et al., 2015)
ZFP629_R	GCCCGGAGAGCTCAGGGTGA	ChIP			
ZFP13_F	GGGTGCGATTATGTGCGATG	ChIP	1	76	This study
ZFP13_R	GTCAAGCCCTACCCGTGC	ChIP			
Polrmt_F	GCTCGTGACTCCATACCGTC	ChIP	1	108	This study
Polrmt_R	GCCTCAAGAAGGGAGACTCG	ChIP			

Xist_P2_F	TCATGTGACCTGCCCTCTAGT	ChIP	1	109	S. Buonomo
Xist_P2_R	CACCCTACCATAATGCACCA	ChIP			
Xist_P1_F	TAAGGCTTGGTGGTAGGGGA	ChIP	1	97	(Elsasser et al., 2016)
Xist_P1_R	TTTGCTCGTTTCCCGTGGAT	ChIP			
Major Satellites_F	GACGACTTGAAAAATGACGAAATC	ChIP	N.A.	N.A.	(Martens et al., 2005b)
Major Satellites_R	CATATTCCAGGTCCTTCAGTGTGC	ChIP			
Minor Satellites_F	CATGGAAAATGATAAAAACC	ChIP	N.A.	N.A.	(Martens et al., 2005b)
Minor Satellites_R	CATCTAATATGTTCTACAGTGTGG	ChIP			
Telomeres_F	CGGTTTGTTTGGGTTTGGGTTTGGGTTTGGGTTT	ChIP	N.A.	N.A.	(Jang et al., 2015)
Telomeres_R	GGCTTGCCTTACCCTTACCCTTACCCTTACCCT	ChIP			
TFIIIC_F	CCTTCGATAGCTCAGCTGGTAGAGCGGAGG	ChIP	1		(Winter et al., 2000)
TFIIIC_R	CGGAATTGAACCAGCGACCTAAGGATGTCC	ChIP			

Atp5b_F	GGCCAAGATGTCCTGCTGTT	RT	1	106	(Wossidlo et al., 2011)
Atp5b_R	GCTGGTAGCCTACAGCAGAAGG	RT			
Hspcb_F	GCTGGCTGAGGACAAGGAGA	RT	1	93	(Wossidlo et al., 2011)
Hspcb_R	CGTCGGTTAGTGGAATCTTCA	RT			
Gapdh_F	TCCATGACAACTTTGGCATTG	RT	1	72	(Rowe et al., 2010)
Gapdh_R	CAGTCTTCTGGGTGGCAGTGA	RT			
Rex1_F	CGATGCTGGAGTGTCCTCAAG	RT	1	113	(Ficz et al., 2011)
Rex1_R	GCCACACTCTGCACACACGT	RT			
Smarcad1_F	TTCCTGGCATACCTCTTTC	RT	1	155	(Ding et al., 2018)
Smarcad1_R	ATTTGCTTACGCTCTTCTTG	RT			
Xist_F	GCCTCAAGAAGAAGGATTGC	RT	1	75	(Jonkers et al., 2009)
Xist_R	GGGATTGTTTGTCCCTTTGG	RT			
Hccs_F	ACCCCGCGTGAGGAAAAACCG	gDNA	1	170	(Wolstenholme et al., 2013)
Hccs_R	CAACCTGCCCAGCCGAACTCA	gDNA			

Msl3_F	TGGTAGACAGCGCAGGTGGT	gDNA	1	130	(Wolstenholme et al., 2013)
Msl3_R	TGCTTTGCGGCTTCCAAGCACT	gDNA			
Obp1a_F	GGATCAGAATTATGGATCATGTG	gDNA	1	150	(Wolstenholme et al., 2013)
Obp1a_R	GATCATGAGAAGGGGAAGGA	gDNA			
Olfr1_F	TCTATGCCCTGTTCCTGGTC	gDNA	1	164	(Wolstenholme et al., 2013)
Olfr1_R	CAACTTGGGCATTGTGACAG	gDNA			
Ppib_F	TGGAGAGCACCAAGACAGACA	gDNA	1	66	(Wolstenholme et al., 2013)
Ppib_R	TGCCGGAGTCGACAATGAT	gDNA			

Table 2.7.2. List of Primers

3. Results

3.1. Identification and characterization of genome-wide SMARCAD1 binding sites in mouse embryonic stem cells

Closer inspection of the *in vivo* binding sites of SMARCAD1 genome-wide as well as its correlation with other epigenetic marks and chromatin regulators will give further insight into its function in ES cells. Remodellers generally do not directly bind DNA and as such are not trivial to ChIP. As no ChIP-validated antibody was available for SMARCAD1 the first step was to develop FLAG-ChIP-seq. Stable cell lines expressing 3xFLAG-SMARCAD1 constructs were developed to investigate the genome-wide association of SMARCAD1 via ChIP-seq. The FLAG peaks were then used to develop a protocol for ChIP-seq with endogenous SMARCAD1. The SMARCAD1 constructs used, contain a 5'-3xFLAG tag which has been successfully used for ChIP-seq (Gelbart et al., 2005; Volkel et al., 2015). Two distinct constructs were used; a FLAG-WT and a FLAG-ATPase mutant construct of SMARCAD1. The mutation is well characterized and involves a single-nucleotide change (the substitution of a highly conserved arginine with lysine: K523R) within the ATPase domain of SMARCAD1 which abolishes the enzymes' ability to hydrolyse ATP, resulting in a catalytically inactive mutant (Neves-Costa et al., 2009; Richmond and Peterson, 1996; Rowbotham et al., 2011; Steglich et al., 2015). This catalytically inactive mutant has previously been used to study the function of ISW2, another ATP-dependent chromatin remodeller (Gelbart et al., 2005) where it was found to be a more sensitive ChIP reagent in marking its binding sites *in vivo* than its WT counterpart. It was proposed that the WT remodeller is recruited only transiently and only a basal interaction can be detected by ChIP when it "scans" the genome, whereas when the remodeller is no longer able to hydrolyse ATP it is possible to "capture" this transient interaction on chromatin via ChIP (Gelbart et al., 2005). Hence, to increase the probability to capture SMARCAD1 on chromatin, an ATPase mutant was to be used for FLAG-ChIP-seq together with the WT.

3.1.1. Establishment and characterization of FLAG-SMARCAD1 stable ES cell lines

Stable 3xFLAG-SMARCAD1 mouse ES cell lines were established for ChIP-seq. Two different constructs were transfected into PGK12.1 mES cells; 3xFLAG-SMARCAD1

WT and ATPase mutant (Figure 3.1.1a-b). The stable cell lines were sequenced to ensure the presence of the SMARCAD1 ATPase mutant and WT sequence (Figure 3.1.1b).

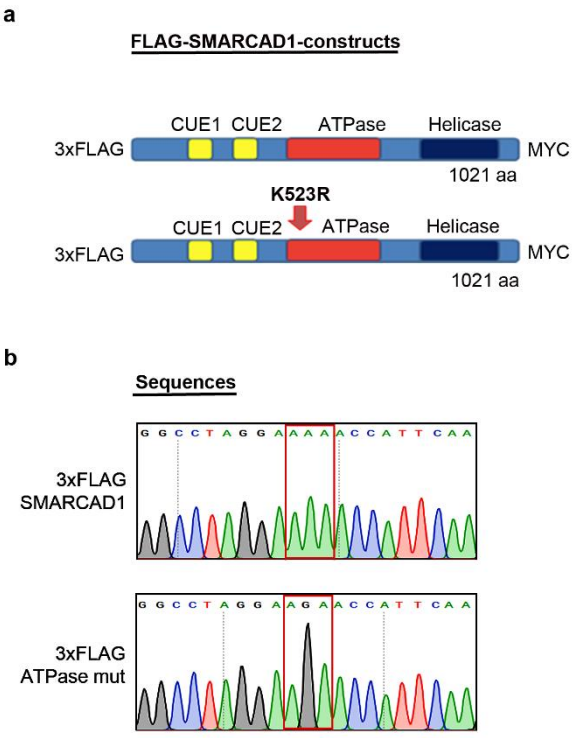

Figure 3.1.1: Overview of the 3xFLAG-SMARCAD1 constructs, WT and ATPase mutant
(a) depicts the two FLAG-constructs transfected into PGK12.1 mES cells. The upper cartoon shows the 3xFLAG-SMARCAD1 WT construct with its main domains and the lower one shows the ATPase mutant with the lysine to arginine substitution at amino acid 523. (b) DNA sequence of both cell lines by Sanger Sequencing. On the top the WT construct with the codon for lysine highlighted and on the bottom the ATPase mutant with a single base substitution resulting in a codon coding for arginine.

The constructs were transfected into wild-type PGK12.1 mES cells and stable clones were selected, >50 WT SMARCAD1 expressing clones were obtained. In contrast, the transfection of the ATPase mutant resulted in a small number of colonies expressing the construct. Five stable clones of the ATPase mutant were obtained from two separate transfections. The WT clone with the lowest overexpression was selected in order to minimise potential off-target effects caused by substantial overexpression, plus to maintain comparable levels to the ATPase mutant which are expressed at much lower levels (Figure 3.1.2a). The highest expressing ATPase mutant clone (mt8) was selected. A mix of expression levels was observed by IF in the cell line, hence an attempt was made

to clone out the highest expressing cells. For this, cells were cloned out using 96 well plates and LiCor quantification and the resulting clone is marked as mt8c in Figure 3.1.2a. However, it showed no improved expression levels compared to the initial clone (Figure 3.1.2a, lane 6). Similarly, the transient transfection of 3xFLAG-SMARCAD1 WT and ATPase mutant constructs on top of the stable cell lines expressing the construct did not improve the level of 3xFLAG-SMARCAD1 ATPase mutant expression, whereas WT expression increased (Figure 3.1.2b). To rule out poor sample quality the vectors were repeatedly phenol-chloroform extracted and Nanodrop measurements indicated very pure DNA. It was hypothesized that the ATPase mutant has a dominant negative effect and its expression is silenced. However, a different stable ATPase mutant cell line (with a single FLAG tag) was established simultaneously by the same experimenter for a colleague and in the same cells, and showed similar expression to its WT construct. Additionally, another stable ATPase mutant cell line in different cells (E14) was successfully established by colleagues in the lab.

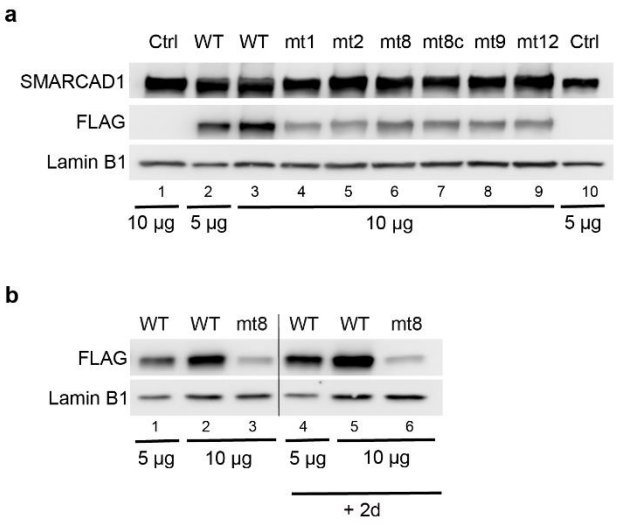

Figure 3.1.2: Expression analysis of FLAG-SMARCAD1 (WT and ATPase mutant) in stable ES cell clones.

(a) Expression analysis of SMARCAD1 by Western blot comparing a 3xFLAG-SMARCAD1 expressing cell line (WT; lane 2 & 3) compared to all 6 clones of the 3xFLAG-SMARCAD1 ATPase mutant. Shown as a control; a pool expressing the empty vector backbone (lane 1&10). (b) Western blot depicting the stable cell lines selected (WT; lane 1 & 2, and mt; lane 3), before and 2 days after the transient transfection of the corresponding vectors on top of the stable expression. 5µg and 10µg from total protein extracts.

The stable 3xFLAG-SMARCAD1 WT cell line depicted in Figure 3.1.2 was selected for further experiments together with the 3xFLAG empty vector pool as a control. The catalytically inactive mutant was not included in the further analysis due to the problems encountered when trying to establish a stably expressing cell line. At a later stage in this study an ATPase mutant in a different ES cell line (E14) is employed.

The pluripotency of the WT cell line is not affected by the expression of the construct as shown by the pluripotency marker Nanog and neither is the expression of the transcriptional regulator KAP1, the stoichiometric binding partner of SMARCAD1 (Figure 3.1.3a). In addition, the SMARCAD1 WT construct localizes similar to the endogenous SMARCAD1 protein; to the nucleus of mES cells overall and with distinct enrichment at DAPI-dense loci (Figure 3.1.3b) which represent the heterochromatin of mouse ES cells. Taken together, this illustrates that the tagged overexpressed protein is a good proxy for the native one. The stable 3xFLAG-SMARCAD1 WT cell line was selected for ChIP-seq together with the 3xFLAG empty vector pool as a control.

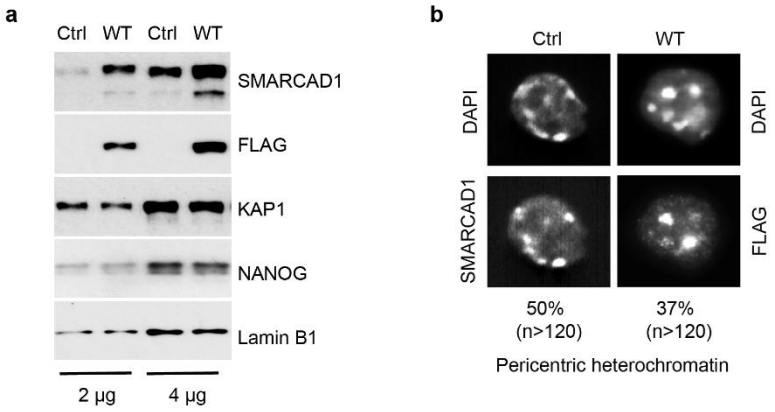

Figure 3.1.3: The 3xFLAG-WT protein behaves like the endogenous SMARCAD1 protein. (a) Western blot analysis showing PGK12.1 mES cells stably expressing 3xFLAG-SMARCAD1 protein. The level of SMARCAD1 and FLAG-protein expression is shown in two different dilutions. Pluripotency is not affected as shown by NANOG. Shown as a control; a pool expressing the empty vector. Lamin B1 is shown as a loading control. 2µg and 4µg from total protein extracts from PGK12.1 mES cells. (b) Indirect Immunofluorescence of the 3xFLAG-SMARCAD1 protein and endogenous SMARCAD1 in PGK12.1 mES cells. Soluble proteins were extracted with detergent prior to fixation. Endogenous SMARCAD1 localizes to pericentric heterochromatin (DAPI-dense foci) in half of the cells (n>120) and the WT construct in 37% (n>120).

3.1.2. Validation of identified FLAG-SMARCAD1 binding sites

The FLAG-ChIP-seq resulted in 3317 detected binding sites after filtering. The results of the FLAG-ChIP-seq will be discussed in more detail later. The focus in this section is on how selected binding sites were validated by ChIP-qPCR and the information used to test antibodies against the endogenous, untagged protein. The obtained peaks were used to optimize the ChIP-seq protocol for pull-down with an antibody against the endogenous SMARCAD1 protein since the antibodies against SMARCAD1 available at this time point had not been validated for ChIP. The top 100 FLAG-SMARCAD1 peaks were investigated manually (representative screenshot shown in Figure 3.1.4a, left panel) and five peaks were selected for primer walking by ChIP-qPCR (representative example: Fig. Figure 3.1.4a, right panel). FLAG-SMARCAD1 enrichment was confirmed to be restricted to the peak and progressively lost with primers lying progressively further away from the SMARCAD1 peak.

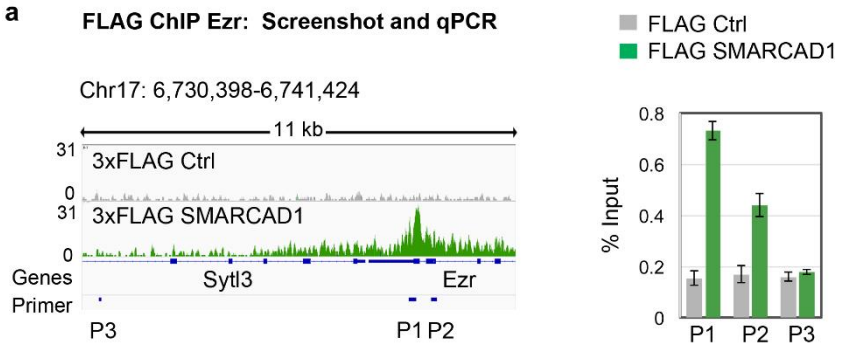

Figure 3.1.4: 3xFLAG-SMARCAD1 is enriched at defined binding sites in mouse ES cells.
Left panel: example of a genome browser screenshot of a FLAG ChIP-seq peak over the Ezr gene in PGK12.1 3xFLAG SMARCAD1 mES cells. 3xFLAG empty vector ChIP-seq shown as control. Genomic location is indicated. Underneath three primer pairs are displayed used for primer walking as depicted in the right panel with ChIP-qPCR with a FLAG antibody. Enrichment is shown as percent input (+/-SE).

Five antibodies against the endogenous SMARCAD1 protein were selected for ChIP validation. They were tested extensively by the Mermoud lab for specificity in IF and Western blot and all antibodies were specific for SMARCAD1, detecting no signal in the knockdown. Subsequently, two FLAG-SMARCAD1 binding sites were selected to test

the antibodies by ChIP-qPCR in SMARCAD1 knockdown and control mouse ES cells. The binding sites were validated with five SMARCAD1 antibodies and shown to be specific for endogenous SMARCAD1 binding (*Ezr* and *Srrm2* shown in Figure 3.1.5a, b). Three antibodies were selected for further optimization by double-crosslinking in a different ES cell line in order to improve the levels of bound protein to chromatin, since remodellers traditionally do not bind DNA directly. Double-crosslinking is an efficient "two-step" ChIP method that involves the fixation of protein-protein interactions followed by protein-DNA fixation with formaldehyde as described in the standard method. For protein-protein crosslinking NHS-ester crosslinking agents are used, such as DSG, which results in irreversible covalent amide bonds (Tian et al., 2012). Two antibodies showed robust enrichment using double-crosslinked compared to single-crosslinked cells (Figure 3.1.5c). In addition, the enrichment was specific to the SMARCAD1 peak and not detected over unspecific regions downstream of the binding site.

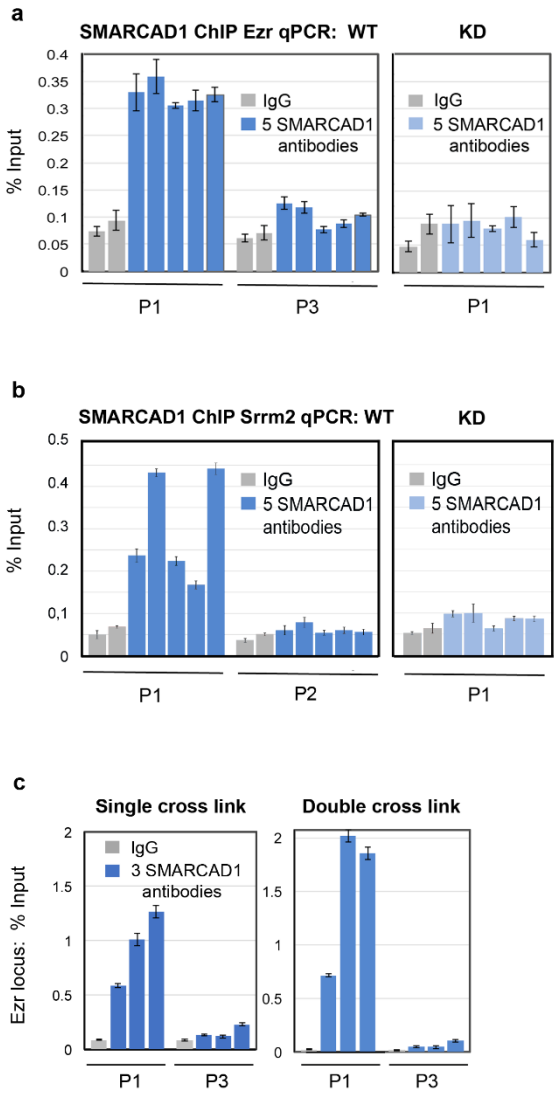

Figure 3.1.5: SMARCAD1 ChIP-qPCR detects specific peaks and can be optimized by double-crosslinking.

(a) & (b) ChIP-qPCR on a candidate locus is depicted (*Ezr* & *Srrm2*) with five different SMARCAD1 antibodies (left to right: A301-593A, A301-592A, Anti-CUE, ab67548, PAB15737) in single-crosslinked chromatin from PGK12.1 SMARCAD1 knockdown and Control mES cells. The primers are as shown in Figure 3.1.4. Enrichment is shown as percent input +/-SE. (c) FA-DSG crosslinking leads to a robust enrichment of SMARCAD1 binding. Exemplary ChIP-qPCR with three antibodies against SMARCAD1 in FA-crosslinked chromatin as well as in DSG-FA-crosslinked chromatin (E14 mES cells Antibodies from left to right: A301-592A, PAB15737 and HPA016737. P3 is shown as a negative control for a binding site and IgG as an antibody control. Enrichment is shown as percent input (+/-SE).

Endogenous ChIP-seq was performed with the PAB15737 antibody and double-crosslinked chromatin in PGK12.1 control and SMARCAD mES knockdown cells. This SMARCAD1 knockdown cell line has been well characterised by Jacqueline Mermoud and the results have been published in (Sachs et al., 2019).

3.1.3. Genome-wide binding sites of SMARCAD1 in mouse ES cells

In addition to FLAG and SMARCAD1 ChIP-seq, H3K9me3 ChIP-seq was performed in the same mES cell line (PGK12.1) to determine its overlap with SMARCAD1 binding sites. To determine the high affinity binding sites of SMARCAD1 in mES cells, peaks occurring in both the FLAG-ChIP-seq as well as in the endogenous ChIP-seq were used. It was, however, determined beforehand that they are of similar high quality (Table 3.1.1). The bioinformatic analysis was performed in collaboration with bioinformaticians Andrea Nist and Boris Lamp, who performed the sequencing, the read alignment and the peak calling. The figures rendered by Boris Lamp are designated as such.

The number of effective reads sequenced is similar between the SMARCAD1 ChIP-seqs as well as in the backgrounds used, meaning all libraries were sequenced at comparable depth. The number of unfiltered and filtered peaks is also indicated. The endogenous and FLAG-ChIP-seq detect a similar number of filtered SMARCAD1 binding sites (5727 and 3317, respectively) and in the SMARCAD1 knockdown only 63 peaks are detected (Table 3.1.1).

ChIP Seq	endogenous SMARCAD1	SMARCAD1 knockdown	FLAG SMARCAD1	H3K9me3
ESCs	PGK12.1	PGK12.1	PGK12.1	PGK12.1
Cross-linking	double	Double	double	Double
ChIP antibody	SMARCAD1	SMARCAD1	FLAG	H3K9me3
Effective reads	7.93E+07	6.52E+07	5.05E+07	5.47E+07
Used background	IgG WT	IgG KD	FLAG Ab	IgG WT
	Input WT	Input KD	Input of Control	Input WT

Effective background reads	6.71E+07	6.44E+07	4.66E+07	6.71E+07
	7.96E+07	6.37E+07	6.22E+07	7.96E+07
Peak caller	MACS	MACS	MACS	SICER
unfiltered peaks	15418	287	8786	55183
filtered peaks	5727	63	3317	10849

Table 3.1.1. Summary of ChIP-seqs in PGK12.1 mES cells

FLAG-SMARCAD1 and endogenous SMARCAD1 ChIP-seq in mES cells summarising the type of crosslinking, the antibodies used.

The tagged-SMARCAD1 and endogenous ChIP-seq were investigated for the genomic distribution of SMARCAD1. 2380 peaks were found in both datasets (Figure 3.1.6c). These "high confidence" peaks were selected for further analysis. Their genomic distribution was analysed and it was found that they lie predominantly at intergenic sites in the mouse genome (Figure 3.1.6a). They are found at other interesting sites, such as promoters, and this will be discussed in more detail later. Most importantly, the SMARCAD1 binding sites identified, designate specific peaks since they disappear in the SMARCAD1 knockdown shown in the cumulative plot of all SMARCAD1-enriched sites (Figure 3.1.6b).

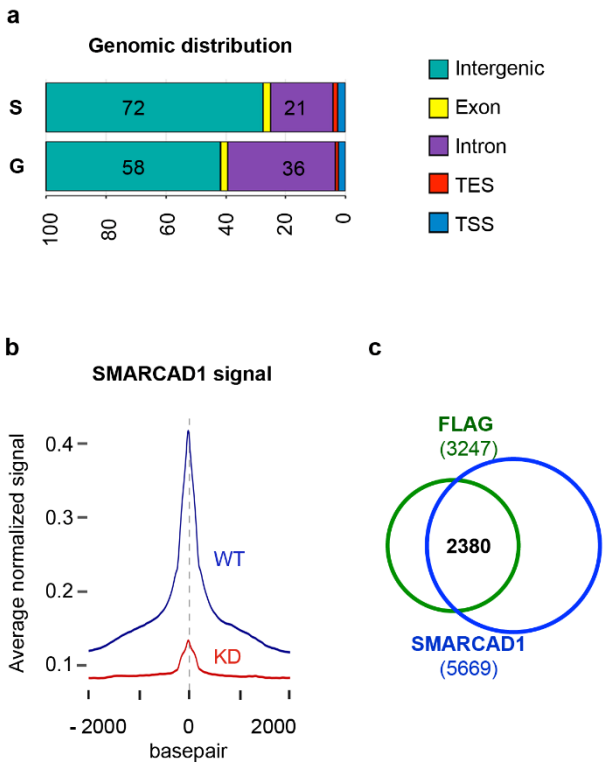

Figure 3.1.6: FLAG-SMARCAD1 and endogenous SMARCAD1 ChIP-seq identifies predominantly overlapping binding sites.

(a) Genomic distribution of SMARCAD1 (S) compared to the distribution of elements in the mouse genome (G). The numbers depict percentages. SMARCAD1 predominantly binds intergenic regions. SMARCAD1 genomic distribution compared to the distribution of elements in the murine genome. The largest proportion (72%) of SMARCAD1 binding sites lie at intergenic sites followed by intronic regions (21%). (b) Cumulative plot of all called SMARCAD1 peaks in mES cells and SMARCAD1 knockdown cells showing the specificity of SMARCAD1 binding sites. (c) Venn diagram depicting the 5669 peaks detected in the endogenous SMARCAD1 ChIP-seq after filtering, of which 2380 are also present in the FLAG-SMARCAD1 ChIP-seq (from a total of 3247 peaks). The genomic distribution, cumulative plot, and Venn diagram were generated in collaboration with Boris Lamp and Philipp Bergmaier.

The ChIP-seq of the WT construct and the endogenous SMARCAD1 protein share a large number of peaks. However, the fact remains that the endogenous ChIP detected more peaks than the FLAG ChIP-seq and a substantial number of peaks do not overlap. If these peaks are investigated manually however, it becomes clear that the endogenous peaks do also display FLAG enrichment, but at a level underneath the assigned peak threshold. Therefore, it must be concluded that the FLAG-ChIP is generally less sensitive but the

actual overlap is potentially much larger than observed, and the two methods do not detect different binding sites.

3.1.4. SMARCAD1 binding sites are associated with heterochromatic factors and modifications

To determine the role of SMARCAD1 in ES cells the overlapping peaks from the FLAG-SMARCAD1 and endogenous SMARCAD1 ChIP-seq, considered high-confidence peaks, were investigated for distinct chromatin features. Primarily, the overlap with the H3K9me3 ChIP-seq from this study, and a published KAP1 (Elsasser et al., 2015) dataset was determined. Published datasets of other heterochromatic factors and modifications were also included (Figure 3.1.7). SMARCAD1 binding sites are predominantly associated with specific repressive histone marks genome-wide, such as H3K9me3 (our dataset) and H4K20me3 (Mikkelsen et al., 2007). They are not linked with regions containing the polycomb-associated repressive mark H3K27me3 nor with sites of active chromatin denoted by histone marks such as H3K27ac or H3K4me3. SMARCAD1 binding sites also overlap with the transcriptional regulator KAP1, and SETDB1, a histone methyltransferase, as well as the histone variant H3.3. Other histone methyltransferases are not enriched in these regions (G9a, Suv39h1 and Suv39h2). Indeed 85.5% of SMARCAD1 sites overlap with H3K9me3 and 87.5% with KAP1 binding sites (Figure 3.1.7b), (KAP1 dataset: (Elsasser et al., 2015). Altogether, the vast majority of high-confidence SMARCAD1 sites are associated with both H3K9me3 and KAP1 (82%) (Figure 3.1.7c). The overlap of SMARCAD1 sites with H3K9me3 might indicate a role in constitutive heterochromatin. Whereas H3K9me3 is important for many types of constitutive heterochromatin, the presence of H4K20me3, KAP1 and SETDB1 is indicative of transposable elements.

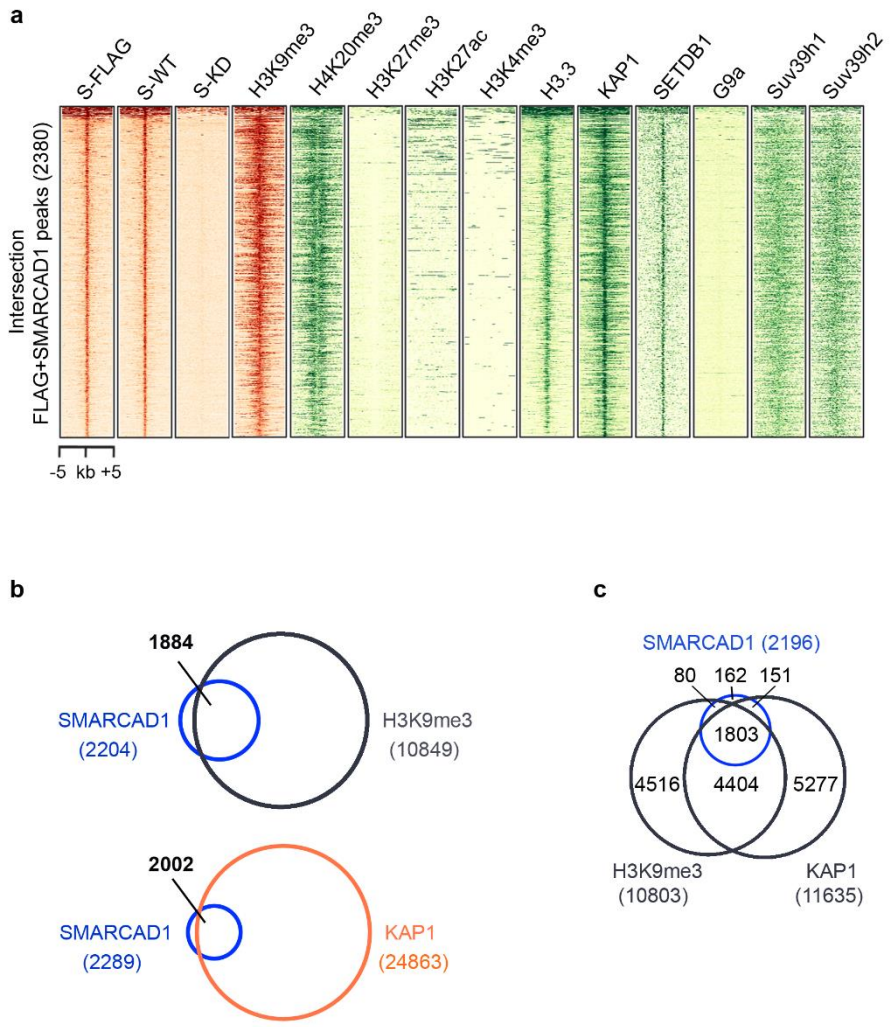

Figure 3.1.7: SMARCAD1 binds heterochromatic regions together with ERV regulators.

(a) heatmaps made by Boris Lamp of high-confidence SMARCAD1 peaks (intersections of endogenous and FLAG-SMARCAD1) and SMARCAD1 knockdown showing the specificity of SMARCAD1 binding sites. Also shown is the association of SMARCAD1 with the histone marks H3K9me3 (this study) and H3K20me3, the histone variant H3.3, and the ERV regulators KAP1 and SETDB1. Genomic scale depicted on the bottom of the heatmaps. b) and c) Venn diagrams made by Philipp Bergmaier showing the overall shared sites between SMARCAD and H3K9me (1884) and between SMARCAD1 and KAP1 (2289). Also shown are the sites shared between both KAP1 and H3K9me3, which make out the vast majority of all SMARCAD1 sites (82%). Intersection of FLAG and SMARCAD1; 2196 peaks) with the H3K9me3 profile (10803 peaks; this study) and KAP1 binding sites; 11635 peaks (Elsasser et al., 2015). Publicly available external ChIP-Seq data were obtained from GEO using the following accession numbers: GSM1555120 (KAP1), GSM1429923 (KAP1 input), GSM307622 (H4K20me3), GSM1033638 (H3K27me3), GSM594578 (H3K27ac), GSM1033636 (H3K4me3), GSM1555116 (H3.3), GSM459273 (SETDB1), GSM1215219 (G9a), GSM1375157 (SUV39H1), GSM1375158 (SUV39H2).

SMARCAD1 binding sites in mES cells were determined in two biological replicates. The shared peaks of the two ChIPs are predominantly localized to intergenic sites but contain a number of interesting peaks at other regions, such as at transcriptional start sites of promoters of imprinted genes, which will be investigated in the next chapters. The binding sites generally fall into two categories, single-copy genes and repetitive elements.

During the course of this study another SMARCAD1 ChIP-seq was published (Xiao et al., 2017). Single-crosslinking was employed by Xiao et al, and consequently, a significantly lower number of peaks was detected compared to the endogenous ChIP-seq discussed in this study (525 compared to 5727 peaks, respectively). Additionally, the knockdown cell line used was sequenced at considerably more depth than the control (8,19E+07 and 1,30E+07 effective reads, respectively) which gives rise to a large number of unspecific peaks in the knockdown. Overall, the coverage is low, hence, the ChIP-seq discussed in this study adds valuable information to the pursuit of defining the binding sites of remodellers in ES cells.

3.1.5. Loss of SMARCAD1 impacts ESC morphology and pluripotency

This study will employ a loss-of-function approach to determine the role of SMARCAD1 on its binding sites. The effect of SMARCAD1 knockdown in ES cells has been characterised by Jacqueline Mermoud and Xiao et al., (Sachs et al., 2019; Xiao et al., 2017). For the purpose of this study the effects on morphology and pluripotency of SMARCAD1 removal were investigated. Additionally, the global expression level of several chromatin modifiers as well as histone modifications were examined as their binding is studied on a local level subsequently.

SMARCAD1 knockdown affects the morphology of ES cells and their proliferation rate in addition of affecting the expression of distinct pluripotency markers (Sachs et al., 2019; Xiao et al., 2017). To confirm this in a distinct cell line from the PGK12.1 SMARCAD1 knockdown line characterised by Jacqueline Mermoud, this cell line in addition to an E14 SMARCAD knockdown cell line were grown three days in 96-well plates. Pictures were taken every three hours and the changes in their morphology quantified. Representative screenshots are shown in Figure 3.1.8a. Initially after plating, both control and knockdown cells show high eccentricity, however over time the control cells form

characteristic spherical ES cell colonies whereas the knockdown cells form flat, irregular colonies, similar to differentiated cells (Figure 3.1.8b).

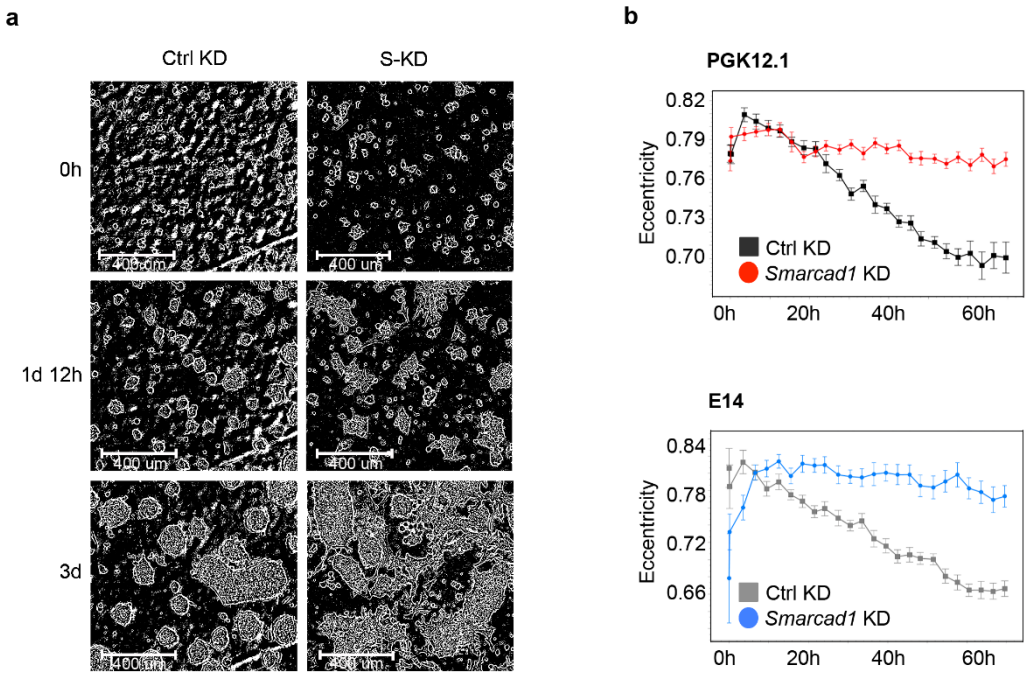

Figure 3.1.8: Upon the knockdown of SMARCAD1 mouse ES cells lose their spherical morphology.

(a) Depicted are snapshots at different time points during the growth of control and stable SMARCAD1 knockdown stem cells (E14) plated simultaneously at the same densities. (b) shown is the corresponding analysis of cell morphology at 3-hour intervals (depicted in average eccentricity) measured in real time with and IncuCyte for three days. PGK12.1 ESCs stably expressing an shRNA targeting Exon 7 and E14 ESCs stably expressing an shRNA targeting the 3'UTR of *Smarcad1* and appropriate control knockdown cells were plated at six different concentrations ranging from 10 K to 312 cells/ 96 well (n=3). A representative example is shown, the error bars depicting the S.D. from triplicates.

3.1.6. The global protein levels of heterochromatin factors are not affected by loss of SMARCAD1

The stable PGK12.1 SMARCAD1 knockdown clone used for endogenous ChIP-seq displays almost complete removal of SMARCAD1 protein levels observed in both Western blot and immunofluorescence (Figure 3.1.9). The pluripotency marker NANOG is reduced in the knockdown, as seen in previous experiments by Jacqueline Mermoud

and Dong Ding. However, proteins and enzymes important for heterochromatin regulation are globally unaffected (Figure 3.1.9a). Of note, the hallmark of constitutive heterochromatin, H3K9me3, is unaffected as are KAP1 and the histone methyltransferase SETB1/ESET which play important roles in H3K9me3 deposition over endogenous retroviral elements (ERVs). Other histone methyltransferases involved in H3K9me3 deposition over other regions are similarly unaffected (G9a & Suv39h1) and the histone deacetylase HDAC1 is also globally unchanged (Figure 3.1.9a).

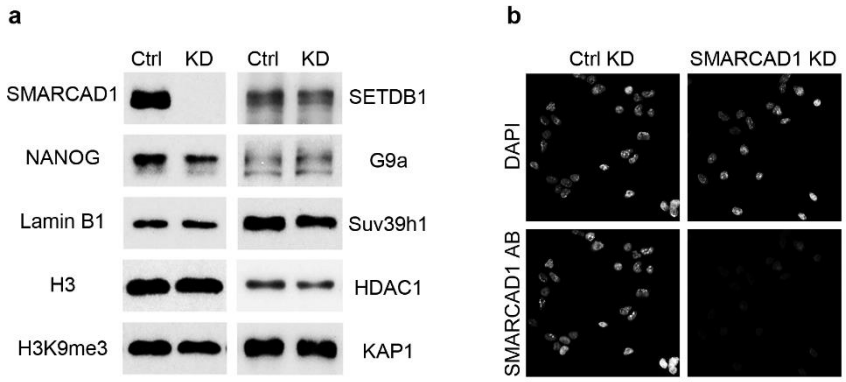

Figure 3.1.9: The stable knockdown of SMARCAD1 in mES cells does not affect the global expression of H3K9me3 and specific histone methyltransferases and deacetylases.

(a) Western blot of stable SMARCAD1 knockdown in PGK12.1 mES cells. Global levels of H3K9me3 are shown, as well as levels of the histone methyltransferases SETDB1, G9a and Suva39. Also shown are the global levels of histone deacetylase I and KAP1. H3 and Lamin B1 are shown as loading controls. (b) Immunofluorescence in PGK12.1 SMARCAD1 knockdown and Control with SMARCAD1 Sigma HPA016737.

3.2. SMARCAD1 is recruited to KAP1 targets by its CUE1 domain

SMARCAD1 interacts with several factors that play roles in repression, most notably the stoichiometric interaction partner of SMARCAD1 in both somatic and ES cells is KRAB-associated protein 1, KAP1 (TRIM28; TIF1β) (Ding et al., 2018). Additionally, several KRAB-ZFPs are found in the proteomic interactome of SMARCAD1 (Mermoud lab, unpublished).

3.2.1. SMARCAD1 binding sites fall into distinct categories

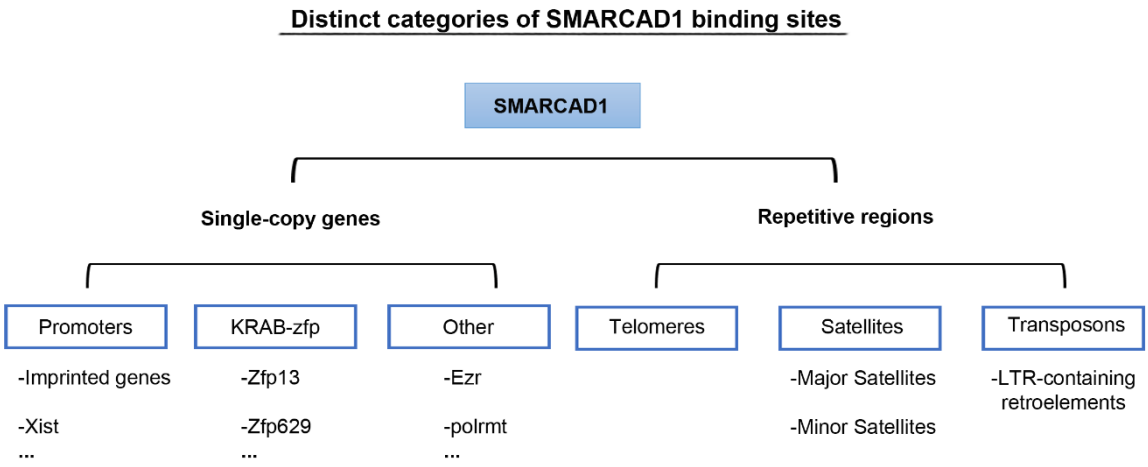

Figure 3.2.1: Types of SMARCAD1 binding sites in mouse embryonic stem cells.
Overview illustrating the different regions bound by SMARCAD1 in mouse ES cells.

SMARCAD1 peaks in mouse embryonic stem cells fall both within repetitive regions and on single-copy genes (Figure 3.2.1). The following examples have all been confirmed by ChIP-qPCR in two mES cell lines and the corresponding SMARCAD1 knockdown (depicted in Figure 3.2.2, Figure 3.2.3, and Figure 3.2.5). Among the single-copy genes were several interesting sites; the promoters bound by SMARCAD1 were often those of imprinted genes (*Fkbp6*, *Peg3*, and *Peg13* (Figure 3.2.2a, left panel). Also bound by SMARCAD1 are the two regulatory promoters of the non-coding gene Xist which is necessary for X-inactivation (Figure 3.2.2a, right panel). SMARCAD1 was also found to bind the 3'end of genes coding for KRAB-ZFPs (Figure 3.2.2a, left panel). Other single-copy genes were found to be highly bound by SMARCAD1 not on their promoters, but within intronic sequences (Figure 3.2.2a, left panel; *Ezr*, *Polrmt*).

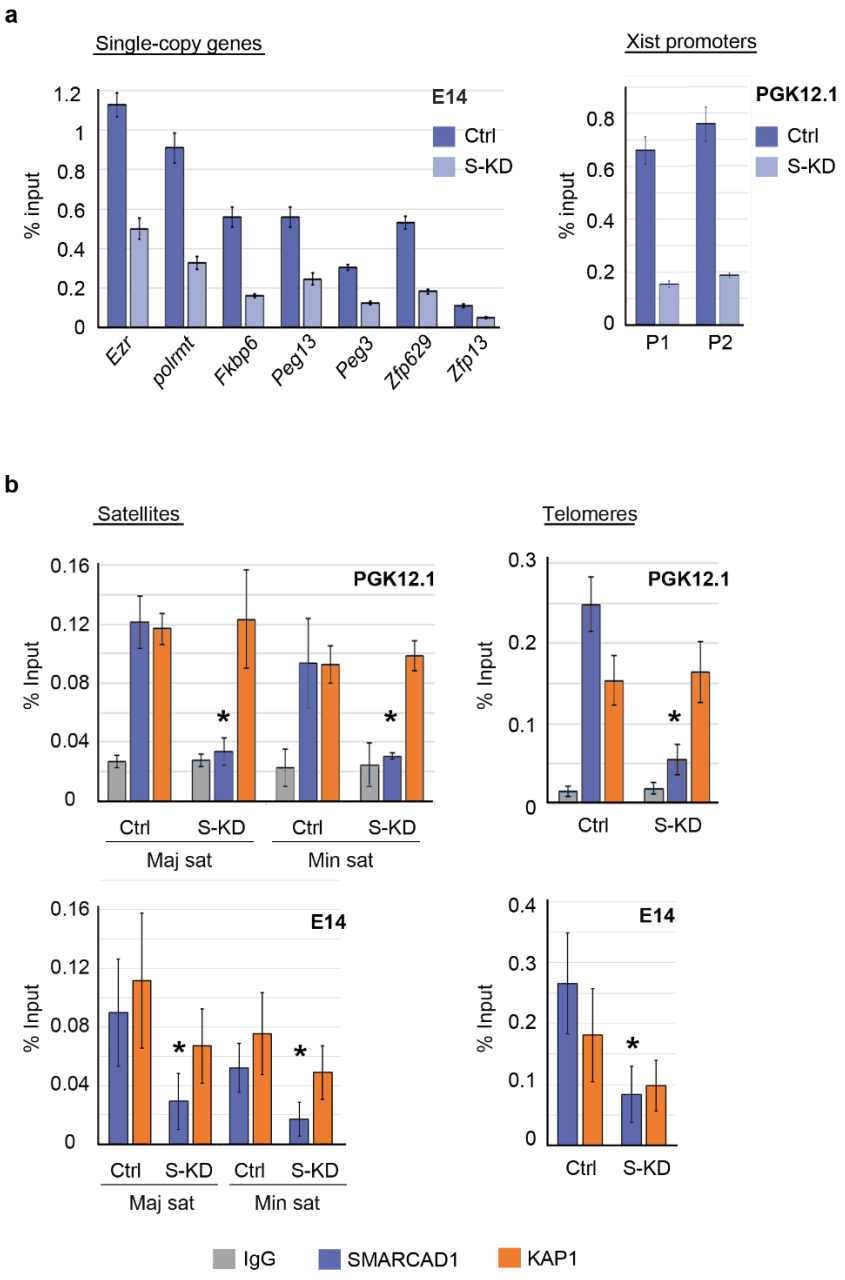

Figure 3.2.2: SMARCAD1 binds single-copy genes and repetitive elements.

(a) Left panel: ChIP-qPCR of SMARCAD1 in E14 control and two-day transient knockdown mES cells. Shown are different examples of single-copy genes bound by SMARCAD1, among them imprinted genes (*Fkbp6, Peg13 and Peg3*), KRAB-Zinc Finger Protein genes (*Zfp629 and Zfp13*). Right Panel: ChIP-qPCR of SMARCAD1 in PGK12.1 stable control and knockdown XX mES cells over the two promoters of the *Xist* gene. Depicted is one representative of two biological replicates (n=2). Enrichment is shown as percent input (+/-SE). (b) SMARCAD1 and KAP1 binding to major and minor satellites (left panels) or telomeres (right panels) in E14 control and two-day transient SMARCAD1 knockdown and in PGK12.1 stable control and SMARCAD1 knockdown mES cells. Depicted is the mean ± S.E. from three biological replicates (n=3) shown as percent input. *P* values are from paired two-tailed Student's *t*-test: $*p < 0.05$.

Repetitive regions are characteristic of constitutive heterochromatin. Among these SMARCAD1 binds telomeres (Figure 3.2.2, upper and lower right panels), major and minor satellites, which make up the centric heterochromatin in the mouse (Figure 3.2.2, upper and lower left panels), and transposons (see next section 3.2.2.).

KAP1 is a stochiometric component of SMARCAD1 complexes, and indeed directly interacts with SMARCAD1 *in vitro* (Ding et al., 2018). Hence, it was tested whether these proteins co-localize in chromatin. Most of the focus of this study will be on the role of SMARCAD1 on transposons, including the stable association of KAP1 on these elements. However, other sites were briefly investigated; KAP1 was found enriched on satellites and telomeres. SMARCAD1 likely plays varying roles on its binding sites and as a first step, it was addressed whether KAP1 binding is affected upon SMARCAD1 loss on different sites. For KAP1 binding on telomeres and major and minor satellites SMARCAD1 appears to play a minor role: in the stable SMARCAD1 knockdown cell line (PGK12.1) KAP1 binding is clearly unaffected in the knockdown (Figure 3.2.2b, upper panels). The E14 cells contain an induced SMARCAD1 knockdown, here shown after two days of SMARCAD1 removal. KAP1 is not significantly affected but it appears somewhat reduced compared to the stable cell line (Figure 3.2.2b, lower panels), implying that SMARCAD1 might play a minor role in KAP1 recruitment or maintenance on these sites probably concurrent with other factors since KAP1 is successfully maintained in a stable cell line, but an immediate effect can be observed in a transient SMARCAD1 knockdown. In conclusion, on major and minor satellites, and on telomeric regions KAP1 association to chromatin is generally stable irrespective of SMARCAD1 binding.

3.2.2. SMARCAD1 binds class I and II ERVs in mES cells

Genome-wide SMARCAD1 is associated with the transcriptional regulator KAP1, H3K9me3, the hallmark of constitutive heterochromatin, as well as H4K20me3 (Figure 3.1.7). These two histone marks are involved in the silencing of ERVs (Matsui et al., 2010; Rebollo et al., 2011). Not only that, but SMARCAD1 also overlaps with SETDB1 binding sites but not with any of the other histone methyltransferases. SETDB1 catalyses the deposition of H3K9me3 on ERVs (Matsui et al., 2010) and is recruited by KAP1 (Cheng et al., 2014b; Frietze et al., 2010; Iyengar and Farnham, 2011; Schultz et al., 2002). Additionally, the histone variant H3.3 is present on ERVs that are bound by

SMARCAD1. This variant has previously been linked to the silencing of ERVs, particularly IAPs (Elsasser et al., 2015).

Taken together the data point to a potential role of SMARCAD1 on ERVs and thus they were investigated as potential targets for the remodeller SMARCAD1. SMARCAD1 binding was plotted to the murine database of repetitive elements by Boris Lamp, and in an unbiased approach the elements with the highest SMARCAD1 binding were analysed, as well as their subcategories. It was discovered that SMARCAD1 peaks are enriched at Intracisternal A-type particles (IAPs), a category of class II ERVs which are active in the mouse (Gerdes et al., 2016; Qin et al., 2010), (Figure 3.2.3). Similarly, H3K9me3 (this study) and KAP1 (Elsasser et al., 2015) are enriched at IAPs in the mouse, mimicking the distribution of SMARCAD1 (Figure 3.2.3) SMARCAD1 is also enriched on VL30 elements, a class I ERV family. The same binding pattern of SMARCAD1 observed in ChIP-seq can be seen in the ChIP-qPCR data. Notably in the ChIP-seq, SMARCAD1 is not enriched over internal sequences of VL30 elements, (annotated as int, Figure 3.2.3), but can be detected when VL30 elements are plotted as a whole. This is reflected in the qPCR data, where the VL30 primer recognizes the primer binding site within the 5'UTR region, and SMARCAD1 enrichment can be observed. Importantly, both methods are "genome-wide" for repeats, as the ChIP-qPCR uses consensus primers. SMARCAD1 binding is also detected over the class II elements ERVK10C and MusD/ETn, while it is not bound to the class III element MuERVL (Figure 3.2.3a).

The non-LTR retrotransposon LINE was not found to be enriched for SMARCAD1 binding. Importantly, this element possesses a very high copy number in the mouse genome (around 600,000 copies, (Wang, 2017)), and yet no SMARCAD1 binding was found, illustrating that the enrichment over IAP elements is not due to their high copy number and is indeed specific. Additionally, the LINE element L1Md_F was investigated for SMARCAD1 binding. L1Md_F is a distinct L1 lineage that appears in the mouse genome around 4.35 million years ago (Castro-Diaz et al., 2014). In contrast to other L1 elements which do not generally require KAP1 for their repression, L1Md_F has been shown to be repressed by KAP1 and KRAB-ZFP mechanisms in both human and mouse ES cells, giving rise to the model that newly emerging L1 lineages are first suppressed by DNA-methylation mechanisms before specific KAP1-recruiting ZFPs are selected (Castro-Diaz et al., 2014). However, no SMARCAD1 binding was detected on LINE1 elements, including L1Md_F in PGK12.1 mouse ES cells by ChIP-seq and ChIP-qPCR

(Figure 3.2.3). KAP1 binding, while above an intergenic control site (background), is still relatively low. Nonetheless, H3K9me3 is clearly enriched on L1Md_F elements compared to other LINE1 sites. This suggests that SMARCAD1 does not play a role in LINE1 repression, however it might be worth to investigate L1Md_F in other cell lines, since KAP1 binding is also relatively low in PGK12.1 mES cells.

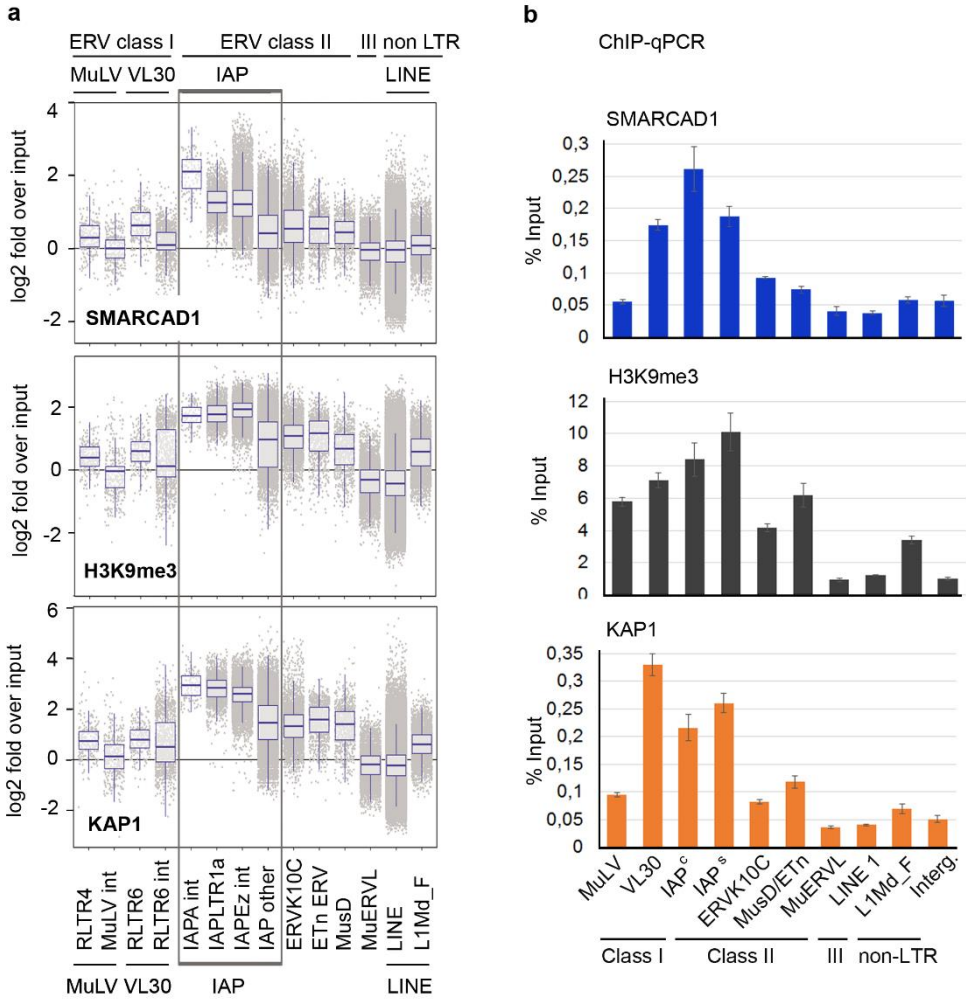

Figure 3.2.3: SMARCAD1 binds lass I & II ERV elements together with KAP1 and H3K9me3.

(a) Combined box- and jitterplot depicting the binding of SMARCAD1 (top, this study), H3K9me3 (middle, this study) and KAP1 (bottom, (Elsasser et al., 2015)) as $\log_2$ fold over input over different types of class I, II and III ERV elements as well as non-LTR retrotransposons (LINEs) in PGK12.1 mES cells. Each grey dot represents a single element. Repeat families shown are characterized according to the UCSC RepeatMasker annotation. "int" represents an internal ERV element. General annotations such as "IAP" or "ERVK10C" denote all elements of that type. IAPA_MM-int, IAPEz-int and IAPLTR1a are the most enriched elements, followed by "IAP other" which represents all other IAP elements. (b) Representative ChIP-qPCR of three biological

replicates of SMARCAD1, H3K9me3 and KAP1 over class I, II and III ERV elements as well as non-LTR retrotransposons in PGK12.1 mES cells. This dataset is used again in Figure 3.3.1 for SMARCAD1 and KAP1 and Figure 3.3.2 for H3K9me3 as one of three biological replicates. All repeat annotations denote consensus primers for all elements of that category except IAP^{s} which represents a specific IAP element within the *Mier3* gene (IAP^{c} represents the consensus primer). Also shown is a subclass of LINE elements, L1Md_F, where high KAP1 binding was reported (Castro-Diaz et al., 2014). An intergenic single-copy site is shown as control. Enrichment is shown as percent input (+/- SE).

SMARCAD1, KAP1 and H3K9me3 enrichment overlap in their peak distribution as shown here in exemplary screenshots on two examples: IAPEz-int element within an exon of the *Mier3* gene (Figure 3.2.4a) and on an IAP element within the *Prnp* gene (Figure 3.2.4b). Regions bound by SMARCAD1 are very broad in general, almost 20kb is some cases, which is very distinct from the precise peaks found in transcription factors, and more characteristic of histone marks. The IAP element within the *Mier3* gene is flanked by its LTRs, whereas the one within the *Prnp* gene denotes a truncated version of an IAP element. The element is next to an MERVL element, noticeably with no SMARCAD1, KAP1 or H3K9me3 enrichment (Figure 3.2.4b).

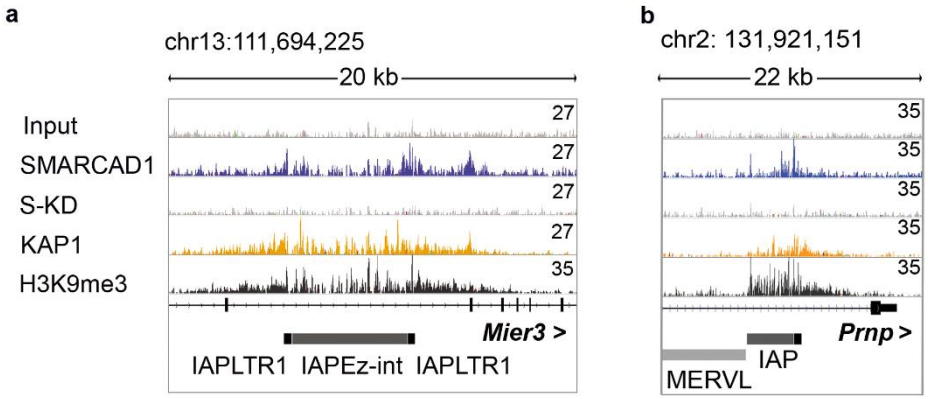

Figure 3.2.4: SMARCAD1, KAP1 and H3K9me3 bind large regions containing repetitive elements.

Screenshots depicting SMARCAD1, KAP1 and H3K9me3 called peaks in the IGV Genome Browser over IAP elements. (a) displays an element of the subcategory IAPEz flanked by its LTR domains, within an exon of the *Mier3* gene. (b) displays and IAP element within the *Prnp* gene. On the right it has its LTR domain and on the left, it is truncated by an MERVL element, a class III ERV. Chromosomal position is indicated. The arrowhead indicates the direction of gene transcription.

A challenge in the repeat field is to identify individual copies. To overcome this, primers for ChIP-qPCR were designed and tested to identify SMARCAD1 binding. Primers were devised to lie just a few nucleotides outside an IAP element, thus denoting a single, specific binding event, here shown for the SMARCAD1 peak within the *Mier3* gene (Figure 3.2.5a). The other primers employed are consensus primers encompassing either all IAP elements (IAP primer), or a subset (IAPEz, IAP1). IAPEz is the most abundant subcategory of IAPs found within or near genes in the mouse exhibiting approximately 200 insertions, predominantly intronic (Maksakova et al., 2006). SMARCAD1 (this work) and KAP1 (Elsasser et al., 2015) are both enriched over IAPEz elements (Figure 3.2.5a). IAP1 is the most recent IAP element that has emerged in the murine genome and in neuronal progenitor cells its expression is deregulated upon the loss of KAP1 (Fasching et al., 2015). In mES cells however, it shows no SMARCAD1 or KAP1 enrichment (Figure 3.2.5a), showcasing an example of tissue-specific IAP regulation. It was previously suggested that IAPs take on specific regulatory roles later in development and require distinct regulatory processes for certain subcategories.

ERVs in general are genomic elements that structurally retain their resemblance to retroviruses, but their structure is rarely completely intact, most ERVs are truncated and the *env* gene is very frequently lost. Solitary LTRs are also quite common (Cohen et al., 2009; Goke and Ng, 2016; Kassiotis, 2014). With consensus primers for distinct structural elements of IAPs it is possible to determine the distribution of SMARCAD1 over IAPs genome-wide in mouse ES cells. Figure 3.2.5b reveals that the highest SMARCAD1 enrichment is found at the 5'UTR followed by the *gag* gene; a distribution also exhibited by KAP1 in this study and in Rowe et al., 2010.

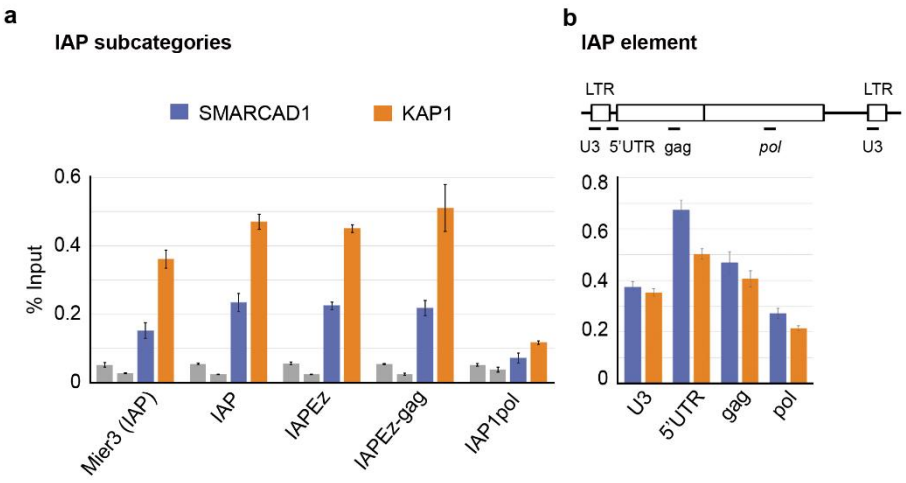

Figure 3.2.5: SMARCAD1 and KAP1 bind distinct IAP subcategories and show a similar enrichment pattern over IAP elements.

(a) ChIP-qPCR in E14 mES cells of SMARCAD1 and KAP1 over different IAP families, a specific IAP element within the *Mier3* gene, all IAP elements, the subcategory IAPEz, the *gag* sequence within IAPEz elements as well as the young IAP element IAP1pol. Enrichment is shown as percent input (+/-SE). (b) Top: IAP structure depicting the primers used in the ChIP-qPCR. Bottom: ChIP-qPCR in E14 mES cells showing SMARCAD1 and KAP1 distribution over IAP elements genome-wide. Enrichment is shown as percent input (+/-SE).

3.2.3. SMARCAD1 and KAP1 are co-enriched at all tested SMARCAD1 binding sites in mESCs

The data obtained in this study illustrate that SMARCAD1 and KAP1 share target sites. Among them imprinted genes, KRAB-ZFP genes, satellites, telomeres and class I and II ERV elements. It is not known whether they occupy their shared targets simultaneously or independently. To address this, sequential ChIP on representative binding sites was performed from both directions; a SMARCAD1 ChIP followed by KAP1 and vice versa (Figure 3.2.6). The results reveal that the two proteins do indeed co-occupy the 5 tested sites of SMARCAD1-bound elements; promoters of imprinted genes, the 3'end of ZFPs, *Ezr*, and retrotransposons (Figure 3.2.6). This strongly suggests a functional link between SMARCAD1 and KAP1.

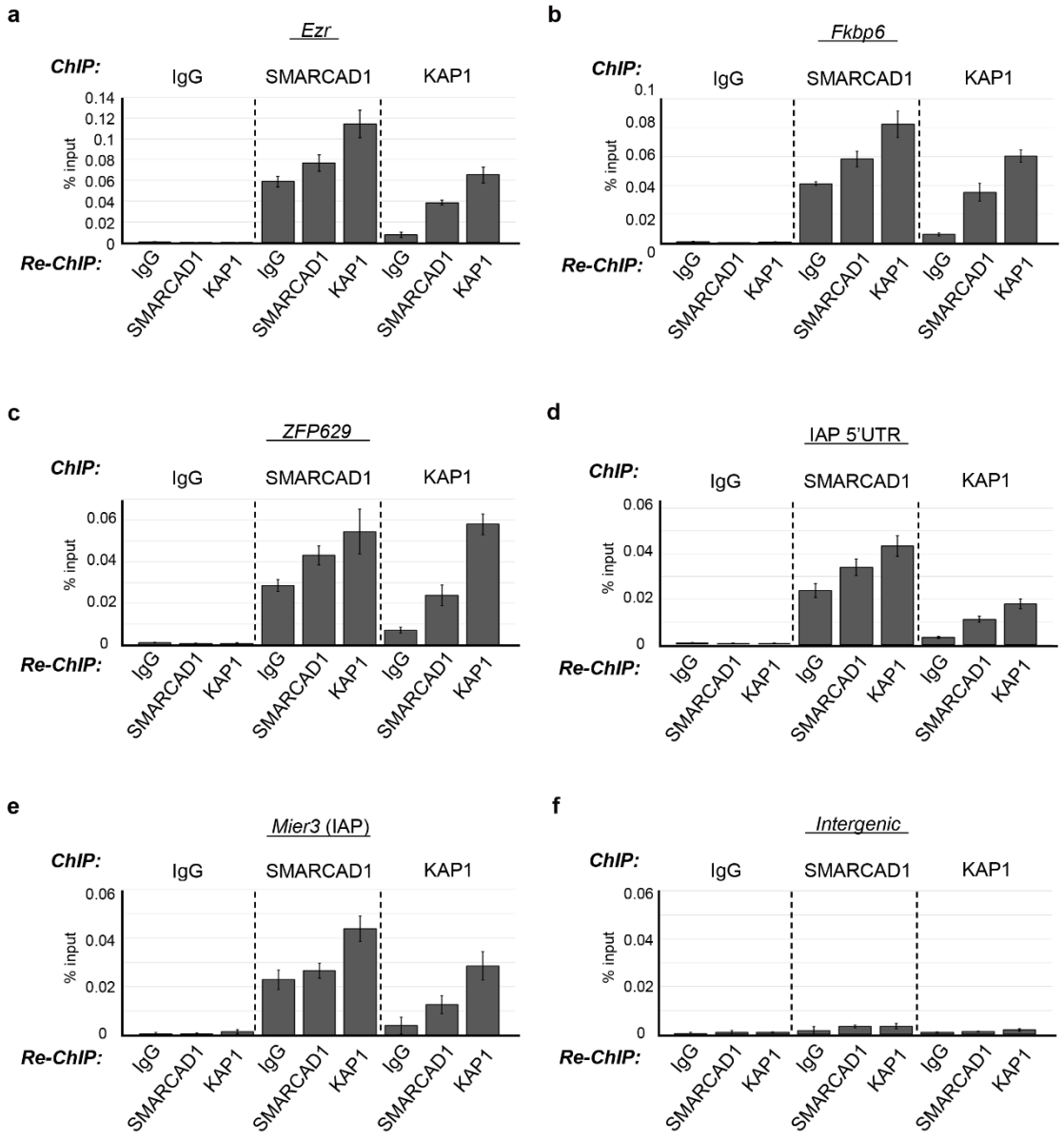

Figure 3.2.6: SMARCAD1 and KAP1 co-occupy exemplary SMARCAD1 binding sites.

Depicted is sequential ChIP to illustrate the co-occupancy of SMARCAD1 and KAP1 (a) within a single-copy gene, (b) on the promoter of an imprinted gene and (c) on the 3' end of a KRAB-Zinc Finger gene, (d) a consensus IAP element, (e) a specific IAP element and (f) an intergenic site is shown as negative control. The first ChIP is indicated on the top of each panel, followed by the re-ChIP annotated underneath. IgG is used as a control for both steps. Enrichment is shown as percent input (+/- SE, n=2).

3.2.4. SMARCAD1 is recruited to chromatin via its CUE1 domain-mediated interaction with KAP1

SMARCAD1 binding sites are co-enriched with KAP1 (Figure 3.1.7 and Figure 3.2.6), both for single-copy genes and repetitive regions. To investigate the order of recruitment to chromatin the strategy was to investigate the binding of SMARCAD1 and KAP1 to chromatin in the absence of their respective partner protein. The effect of the SMARCAD1 knockdown is discussed in chapter 3.3.1.

Initially a transient KAP1 knockdown was performed to determine whether SMARCAD1 association to chromatin is dependent on the presence of KAP1. KAP1 protein levels were successfully reduced (Figure 3.2.7). However, this was accompanied by a reduction of the global protein levels of SMARCAD1 (Figure 3.2.7). This effect was subsequently studied in depth by a colleague in the lab in three cell lines (Ding et al., 2018). KAP1 knockdown had an effect on SMARCAD1 at the RNA level, even when an exogenous SMARCAD1 construct under the chicken β actin promoter was employed (Ding et al., 2018). The KAP1 knockdown could therefore not be confidently used to determine SMARCAD1 binding on chromatin and the sequence of binding events. Instead, the interaction between KAP1 and SMARCAD1 was disrupted.

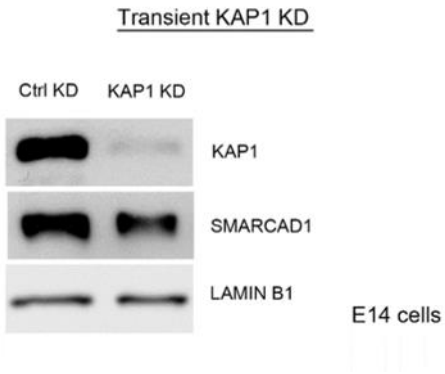

Figure 3.2.7: SMARCAD1 protein levels are reduced globally after 4 days of KAP1 knockdown.

Western blot of a 4d transient KAP1 knockdown in mESCs from global whole cell protein extracts. KAP1 knockdown is known to affect the pluripotency of mES cells (Cheng et al., 2014a), accordingly Nanog is seen reduced. SMARCAD1 protein levels are also affected, an effect subsequently well documented by Dong Ding (Ding et al., 2018). Lamin B1 is shown as loading control.

The direct interaction of SMARCAD1 and KAP1 was mapped by colleagues in this laboratory to the RING B-box coiled-coil (RBCC) domain of KAP1 and to the proximal coupling of ubiquitin conjugation to ER degradation (CUE1) domain of SMARCAD1 (Ding et al., 2018) (Figure 3.2.8a). Mutations in two key residues of the CUE1 domain abolish the interaction of SMARCAD1 and KAP1 *in vitro* and severely perturb it *in vivo*. A cell line was established in this laboratory containing the CUE1-mutant construct over a doxycycline-induced knockdown of endogenous SMARCAD1. The CUE1-mutant is expressed at the same level than a WT-construct in the cells (Figure 3.2.8b).

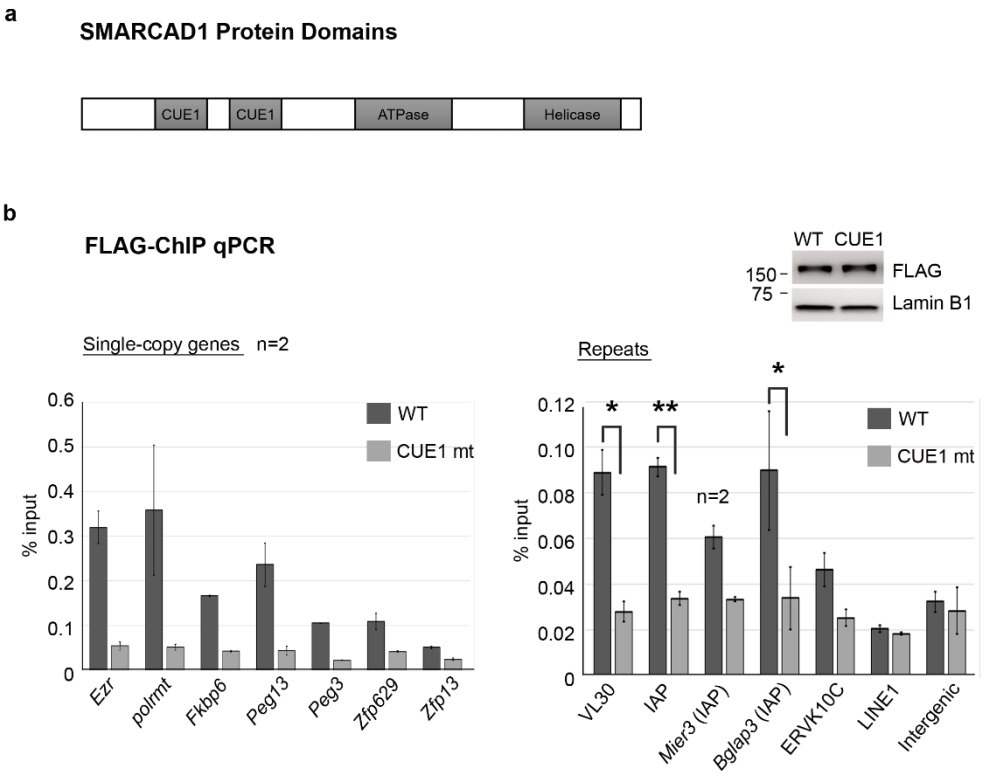

Figure 3.2.8: SMARCAD1 is recruited to chromatin via its CUE1 domain.

(a) Cartoon depicting the domains of SMARCAD1. (b) Left panel: ChIP-qPCR of a FLAG-SMARCAD1 (WT) and a FLAG-SMARCAD1 CUE1 mutant (CUE1 mt) over different subcategories of single-copy genes bound by SMARCAD1 in E14 mES cells (Imprinted genes, *Zfp*-genes, *Ezr*, and *Polrmt*). Shown is the mean enrichment of two biological replicates (+/-SE) as percent input. Right panel: WT and CUE1 mt binding to different subcategories of ERVs (VL30: class I, IAP, *Mier3* (IAP), *Bglap3* (IAP) and ERVK10C: class II), and to LINE1 elements in E14 mES cells. An intergenic site is shown as negative control. Shown is the mean enrichment of three biological replicates (+/-SE). except for *Mier3* (IAP), n=2. *P* values are from paired two-tailed Student's *t*-test: $*p < 0.05$, $**p < 0.01$. The Western blot on top illustrates equal expression levels of the WT and mutant construct in mES cells, expressed over an inducible knockdown of SMARCAD1.

SMARCAD1 CUE1 mutant provided the opportunity to investigate whether SMARCAD1 binding to its target sites is dependent upon its interaction with KAP1 via the CUE1 domain. Consistent with this model it was observed that the SMARCAD1 CUE1 domain is essential for tethering SMARCAD1 to the nucleus (Ding et al., 2018). The binding to chromatin of the mutant compared to a wild-type construct was severely reduced on both single-copy genes (Figure 3.2.8, left panel) as well as on repetitive elements, specifically different types of ERVs (Figure 3.2.8, right panel). This demonstrates that SMARCAD1 recruitment to chromatin is dependent on an intact CUE1 domain and likely its interaction with KAP1.

Collectively, the top SMARCAD1 binding sites are predominantly in heterochromatic regions, both facultative and constitutive. Many of these SMARCAD1 peaks are co-enriched with KAP1 as shown by sequential ChIP on exemplary SMARCAD1 binding sites. The recruitment of SMARCAD1 to chromatin was investigated via a KAP1-interaction mutant, since a KAP1 knockdown proved to result in a double SMARCAD1-KAP1 knockdown, and it was determined that SMARCAD1 is indeed recruited to the investigated binding sites by its interaction with KAP1.

3.3. SMARCAD1 remodelling is required for the association of repressive factors and histone modifications to class I and II ERVs

SMARCAD1 binds to class I and II ERVs together with KAP1, exhibiting the highest enrichment on IAPs. Class I and II ERVs are also bound by H3K9me3, in this study as well as in other H3K9me3 published datasets (Bulut-Karslioglu et al., 2014). In addition, the interaction of SMARCAD1 and KAP1 is required for the binding of SMARCAD1, on both single-copy targets and repetitive regions. If the interaction is interrupted by the CUE1 mutation in SMARCAD1, it is no longer stably bound to the tested targets.

3.3.1. KAP1 and the histone methyltransferase SETBD1 are reduced over class II ERVs upon the loss of SMARCAD1

To address the function of SMARCAD1 on ERVs, a loss-of-function approach was taken. What effect does the loss of SMARCAD1 have on KAP1 association and other factors involved in the establishment of heterochromatin on these sites, specifically the histone methyltransferase SETDB1? This question was investigated in two different mES cell lines, female and male, and under two different conditions of knockdown, stable and transient. Hence, any effects observed are not cell line specific. Additionally, the transient knockdown will reveal immediate effects of SMARCAD1 removal which might have been compensated for in the stable knockdown.

ChIP-qPCR experiments in the stable knockdown show that SMARCAD1 is necessary for the binding or stabilization of KAP1 on ERVs (Figure 3.3.1a, b). KAP1 binding is significantly reduced over IAP elements as depicted by the consensus primer (Figure 3.3.1a) as well as on a specific IAP element investigated. Primers for different structural elements of IAPs highlight this decrease over the entire element (Figure 3.3.1b). Other class II ERVs were investigated, such as ERVK10C, show a slight but not significant reduction of KAP1. Class I ERV elements display SMARCAD1 binding, however KAP1 is not affected upon its removal on these sites in the stable SMARCAD1 knockdown. Class III elements show very little to no SMARCAD1 binding and KAP1 is unchanged upon SMARCAD1 loss, similarly to what was observed on non-LTR retrotransposons such as LINEs. The transient SMARCAD1 knockdown is discussed in more detail in section 3.3.5, briefly; KAP1 is significantly reduced over ERVK10C elements and also

shows a slight but not significant reduction over VL30 elements not seen in the stable knockdown (Figure 3.3.8).

KAP1 serves as a recruiting platform for several factors important for ERV control (Cheng et al., 2014b; Iyengar and Farnham, 2011; Jang et al., 2018; Rowe et al., 2010). Among them the histone methyltransferase SETDB1. In order to determine the place of SMARCAD1 in this sequence of binding events, SETDB1 binding over ERVs was investigated upon SMARCAD1 knockdown. The result is that SETDB1 is reduced over consensus and specific IAP sites and over the class II ERV element MusD/ETn, which displays no KAP1 reduction upon SMARCAD1 loss (Figure 3.3.1c). As seen with KAP1, SETDB1 is primarily decreased over the 5'UTR of IAP elements (Figure 3.3.1d). The global protein levels of KAP1 and SETDB1 are unchanged upon the knockdown of SMARCAD1 as shown earlier (Figure 3.1.9).

Taken together, this implies SMARCAD1 is required for the association of KAP1 to class II ERVs and in the sequence of binding events it is recruited prior to SETDB1, since SETDB1 is reduced upon the loss of SMARCAD1. On IAP and ERVK10C elements both KAP1 and SETDB1 are affected, whereas on the class II element MusD/ETn, only SETDB1 is reduced (Figure 3.3.1) possibly suggesting distinct mechanisms on different sites.

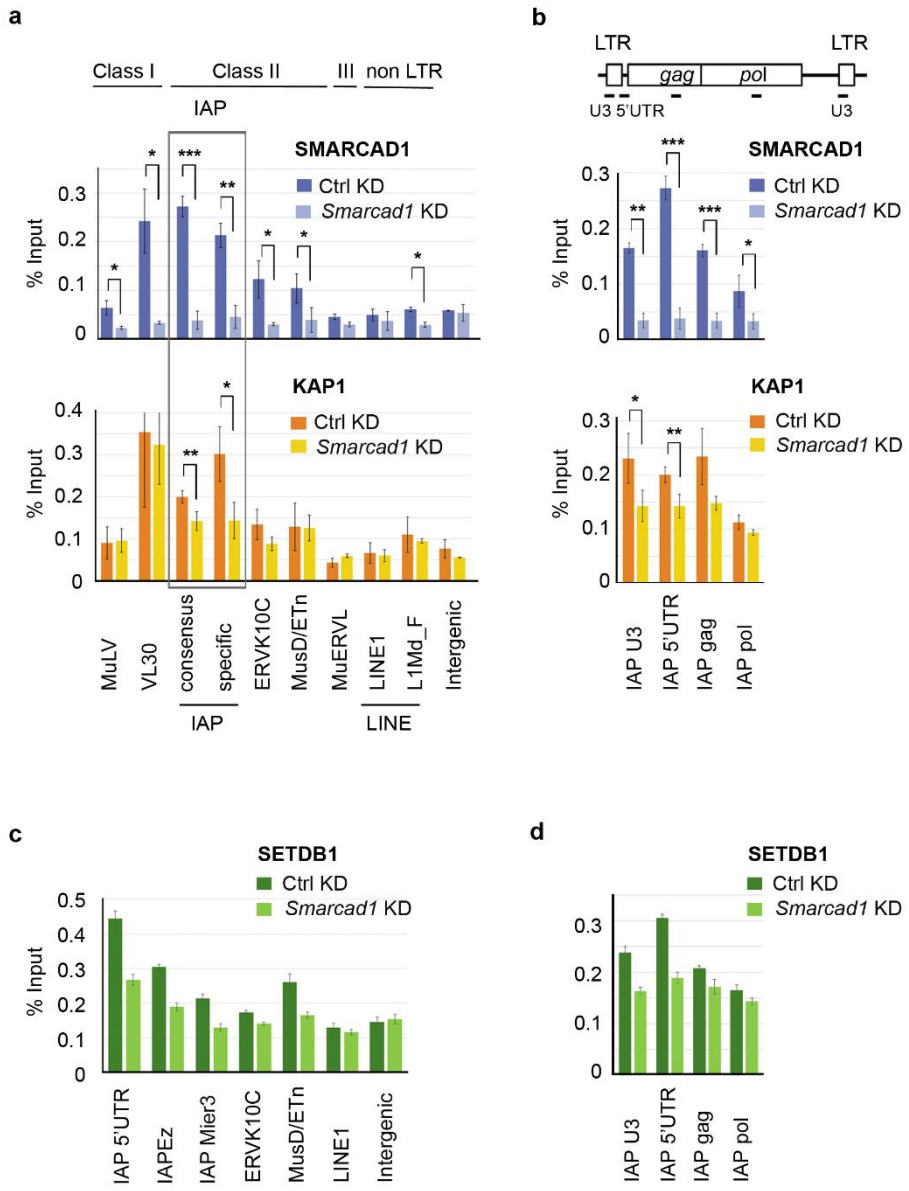

Figure 3.3.1: The ERV regulators KAP1 and SETDB1 are reduced upon the loss of SMARCAD1 at class II ERVs.

(a) ChIP-qPCR of SMARCAD1 and KAP1 binding over different ERV classes in control and stable SMARCAD1 knockdown mES cells (PGK12.1). Depicted is the mean enrichment of three biological replicates (+/-S.E.). An intergenic site is shown as negative control. Enrichment is shown in percent input. (b) ChIP-qPCR of SMARCAD1 and KAP1 binding over different structural elements of IAPs (as depicted in the cartoon on top) in control and stable SMARCAD1 knockdown mES cells (PGK12.1). Depicted is the mean enrichment of three biological replicates (+/-S.E.). *P* values are from paired two-tailed Student's *t*-test: $*p < 0.05$, $**p < 0.01$, $***p < 0.001$ (c) representative ChIP-qPCR of two biological replicates showing SETDB1 binding over different ERV classes upon stable knockdown of SMARCAD1 and control in mES cells (PGK12.1). Enrichment is shown in percent input (+/-S.E.). (d) SETDB1 binding over

different structural elements of IAP in control and stable SMARCAD1 knockdown mES cells (PGK12.1). Representative example of two biological replicates. Enrichment is shown in percent input (+/-S.E.).

3.3.2. Repressive histone marks H3K9me3 and H4K20me3 are decreased in SMARCAD1 knockdown on class II ERVs

SETDB1 association to chromatin is reduced upon the loss of SMARCAD1. It being the histone methyltransferase that catalyses H3K9me3, this modification was investigated next. Upon the loss of SMARCAD1 the global protein levels of H3 and H3K9me3 are unaffected (Figure 3.1.9), as well as global levels of H4K20me3 as shown by colleagues repeatedly (Sachs et al., 2019). However, a significant reduction of H3K9me3 was observed over IAP elements, this effect was seen with consensus primers and over the span of the entire element (Figure 3.3.2a, b) and was particularly pronounced over specific IAP elements (Figure 3.3.2a). This was confirmed in a second mES cell line (Figure 3.3.9). The repressive histone mark H4K20me3 was similarly reduced in two different SMARCAD1 knockdown cell lines over IAP elements. Compared to H3K9me3 the decrease of H4K20me3 on IAPs genome-wide is mild, on two specific IAPs however it is very pronounced (Figure 3.3.2c). H3K9me3 is also mildly, but not significantly, reduced over MuLV, ERVK10C, and MusD/ETn. The small effect on ERVK10C elements is not surprising since it is bound by SMARCAD1, and KAP1 is also mildly reduced. However, MuLV displays very little or no SMARCAD1 binding, and MusD/ETn exhibits a reduction of SETDB1 upon stable SMARCAD1 knockdown, but not KAP1. Overall, this suggests the reduced H3K9me3 is not always concomitant of reduced KAP1, but on certain elements a reduction of SMARCAD1 alone results in less H3K9me3, whereas on the majority of sites both SMARCAD1 and KAP1 play a role in the stable association of SETDB1 and subsequently H3K9me3.

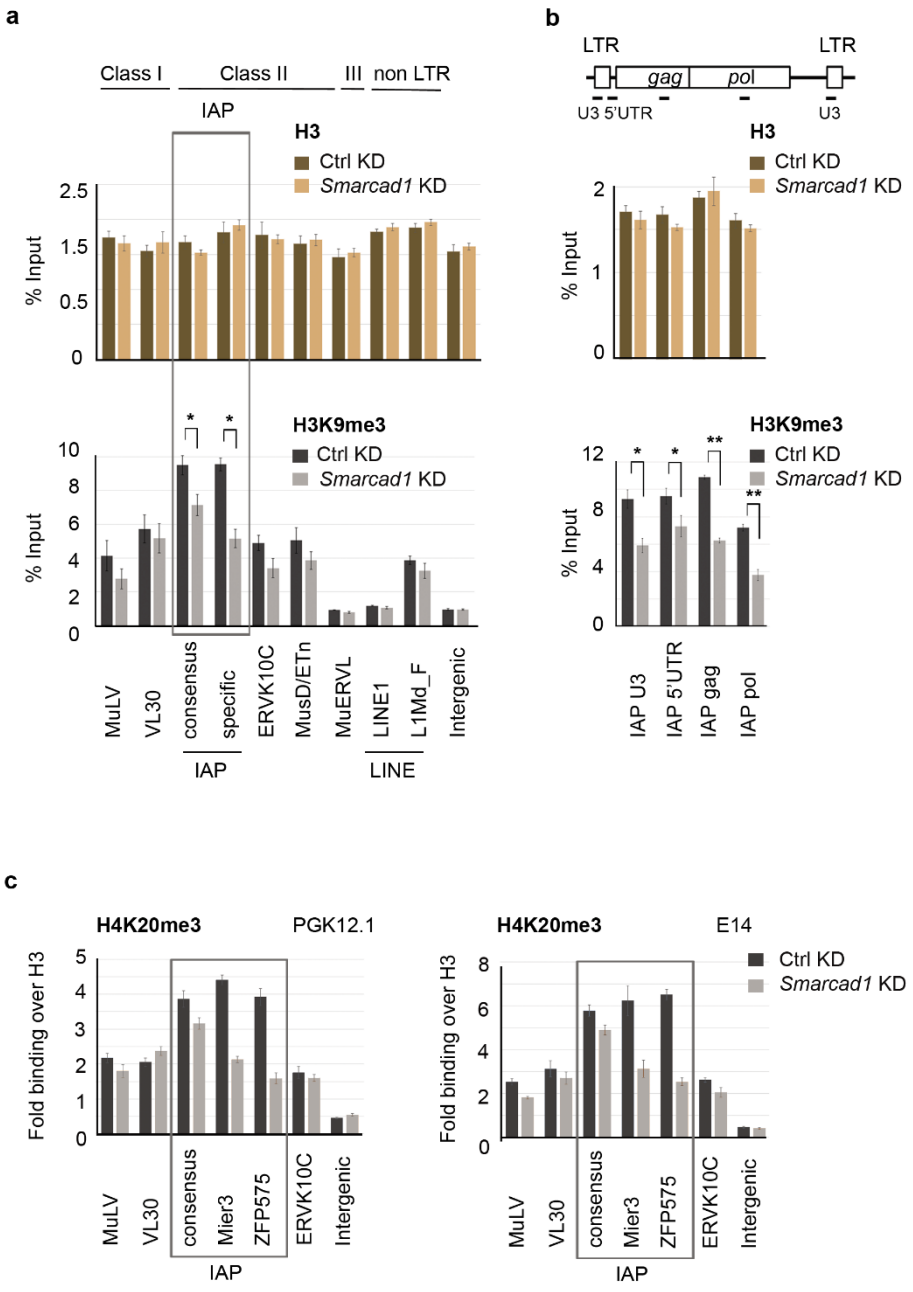

Figure 3.3.2: The repressive histone marks H3K9me3 and H4K20me3 are decreased upon loss of SMARCAD1 over class II ERVs.

(a) ChIP-qPCR of H3 and H3K9me3 binding over different ERV classes in control and stable SMARCAD1 knockdown mES cells (PGK12.1). Depicted is the mean enrichment of three biological replicates (+/-S.E.). An intergenic site is shown as negative control. Enrichment is shown in percent input. (b) ChIP-qPCR of H3 and H3K9me3 binding over different structural elements of IAPs (as depicted in the cartoon on top) in control and stable SMARCAD1 knockdown mES cells (PGK12.1). Depicted is the mean enrichment of three biological replicates (+/-S.E.). *P* values are from paired two-tailed Student's *t*-test: $*p < 0.05$, $**p < 0.01$, $***p < 0.001$. (c) ChIP-qPCR showing H4K20me3 binding over different ERV classes in two

different cell lines, the left panel depicts a stable SMARCAD1 knockdown and the right a transient knockdown. Enrichment is shown as fold binding over H3 (+/-S.E.).

A comparable reduction of KAP1 and H3K9me3 is observed upon the transient removal of SMARCAD1 (Figure 3.3.8 and Figure 3.3.9). KAP1 is significantly reduced over IAP elements and over ERVK10C. In contrast to the stable knockdown, it is additionally reduced over the class I ERV VL30, yet not significantly, perhaps suggesting a redundant mechanism of regulation in the stable knockdown that does not initiate quickly enough in the transient knockdown (Figure 3.3.8). H3K9me3 was also investigated after 2 days of SMARCAD1 knockdown, however almost no H3K9me3 reduction was observed at this time point (data not shown) likely because it is passively lost throughout cell divisions. After 4 days of SMARCAD1 loss the effects on H3K9me3 association were particularly noticeable on three specific IAP elements but also on ERVK10C and importantly, H3K9me3 was recovered upon introduction of exogenous SMARCAD1 (Figure 3.3.9)

3.3.3. SMARCAD1 is required to retain the histone variant H3.3 over class I and II ERVs

The histone variant H3.3 has generally been linked to regions of high nucleosome turnover and has been associated with gene activation (Ahmad and Henikoff, 2002; Goldberg et al., 2010). It has however, also been shown that this variant localizes to facultative and constitutive heterochromatin and is enriched on repetitive regions, particularly class I and II ERVs, together with H3K9me3 and SETDB1 (Elsasser et al., 2015). Loss of H3.3 resulted in decreased H3K9me3 over ERVs and a moderate deregulation of IAPs (Elsasser et al., 2015).

Thus, the presence of this histone variant was investigated on ERVs upon the loss of SMARCAD1. Notably, the loss of SMARCAD1 has no impact on the global protein levels of H3.3 Figure 3.3.3a). Nevertheless, locally on chromatin, stable association of H3.3 is affected in both stable and transient SMARCAD1 knockdown (PGK12,1 and E14 cells, respectively, (Figure 3.3.3b). The effect is most predominant on IAP elements, genome-wide (consensus IAP primer) as well as on specific IAP elements (within the *Bglap3*, *Mier3* and *Zfp575* genes). H3.3 binding is also affected on ERVK10C, another

ERV class II element, and over the class I element VL30 in the transient SMARCAD1 knockdown (E14 cells) as seen for KAP1. The role of ATRX in the regulation of ERVs has been debated extensively (He et al., 2015; Hoelper et al., 2017; Sadic et al., 2015). In this project, ATRX was not found enriched on IAP elements while it was found bound at high levels to telomeres in the same experiment (Figure 3.3.3c). The possibility remains that this is cell line specific.

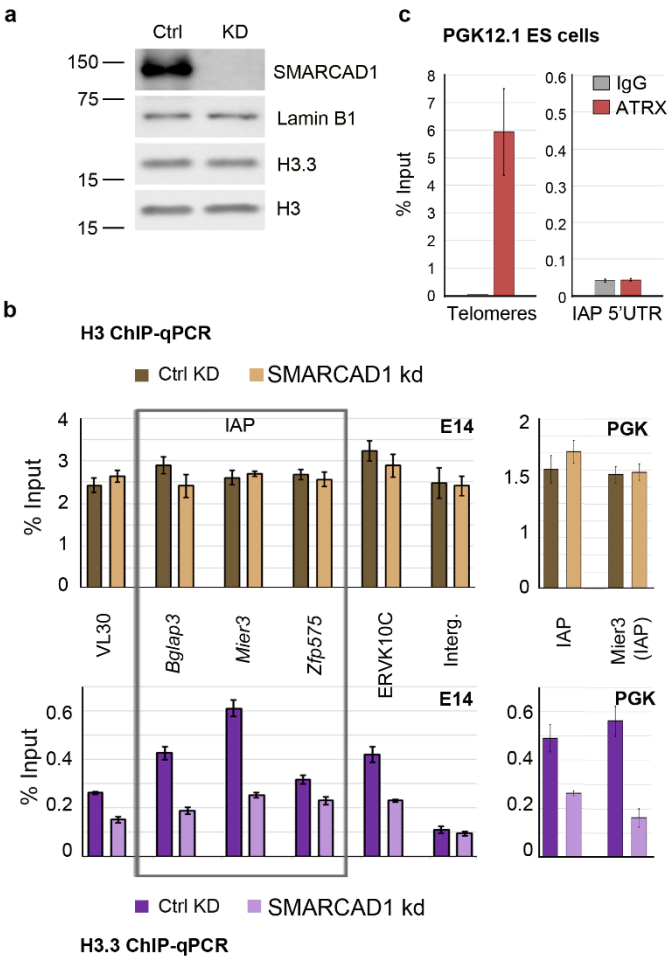

Figure 3.3.3: The histone variant H3.3 is reduced over ERVs upon the loss of SMARCAD1. (a) Western blot of stable SMARCAD1 knockdown in PGK12.1 mES cells showing H3.3 is not globally affected. Lamin B1 and H3 are shown as loading controls. (b) ChIP-qPCR depicting H3.3 binding (lower panels) with H3 ChIP-qPCR as a negative control (upper panels) in a 4-day transient SMARCAD1 knockdown (right panels, E14, n=2) and in the stable knockdown (left panels, PGK, n=2). Class I (VL30) and class II (IAP, ERVK10C) ERVs are depicted. An intergenic site is shown as negative control. Enrichment is shown as percent input (+/-S.E.). (c) ATRX does no bind the 5'UTR of IAP elements in PGK12.1 ES cells. ChIP-qPCR depicting ATRX over telomeric regions and IAP elements from the same experiment (n=1). Enrichment is shown as percent input (+/- SE) and IgG is depicted as a negative binding control.

3.3.4. *The loss of SETDB1 has no effect on SMARCAD1 binding*

Whereas for LINE1 elements the histone methyltransferase G9a is required for H3K9me3, on ERVs SETDB1 is the enzyme that is necessary for H3K9me3 deposition and silencing. It is recruited after KAP1 and is crucial for further association of silencing factors, such as HP1, and final repressive histone tail methylation (Cheng et al., 2014b; Iyengar and Farnham, 2011; Maksakova et al., 2013; Matsui et al., 2010; Schultz et al., 2002). Is SMARCAD1 binding affected by the loss of SETDB1 and subsequent reduction of H3K9me3 on chromatin?

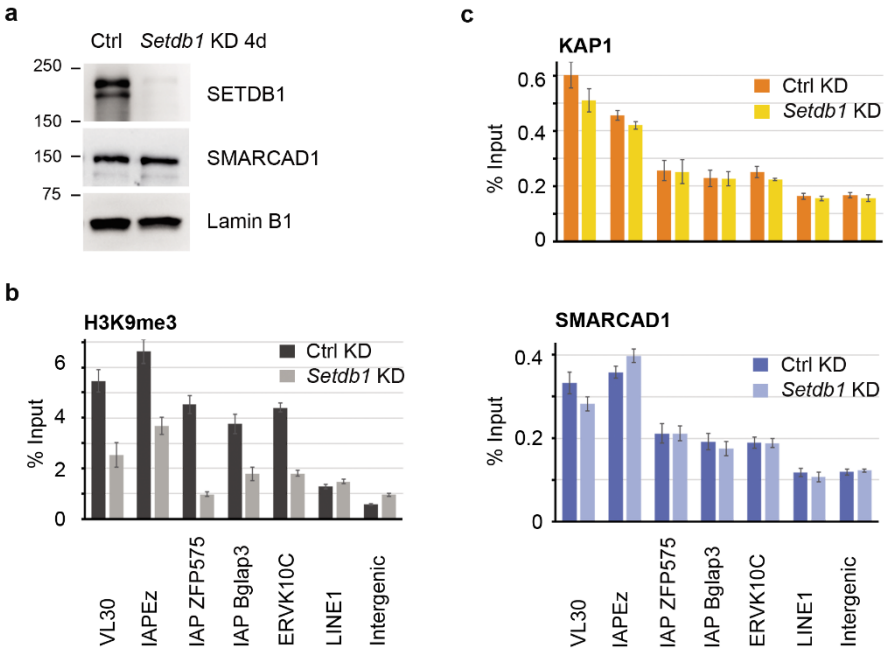

Figure 3.3.4: SMARCAD1 binding over ERVs is unaffected upon the loss of SETDB1.
(a) Western blot showing a 4-day transient SETDB1 knockdown in E14 mES cells. Global SMARCAD1 expression is unaffected. Lamin B1 is shown as negative control. (b) representative ChIP-qPCR of two biological replicates of H3K9me3 over ERVS upon SETDB1 knockdown. LINE1 and an intergenic site are shown as negative controls. (c) representative ChIP-qPCRs of two biological replicates of KAP1 and SMARCAD1 over ERVs upon a 4-day SEDTDB1 knockdown. LINE1 and an intergenic site are shown as negative controls. Enrichment is shown as percent input (+/-S.E.).

A 4-day transient SETDB1 knockdown (Figure 3.3.4a) was established in E14 mouse ES cells which showed the expected reduction of H3K9me3 over class I and II ERVs while LINE1 remains unaffected (Figure 3.3.4b). The Western blot also illustrates that global

protein levels of SMARCAD1 are not decreased. As expected KAP1 binding over ERVs remains at the same levels seen in control, in accord with a sequence of binding events that placed KAP1 before SETDB1 (Rowe et al., 2010; Schultz et al., 2002) (Figure 3.3.4c). Interestingly, SMARCAD1 binding is similarly unaffected upon SETDB1 knockdown and the subsequent decrease of H3K9me3, placing it upstream of SETDB1 and H3K9me3, together with KAP1 binding.

3.3.5. Chromatin remodelling by SMARCAD1 is required for the control of ERVs

SMARCAD1 is required for the stable association of KAP1 (Figure 3.3.1), SETDB1 (Figure 3.3.1), and repressive histone modifications (Figure 3.3.2). Since SMARCAD1 is an ATP-dependent remodeller this raised the question whether the catalytic activity of SMARCAD1 is required for the binding of repressive factors to ERVs.

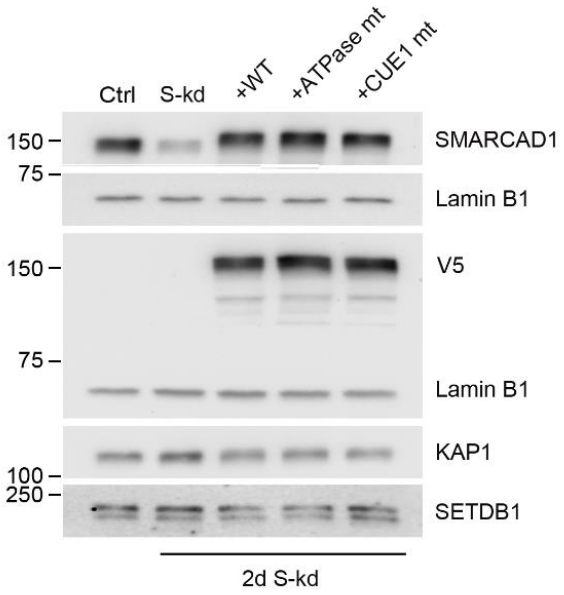

Figure 3.3.5: The SMARCAD1 constructs are expressed equally and KAP1 and SETDB1 protein levels are unaffected.

Western blot from total extracts depicting the stable expression of three different SMARCAD1 constructs on top of a doxycycline-inducible SMARCAD1 knockdown in E14 mES cells. The original control cell line is shown and the inducible SMARCAD1 knockdown in the two lanes on the right. The constructs reintroduce SMARCAD1 expression. V5 illustrates the equal expression of the three constructs. KAP1 and SETDB1 are unaffected. Lamin B1 is shown as a loading control. Molecular weight markers are indicated.

Colleagues in the lab had established an inducible SMARCAD1 knockdown in E14 mES cells and cell lines which reconstituted this knockdown with either the WT SMARCAD1, an ATPase mutant or a CUE1 mutant. Before investigating the effect on the association of repressive factors, it was confirmed that the constructs are expressed at equal levels (Figure 3.3.5). Furthermore, KAP1 and SETDB1 are unaffected in all cell lines used (Figure 3.3.5). This was confirmed by multiple Western blots in studies carried out by colleagues in the lab.

To address the question whether the catalytical activity of SMARCAD1 plays a role on ERVs, ChIP-seq was performed on the WT and ATPase mutant. Indeed, it was found that the mutant binds class I and II ERV elements at significantly lower levels than the WT, most notably IAP elements (Figure 3.3.6b, c) but also other ERV elements such as VL30 and ERVK10C show significantly reduced binding in two ATPase mutants compared to the WT construct (Figure 3.3.6c). This is most clearly seen in the FLAG ChIP-qPCR (Figure 3.3.6c), where the two ATPase mutants bind at very similarly reduced levels, the binding is however not abolished entirely. For simultaneous ChIP (-seq & -qPCR) in a large number of cell lines it was ensured that chromatin shearing gave similar reliable size distributions in order to successfully compare binding across cell lines over multiple experiments (Figure 3.3.6a).

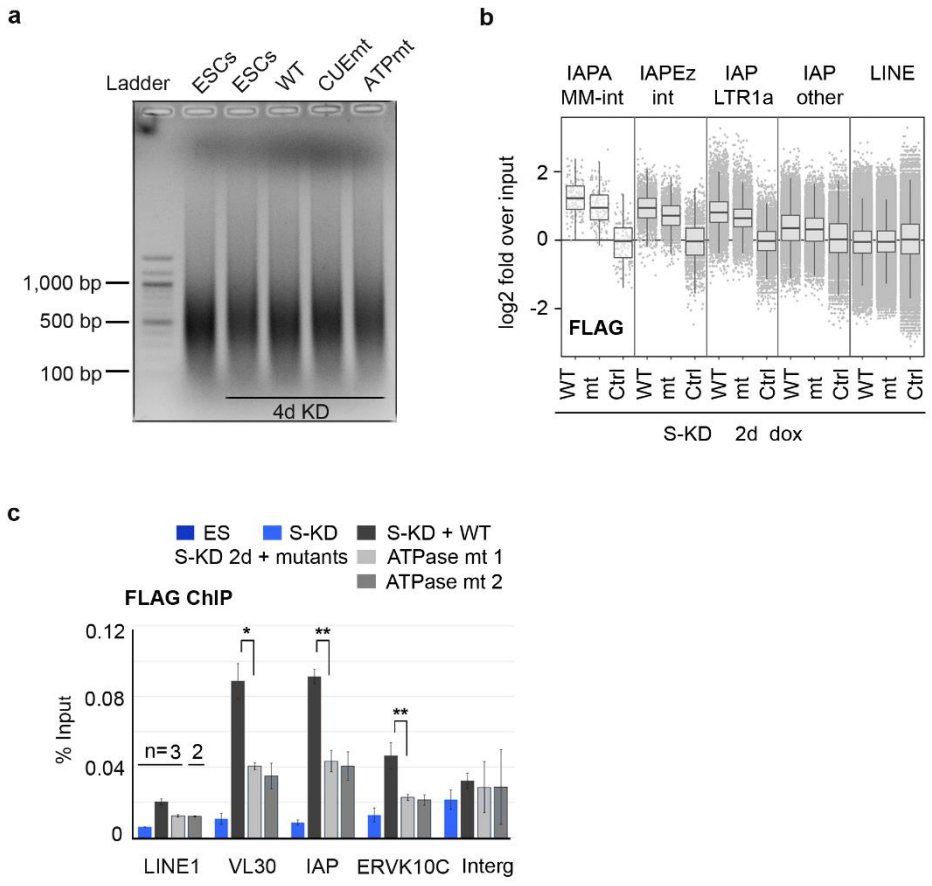

Figure 3.3.6: The SMARCAD1 ATPase mutant is not stably associated to chromatin on ERVs.

(a) Gel electrophoresis depicting the size distribution of sheared chromatin. Loaded on the gel is chromatin from E14 mES cells; control, 4d SMARCAD1 knockdown, and three cell lines containing different SMARCAD1 constructs (WT, CUE1 mutant, and ATPase mutant). A DNA ladder is loaded on the 1st lane from the left and fragment size ranges from 300bp to 900bp approximately, in all five cell lines. (b) Box plot portraying FLAG-SMARCAD1 constructs (WT and ATPase mt) binding to IAP subcategories in a 2d transient SMARCAD1 knockdown. The control contains the SMARCAD1 knockdown and no FLAG-construct. Enrichment is depicted as log2 fold over input. (c) FLAG ChIP-qPCR showing average enrichment from three biological replicates in four different cell lines (+/- S.E.) with the exception of the 2nd ATPase mutant on LINE1 elements (n=2). SMARCAD1 knockdown is shown as a control. The SMARCAD1 WT construct is successfully bound to class I and II ERVs (VL30, IAP and ERVK10C, respectively) whereas the SMARCAD1 ATPase mutant in two different cell lines is not stably bound to the repetitive elements. An intergenic site is shown as a negative control for binding. *P* values are from paired two-tailed Student's *t*-test: $*p < 0.05$, $**p < 0.01$.

This reduced chromatin association of the FLAG-ATPase mutants is reflected in the overall binding of SMARCAD1. SMARCAD1 binding is reduced in its transient knockdown and reconstituted by the WT construct on ERVs. Albeit this reconstitution is not entirely complete in rare cases (Figure 3.3.7 IAP, and data not shown). The two ATPase mutants however do not reconstitute SMARCAD1 association, its binding is even slightly lower than in the transient knockdown (Figure 3.3.7). This could suggest that some endogenous protein is "blocked" from binding in the ATPase mutants, however the effect compared to the levels in the knockdown is quite mild, and would require further investigation.

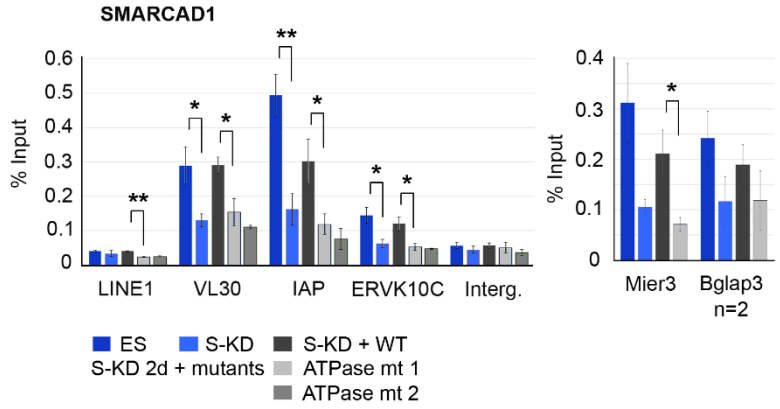

Figure 3.3.7: The catalytic mutant does not reconstitute SMARCAD1 binding over ERVs.
SMARCAD1 ChIP-qPCR of three biological replicates in five different cell lines over class I and II ERVs. *Mier3* and *Bglap3* represent specific IAPs within the annotated genes. Average enrichment is depicted (+/- S.E.) with the exception of the IAP within *Bglap3,* which was done in two biological replicates and the 2nd ATPase mutant, which was also done twice on all sites shown and accordingly has no *p*-value. Expected levels of SMARCAD1 association are seen in the ES cell line, and reduced upon SMARCAD1 knockdown. SMARCAD1 association to its binding sites is restored by the WT construct but not by the two different ATPase mutants. *P* values are from paired two-tailed Student's *t*-test: $*p < 0.05$, $**p < 0.01$.

How are the other ERV regulators affected in the SMARCAD1 mutants? KAP1 and H3K9me3 were investigated to address this question. KAP1 is no longer stably associated to class I and II ERVs in the ATPase mutant (Figure 3.3.8a, b). This is particularly pronounced on IAP elements, both with consensus and specific primers (Figure 3.3.8b)

and can also be seen in the box plot of KAP1 enrichment over IAPs from the ChIP-seq (Figure 3.3.8a). In contrast KAP1 binding is reconstituted when a WT SMARCAD1 construct is introduced in a cell line with a transient SMARCAD1 knockdown, clearly showing that the reduced KAP1 in the SMARCAD1 knockdown is a direct consequence of the loss of remodelling on the investigated sites. The ATPase mutants, on the other hand, do not reconstitute KAP1 binding and show even less KAP1 binding than seen in the SMARCAD1 knockdown. Interestingly, in the CUE1 mutant KAP1 binding levels are reliably reconstituted to initial levels, suggesting that the interaction of KAP1 and SMARCAD1 is not necessary for KAP1 association to chromatin, however the presence of a certain level of SMARCAD1 bound together with KAP1 appears to be necessary since in a near-complete SMARCAD1 knockdown, KAP1 is still reduced over their shared binding sites.

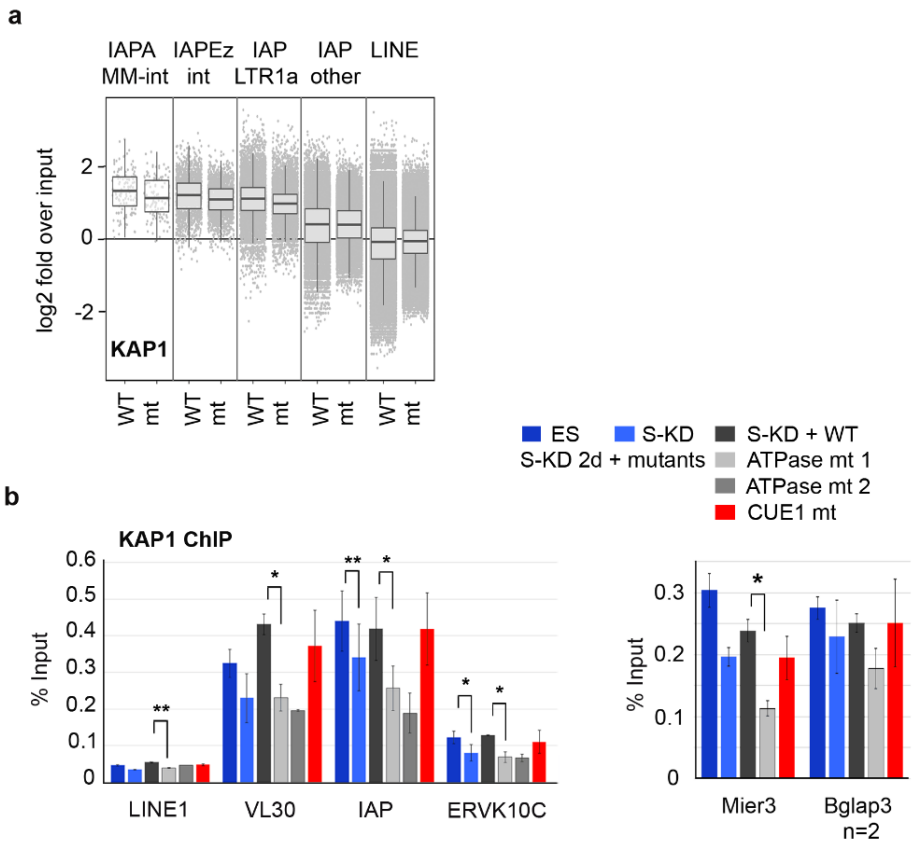

Figure 3.3.8: A catalytic SMARCAD1 mutant results in reduced KAP1 association.

(a) KAP1 binding over different IAP subcategories in a SMARCAD1 WT and a SMARCAD1 ATPase mutant cell line depicting reduced KAP1 association in the mutant. Enrichment is shown as log2 fold over input. (b) KAP1 ChIP-qPCR in three biological replicates over class I (VL30) and II (IAP & ERVK10C) ERVs with *Mier3* and *Bglap3* representing specific IAPs within the annotated genes. KAP1 binding in five different cell lines was investigated; Control, transient SMARCAD1 knockdown and three cell lines with SMARCAD1 constructs expressed over the knockdown: WT, two ATPase mutants, and a CUE1 KAP1-interaction mutant. Average Enrichment is shown (+/- S.E.). *P* values are from paired two-tailed Student's *t*-test: $*p < 0.05$, $**p < 0.01$. *Bglap3* was performed in biological duplicates as was the 2nd ATPase mutant, accordingly no *p*-values were calculated.

When it comes to H3K9me3, it is reduced after four days of transient SMARCAD1 knockdown (Figure 3.3.9). H3K9me3 association to chromatin over class I and II ERVs is restored in the cell line containing the WT SMARCAD1 construct, again illustrating the direct effect of SMARCAD1. Silencing is however not restored in the ATPase mutant

and neither is H3K9me3 fully reconstituted in the CUE1 KAP1-interaction mutant, even though KAP1 binding is back to normal levels which likely implies that the loss of SMARCAD1 remodelling alone has an impact on ERV repression apart from its impact on KAP1 association to chromatin (Figure 3.3.9).

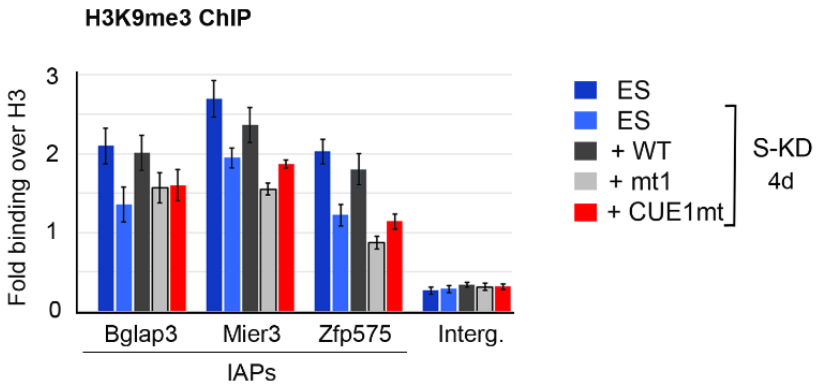

Figure 3.3.9: H3K9me3 is not reconstituted by the SMARCAD1 ATPase and CUE1 mutant. ChIP-qPCR showing H3K9me3 association to specific IAP elements in control, 4-day SMARCAD1 knockdown and reconstitution with WT, ATPase and CUE1 mutants. Enrichment (+/-S.E.) is shown as fold binding over H3 from the average of two biological replicates. An intergenic site is shown as negative control.

Taken together these data illustrate the role of SMARCAD1 in the formation of the repressive complex on retrotransposons, and specifically that it is its catalytic function as an ATP-dependent chromatin remodeller that is required for the stable assembly of KAP1 and H3K9me3 over these sites. SMARCAD1 knockdown affects KAP1, SETDB1 binding and the association of repressive histone marks and variants over ERVs. However, the effects are slightly different in cell lines with a stable SMARCAD1 knockdown compared to inducible knockdowns. Over IAPs the effect is very consistent. When it comes to the type I ERV element VL30, it is interesting to note that in the stable knockdown KAP1 is not affected but in the transient knockdown of SMARCAD1 it is reduced. In stable clones, cells have a lot of time to react to the knockdown of a protein. Possibly the lack of KAP1 reduction in the stable clone reveals a redundant mechanism that has kicked in to “rescue” the repression of VL30. Another interesting observation is

that over MusD/ETn KAP1 binding is not reduced but SETDB1 and H3K9me3 are affected albeit not significantly. This might hint towards slightly different repressive mechanisms over different elements. Notably, H3K9me3 is not affected over ERVK10C in the SMARCAD1 knockdown. Overall, SMARCAD1 and KAP1 are both important for the repression of IAPs but might play slightly different roles over other elements.

SMARCAD1 binding to chromatin is required for the successful heterochromatisation of its ERV targets and the remodelling activity of SMARCAD1 plays a role in this mechanism. Further H3K9me3 is not fully reconstituted in the KAP1 interaction mutant implying that KAP1 alone is not enough for the complete repression of these elements. This is supported by expression data from the Mermoud lab, which reveals the de-repression of IAP and ERVK10C elements upon SMARCAD1 knockdown. Additionally, the de-repression of IAP elements is rescued by the WT but not the ATPase mutant (Sachs et al., 2019)

3.4. SMARCAD1 binding sites in other cell types

SMARCAD1 is expressed in MEFs and similar to ES cells it is localized throughout the nucleus and enriched on DAPI-dense loci which represent the pericentric heterochromatin in murine cells (Figure 3.4.1a). Moreover, SMARCAD1 is known to have functions in the maintenance of heterochromatin in somatic cells (Rowbotham et al., 2011), and its binding partner KAP1 has been shown to regulate ERVs in adult human cells (Tie et al., 2018). It was assessed whether SMARCAD1 is bound to the same ERV subfamilies that are enriched in mES cells. However, the result is that SMARCAD1 is not bound to class I and II ERV target sites in MEFs (Figure 3.4.1b). At the same time, ChIP was performed with another antibody on a known target site to ensure that the quality of the chromatin and the ChIP worked in principle (Figure 3.4.1b).

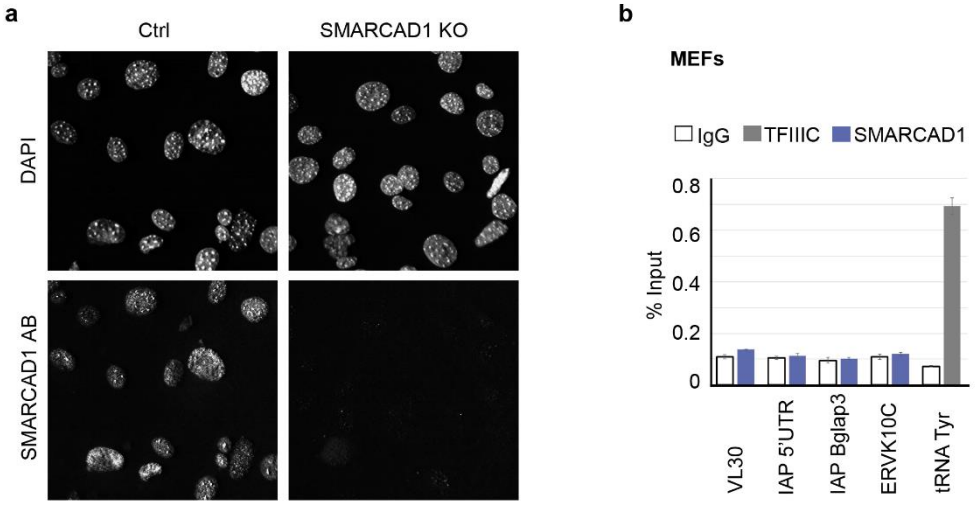

Figure 3.4.1: SMARCAD1 is not enriched on class I and II ERVs tested in MEFs.

(a) Immunofluorescence staining of endogenous SMARCAD1 in mouse embryonic fibroblasts (MEFs), Control and SMARCAD1 KNOCK-OUT. (b) ChIP-qPCR showing SMARCAD1 binding on class I and II ERVs, bound by SMARCAD1 in mES cells, in mouse embryonic fibroblasts. TFIIIC is shown as a positive control for enrichment on a tRNA gene and IgG is a negative binding control. Enrichment is shown as percent input (+/-S.E.)

To continue the investigation whether SMARCAD1 plays a role in the control of ERVs in adult tissues, SMARCAD1 binding was examined during early mES cell differentiation. A differentiation protocol for mES cells by removal of LIF (Leukaemia

Inhibitory Factor), which inhibits differentiation, was performed to test whether SMARCAD1 still binds to chromatin.

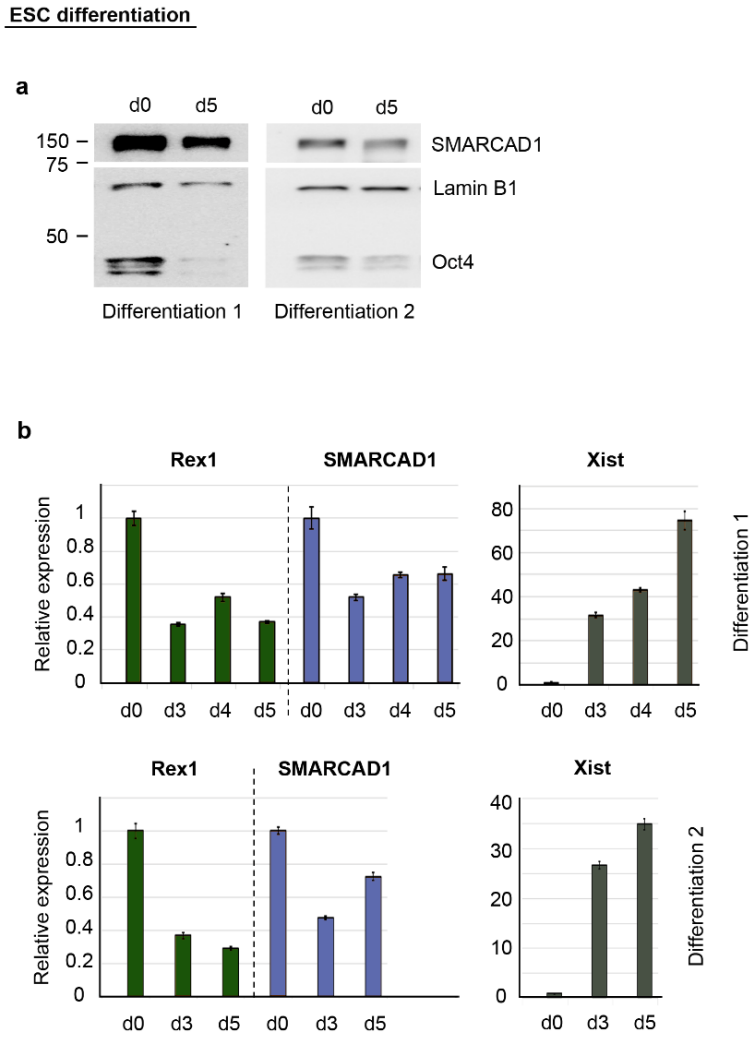

Figure 3.4.2: A five-day differentiation timecourse in female ES cells results in the expected decreased expression of pluripotency markers, SMARCAD1 and the increase of the ncRNA Xist.

Protein (a) and RNA levels (b) of indicated factors were examined during a 5-day differentiation timecourse. (a) Western blot showing the reduction of the pluripotency marker Oct4 after 5 days (d5) in two separate mES cell differentiation experiments. SMARCAD1 is also shown slightly reduced during differentiation. Lamin B1 is shown as a loading control. (b) Relative RNA expression levels by qPCR of the pluripotency marker Rex1 as well as SMARCAD1. Xist is also portrayed in its expected upregulation during female mES cell differentiation. RNA expression levels are depicted as relative expression (+/-SE).

A time course of differentiation was established and to control for successful differentiation the following markers were assessed: pluripotency markers *Oct4* and *Rex1* (Figure 3.4.2). These markers are expected to decrease upon the exit of pluripotency. Another marker for differentiation is the upregulation of the ncRNA Xist, responsible for X inactivation in female ES cells during differentiation, which was also included in the controls (Figure 3.4.2). Pluripotency markers are down both on protein and RNA level (Figure 3.4.2a, b). SMARCAD1 itself is also downregulated (Figure 3.4.2b) in accordance with the previous observation by Jacqueline Mermoud that it is generally expressed at lower levels in somatic cells (Sachs et al., 2019; Xiao et al., 2017). The ncRNA Xist is upregulated in accordance with the onset of X inactivation expected during the differentiation of female ES cells (Figure 3.4.2b). It was possible to conclude that 5 days of LIF depletion resulted in sufficiently differentiated cells.

To address the question whether SMARCAD1 binds its targets in differentiated mES cells, chromatin was collected on day 5 of differentiation. In contrast to MEFs, SMARCAD1 is still bound to class II ERV elements in differentiated mES cells. (Figure 3.4.3a). Additionally, SMARCAD1 is bound to the major and minor satellites of pericentric heterochromatin and to telomeres, albeit on satellites the enrichment is minor. Similarly, SMARCAD1 binding is maintained on the two regulatory promoters of *Xist* (P1 & P2) during mES cell differentiation (Figure 3.4.3c).

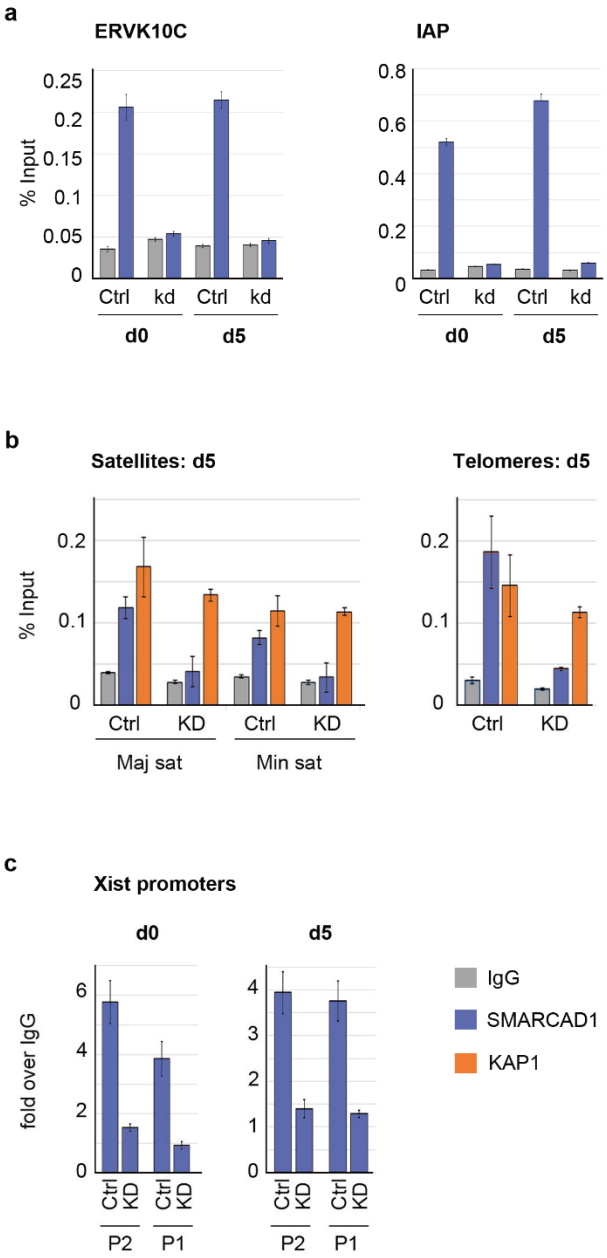

Figure 3.4.3: SMARCAD1 is enriched on its mES cell-specific binding sites after 5 days of ESC differentiation.

(a) ChIP-qPCR of SMARCAD1 bound to IAP and ERVK10C elements in PGK12.1 control and knockdown cells on d0 and d5 of differentiation. Enrichment is shown as percent input (+/-S.E., n=2). (b) ChIP-qPCR of SMARCAD1 and KAP1 bound to major and minor satellites, and telomeres after 5 days of ESC differentiation in PGK12.1 control and SMARCAD1 knockdown cells. Enrichment is shown as percent input (+/-S.E., n=2). See Figure 3.2.2 for levels of SMARCAD1 enrichment on these sites at d0. (c) ChIP-qPCR of SMARCAD1 binding to the two promoters (P1, P2) of the ncRNA *Xist* in PGK12.1 control and knockdown cells on d0 and d5 of differentiation. Enrichment is shown as fold binding over IgG (+/-S.E., n=2).

It has recently been shown that KAP1 plays a role in silencing ERVs in an adult cell line (Ecco et al., 2016; Tie et al., 2018). During the course of this present study, it was hypothesized that the role of SMARCAD1 on ERVs might not be restricted to ES cells. No SMARCAD1 binding was detected on ERVs in mouse embryonic fibroblasts (MEFs) however SMARCAD1 binding on ERVs was also not lost after five days of ES cell differentiation. This implies SMARCAD1 is still bound to its targets upon the reduction of the pluripotency markers REX1 and OCT4. Importantly, the *Smarcad1* gene is bound by NANOG, OCT4 and SOX2 (Efroni et al., 2008; Loh et al., 2006). SMARCAD1 protein levels are globally reduced after 5 days of differentiation (Figure 3.4.2a) but this reduction is not seen on chromatin at this point (Figure 3.4.3).

ERVs are bound and regulated in a tissue-specific manner and SMARCAD1 might play role in ERV regulation in other adult cell lines, such as KAP1, since the ERV family bound by KAP1 in somatic cells is a subcategory of ERVK elements not investigated in this study (Ecco et al., 2016; Tie et al., 2018). Further studies would clarify whether SMARCAD1 plays a role in the regulation of ERVs in tissues other than embryonic stem cells.

4. Discussion

The results of this thesis revealed a role of SMARCAD1 in the establishment of silent chromatin on ERVs in ES cells. SMARCAD1 and KAP1 co-occupy the investigated binding sites and SMARCAD1 is recruited to chromatin by its interaction with KAP1, but in turn also stabilizes the association of KAP1 with chromatin. These interactions lead to the recruitment of SETDB1 and the establishment of the repressive histone modifications H3K9me3 and H4K20me3 (Figure 4.1). The presence of the histone variant H3.3 was confirmed within ERV elements and is found reduced upon the loss of SMARCAD1. The role of H3.3 has previously been investigated in ERV silencing and has been debated since (Elsasser et al., 2015) – discussed in 4.2.

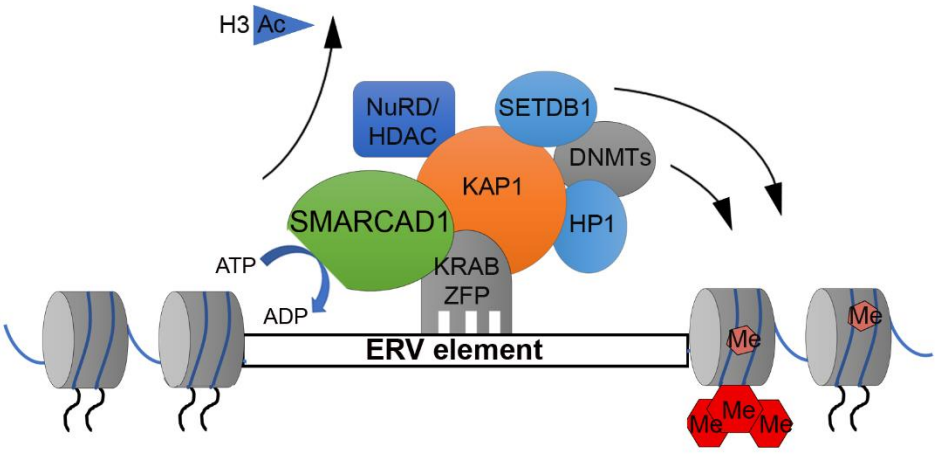

Figure 4.1: Model of the SMARCAD1-KAP1-SETDB1 silencing pathway.

The repressive factors are localized to the chromatin by KRAB-ZFPs where SMARCAD1 and KAP1 stabilize each other and are required for the association of further silencing factors and the establishment of repressive histone modifications. For this function the chromatin remodelling of SMARCAD1 is essential.

The catalytic activity of SMARCAD1 is necessary for stable KAP1 binding, whereas the interaction with SMARCAD1 is not, as seen in the CUE1 mutant. This implies that the presence of SMARCAD1 plays a role in the stabilization of KAP1 on chromatin, whereas in an almost complete knockdown of SMARCAD1 KAP1 is no longer bound at full levels to its binding sites. In the CUE1 mutant, H3K9me3 is not restored to WT levels, illustrating that KAP1 alone is not enough to fully repress transposable elements. This is reflected in the de-repression of ERVs in this instance observed by expression analysis

from the Mermoud lab and hence, SMARCAD1 and KAP1 are both necessary for the full repression of ERVs. KAP1 and H3K9me3 are equally not fully restored in the SMARCAD1 ATPase mutant (as is ERV repression in transcriptome data from the Mermoud lab), illustrating that it is indeed the catalytic activity of SMARCAD1 that plays a role in heterochromatin establishment on these sites (Figure 4.1).

4.1. Targeting remodellers to chromatin

ATP-dependent chromatin remodellers are generally recruited to chromatin by histone modifications or by co-factors. They often contain domains that bind histone modifications, such as bromo- and chromodomains, and a combination of histone modifications is generally involved in the recruitment. (Clapier and Cairns, 2009; Clapier et al., 2017b; Flaus and Owen-Hughes, 2011; Hota and Bruneau, 2016; Muller and Leutz, 2001). SWI/SNF proteins are usually targeted by accessory proteins which modulate their activity, and integrate chromatin remodelling into cellular pathways (Hota and Bruneau, 2016).

SMARCAD1 does not have a DNA binding domain nor does it have a known domain that binds histone modifications. One candidate are CUE motifs known to bind mono- and polyubiquitin (Hurley et al., 2006), hence a role of these domains in recognizing ubiquitylated histones has been debated (Byeon et al., 2013; Densham et al., 2016). However, a preference for ubiquitylated histones over unmodified histones has not been confirmed so far (Awad et al., 2010; Densham and Morris, 2017). The Mermoud lab has shown that SMARCAD1 directly interacts with recombinant non-ubiquitylated KAP1 via its CUE1 domain. Further, there is no competition in KAP1-CUE1 binding experiments upon the addition of ubiquitin (Ding et al., 2018). Hence, this is a novel interaction mode for CUE domains. This study has shown that SMARCAD1 binding to KAP1 targets requires an intact CUE1 domain. We infer that SMARCAD1 is recruited to or stabilised om chromatin by its interaction with KAP1. This is a major mechanism of targeting and relevant for the silencing of ERV elements. Open questions concern whether alternative mechanisms exist to target SMARCAD1 to chromatin.

The ChIP-seq approach has demonstrated that SMARCAD1 and KAP1 share many common target sites and it is likely that the recruitment and stabilization of SMARCAD1-KAP1 genome-wide involves the interaction with KAP1 via the CUE1 domain given that

the SMARCAD1 CUE1 domain is essential for tethering SMARCAD1 to the nucleus (Ding et al., 2018). This does not rule out other mechanisms of recruitment, perhaps on sites not shared with KAP1. Based on manual inspection of the binding sites, SMARCAD1 sites unbound by KAP1 appear to be the exception, but a thorough bioinformatic analysis is likely to give interesting candidates. One such example are the promoters of Xist. KAP1 binding has not been reported on Xist in published ChIP-seq (Elsasser et al., 2015) and hence SMARCAD1 recruitment might involve distinct mechanisms which possibly also play a role elsewhere in the genome. One different mechanism has been suggested by Xiao et al. who reported that SMARCAD1 preferentially associates with H3 arginine 26 citrullination in vitro and is colocalized with this modification genome-wide (Xiao et al., 2017).

Histone modifications other than ubiquitin could also play a role in recruiting SMARCAD1 to chromatin. In the case of H3K9me3 this present study has demonstrated that SMARCAD1 binding to chromatin on ERVs is not reduced upon H3K9me3 reduction and this modification does not play a role in stabilizing SMARCAD1 on ERVs. Interestingly, Xiao et al. have shown that H3S28 phosphorylation inhibits SMARCAD1 binding *in vitro.* It has been suggested that phosphorylation of a given site might result in the dissociation of a reader of lysine methylation on an adjacent residue as part of a "pospho/methyl switch" (Fischle et al., 2003). Indeed, phosphorylation of H3S28 upon stress resulted in displacement of PcG proteins from a subset of their targets (Gehani et al., 2010). Considering that SMARCAD1 binds many other targets outside of constitutive heterochromatin, the next step would be to investigate whether there are stress-regulated genes or elements among them, and if phosphorylation and SMARCAD1 displacement is induced upon stress on these sites.

4.2. The role of chromatin remodelling by SMARCAD1 in the repression of ERVs

This study has shown that the ATPase mutant of SMARCAD1 does not reconstitute the binding of repressive factors on ERVs implying the catalytic activity is required for the assembly of the silencing complex of KAP1-SETDB1-H3K9me3. The Mermoud lab has shown that SMARCAD1-bound ERVs are indeed upregulated upon the transient knockdown of SMARCAD1 and the repression is only restored partially with the ATPase mutant, while it is restored with the WT. So, what is SMARCAD1 doing on these sites

that requires its catalytic activity? In budding yeast, the SMARCAD1 homolog *Fun30* performs histone dimer exchange and nucleosome sliding (Awad et al., 2010; Byeon et al., 2013). In fission yeast, *Fft3* is localized to LTRs residing within the transition zone between euchromatin and silent domains. In its absence nucleosome occupancy is decreased, as shown by Mnase-seq (Steglich et al., 2015; Stralfors et al., 2011). In mES cells, H3 is unaffected upon the loss of SMARCAD1 on all sites investigated by ChIP in this study. However, ChIP would likely only detect large-scale nucleosome eviction and nucleosome sliding not at all. To answer these questions, it would be quite interesting to perform nucleosome remodelling assays *in vitro* to determine what exactly SMARCAD1 remodels. *In vivo* one would perform ATAC-seq or MNase-seq in the SMARCAD1 ATPase mutant to investigate nucleosome occupancy over ERVs to determine the mechanistic nature of ERV de-repression upon the loss of remodelling.

Are there other examples of ATP-dependent chromatin remodellers that play a role in the control of ERVs? One example is the remodeller Chd5 which plays a role in the occupancy of canonical H3.1/2 within MERVL elements and is involved in its regulation (Hayashi et al., 2016). Regarding the H3.3 data revealed in this present study, a more interesting candidate is the remodeller ATRX (alpha-thalassemia/mental retardation syndrome X-linked) which is known to deposit H3.3 within telomeres together with the chaperone DAXX (Wong et al., 2010). The data on ERVs is somewhat inconclusive so far, as it has been shown that ATRX binds a specific sequence within IAPs and promotes heterochromatin formation at retrotransposons (Sadic et al., 2015), it was further found that an ATRX knockdown upregulates the IAP1 subclass in ES cells (He et al., 2015). A third study, however, not only finds no effect of ATRX on the silencing of ERVs but also shows that there are indeed two distinct DAXX complexes, ATRX-DAXX and DAXX-SETDB1-KAP1-HDAC1, with the latter playing a role in the repression of ERVs and this is achieved with the H3.3-dependent stabilisation of DAXX. An H3.3 mutant in their study could still be integrated into chromatin but no longer interacted with DAXX and in this context ERVs were still upregulated (Hoelper et al., 2017).

In this present study, H3.3 is reduced within ERVs upon the knockdown of SMARCAD1 but this has not been tested in the presence of the ATPase mutant, hence it cannot be concluded that it is the catalytic activity of SMARCAD1 affecting it on these sites. It could be speculated that remodelling by SMARCAD1 is involved in H3.3 deposition, possibly in consonance with DAXX instead of ATRX. This would suggest ATRX does

indeed not play a role on these sites as Hoelper et al postulated. The Mermoud lab found DAXX in mass spectrometry proteomic data of the SMARCAD1 ATPase mutant, albeit at rather low amounts (data not published), and KAP1 associates with DAXX (Elsasser et al., 2015). This could be tested by performing ChIP against H3.3 and H3.1/2 simultaneously. Opposing this, is the fact that Hoelper et al found that H3.3 is not required within chromatin for the repression of ERVs. They do speculate that DAXX may serve as a bridging factor between the transcriptional repressing activities of HDACs and the H3K9me3 directed activity of the SETDB1-KAP1 proteins, so maybe it would be worth to investigate the effect SMARCAD1 has on DAXX directly. This present study has not found ATRX on IAP elements whereas in the same experiment ATRX was found highly enriched on telomeres, which would support the conclusion that ATRX does not play a major role on IAPs in the investigated cell line.

The role of H3.3 on ERVs is however also somewhat elusive. Elsässer et al. demonstrated that H3.3 is enriched at ERVs, deposited there by ATRX-DAXX, and in a H3.3 knock-out cell-line ERVs are derepressed which leads to increased retrotransposition. This conclusion has been challenged by Wolf et al., who re-analysed the published data. They confirmed the presence of H3.3 within ERVs but could not confirm the increased retrotransposition and found only mild ERV upregulation genome-wide upon the loss of H3.3 (Wolf et al., 2017). In their reply, Elsässer et al. explain the difference in the data analysis employed, and confirm that albeit the effect on ERV repression of an H3.3 knock-out is not as drastic as what is observed upon the loss of SETDB1, it is still significant and even a subtle difference in ERV expression could be biologically important. They do however withdraw their conclusion about increased retrotransposition upon the loss of H3.3, but maintain that H3.3 plays an unexpected role in ERV silencing (Elsässer et al., 2017).

This present study cannot state whether H3.3 plays a role in the regulation of ERVs. However, H3.3 is reduced upon the loss of SMARCAD1 and combined with the recent paper that showed no ATRX on ERVs it would be interesting to investigate the presence of H3.3 in the SMARCAD1 ATPase mutant to determine if SMARCAD1 plays a role in H3.3 deposition together with DAXX.

4.3. The repression ERVs by SMARCAD1 and the impact on gene regulation

Apart from examining the exact mechanism of ATP-dependent chromatin remodelling by SMARCAD1 on ERVs, there are many other questions to be explored. Colleagues in the Mermoud lab have shown that ERVs are derepressed upon the transient loss of SMARCAD1, and more importantly so are nearby genes. This phenotype is not fully rescued by the ATPase mutant nor by the CUE mutant of SMARCAD1 (Sachs et al., 2019), which is also reflected in the H3K9me3 binding data from this present study. This clearly illustrates that the silencing of retrotransposons stretches over to neighbouring genes and, upon the de-regulation of these elements, nearby genes are expressed. Hence, SMARCAD1 is necessary to maintain correct gene regulation in ES cells. Interestingly, the pluripotency of ES cells tracks very closely with the expression levels of retrotransposons (Gifford et al., 2013; Macfarlan et al., 2012). A comparison of *Oct4* and *Nanog* binding sites in humans and mice has revealed a very low conservation but one fourth of these sites in both harbour ERV elements. Moreover, the expression of the majority of genes was affected by the depletion of *Oct4* (Kunarso et al., 2010). This illustrates the plasticity of regulatory networks of ES cells as well as the role of ERVs within them. Reflecting this plasticity, ERVs additionally often function as tissue-specific promoters and alternative promoters. There are many examples of this phenomenon during development and in tissues such as the brain or the testis (Cohen et al., 2009).

Taking this into account, there are several interesting questions about the role of SMARCAD1. As mentioned in the introduction and the results of this thesis, the loss of SMARCAD1 has a clear effect on the maintenance of pluripotency. While the Mermoud lab has found a direct link to the pluripotency network in proteomic interaction data, the role of SMARCAD1 on ERVs in this context merits additional exploration. SMARCAD1 itself is bound by core pluripotency factors in the genome (Efroni et al., 2008; Loh et al., 2006), reminiscent of an autoregulatory feedback loop. In accordance with these observations, it would be interesting to investigate the de-regulated genes close to SMARCAD1-bound ERVs by RNA-seq. Do they play a role in pluripotency? Additionally, using CRISPR technology it is now feasible to inactivate or delete specific LTRs to determine their effect upon the host transcriptome. Are these genes expressed in development and might therefore drive the exit of pluripotency upon their de-repression?

Possibly this would have to be investigated together with double knockdowns of other factors of ERV silencing to capture larger effects.

From this present study, SMARCAD1 was not found bound to ERVs in MEFs but was also not lost after 5 days of differentiation. For KAP1 it is known that it plays a role in ERV silencing in several adult tissues, including the brain (Ecco et al., 2016; Fasching et al., 2015; Tie et al., 2018). Similar cell lines could be investigated for ERV silencing by SMARCAD1. In addition, the SMARCAD1 binding sites in ES cells could be investigated in several adult cell lines, with published DNAse I hypersensitivity data to correlate in what cell types they are now used as regulatory elements (together with data on active histone marks).

Many questions remain about the extent to which retrotransposons can be repressed or activated in certain tissues or at particular developmental stages and which factors coordinate switches in these activation statuses. SMARCAD1 and its co-factors are potential candidates for these mechanisms. A specific and intriguing example of ERVs in gene regulation is their potential role in innate immunity. The emergence of ERVs coincides with the appearance of adaptive immunity in mammals, suggesting that the lympho-tropism of many retroviruses may have played a role in this process (Rowe and Trono, 2011). It has been shown that some ERV families act as interferon-induced enhancers for immune genes in humans and mice (Chuong et al., 2016). Interferons are pro-inflammatory signalling molecules that are released upon infection to promote the transcription of innate immunity factors.

The interferon-induced enhancers contained within specific ERVs have dispersed in several mammalian genomes independently (Canadas et al., 2018; Chuong et al., 2016) and CRISPR-deletion of a subset in the human genome impaired the expression of adjacent interferon-induced genes, which revealed their involvement in the regulation of essential immune functions (Chuong et al., 2016). Instead of understanding the role of specific ERVs in innate immunity it would be worthwhile to investigate the role of their controlling factors. Indeed, recently it has been shown that the repression of ERVs by KAP1 in adult cells is conserved in human peripheral blood mononuclear cells and the loss of KAP1 leads to the activation of interferon-stimulated genes (Tie et al., 2018). Furthermore, SETDB1 has functions in immune evasion in certain cancer cells (Cuellar et al., 2017). Of course, ERVs in adult cell lines are generally silenced by DNA methylation, and not primarily by repressive histone modifications. Altogether, it would

be interesting to investigate the role of SMARCAD1 in this context, due to its link to DNA methylation (via NP95, discovered by the Mermoud lab) and its role in ERV silencing together with the KAP1-SETDB1 machinery. This has the potential to be employed to reactivate subsets of ERVs to harness their natural ability to trigger the innate immune response.

4.4. Genome-wide binding sites of SMARCAD1 in mES cells

SMARCAD1 binds many targets genome-wide other than ERVs and its exact role on these sites remains to be determined. Some interesting possibilities are discussed here.

4.4.1. SMARCAD1 on pericentric heterochromatin

This present study has demonstrated that SMARCAD1 binds satellites and telomeres via ChIP and its presence there is likely also dependent upon its interaction with KAP1. How KAP1 is recruited to these sites, however, remains a mystery. KAP1 is not affected upon the stable loss of SMARCAD1 on satellites and only mildly so, upon the transient loss of SMARCAD1 in ES cells. To further elucidate the role of SMARCAD1 one would have to look at the effect on histone modifications and whether its absence impacts DNA replication in ES cells. If the stabilization of SMARCAD1 is similarly dependent upon its interaction with KAP1 can be investigated in binding experiments with the CUE1-mutant.

4.4.2. SMARCAD1 on imprinting Control Regions (IRCs)

This study has identified several SMARCAD1 peaks on the promoters of imprinted genes. Interestingly, it has previously been shown that a KRAB-ZFP (ZFP57) together with KAP1, binds the H3K9me3-bearing and DNA-methylated allele of Imprinting Control Regions (IRCs) (Quenneville et al., 2011; Strogantsev et al., 2015). KAP1 and ZFP57 associate with NP95 (or UHRF1), a protein known to be required for the maintenance of DNA methylation in ES cells (Avvakumov et al., 2008; Bostick et al., 2007; Quenneville et al., 2011). The deletion of KAP1 results in the loss of heterochromatin marks on IRCs and the deletion of ZFP57 leads to ICR DNA demethylation (Quenneville et al., 2011).

The Mermoud lab has shown that SMARCAD1 interacts with NP95 in both mES and human cells and it has been confirmed in this laboratory with purified recombinant proteins that this is a direct interaction (unpublished data). Speculatively, SMARCAD1 may link histone modifications established with the help of KAP1, to DNA methylation brought by NP95 and DNMT1. This would also be worth investigating on pericentric heterochromatin and ERVs, since all these elements require histone modifications during development, to protect them from the wave of DNA demethylation, and later in development acquire DNA methylation.

4.4.3. SMARCAD1 on the 3'end of ZFP genes

SMARCAD1 binds to the 3'end of a subset of KRAB-ZFP genes. The 3'end of ZFP genes is one of the first sites were a specific ZFP (ZFP274) was found responsible for the recruitment of KAP1, and they are also regions which co-localize with SETDB1 and H3K9me3 (Frietze et al., 2010). ZFPs recognize specific DNA sequences, so they are the factors that confer binding specificity to their co-factors. In order to attain a complete picture, the ChIP-seq overlap of hundreds of ZFPs would need to be investigated together with SMARCAD1 and other co-factors. To increase the complexity, the expression of ZFPs is highly tissue-specific, as they contribute to cell-specific gene regulation and make up a large network (Ecco et al., 2017; Emerson and Thomas, 2009). To speculate on the role of SMARCAD1 on the 3'end of ZFP genes: it is probable that remodelling plays a role in the establishment or maintenance of heterochromatin on these sites, perhaps to stop recombination of the zinc finger repeats.

Abstract

Chromatin remodellers slide, assemble, eject or edit nucleosomes influencing chromatin structure, DNA accessibility and transcriptional programmes. The SNF2-like remodeller SMARCAD1 is conserved from yeast to human cells and is highly expressed in mouse embryonic stem cells. Upon its loss cells lose their pluripotent phenotype but its function in ES cells is not known.

In order to understand the role of SMARCAD1 in mouse embryonic stem cells a robust ChIP-seq protocol was developed for the tagged and endogenous protein in wild-type and knockdown cell lines. SMARCAD1 binding sites were found predominantly at intergenic sites genome-wide, and overlap with repressive histone modifications. Among the chromatin-bound proteins discovered enriched with SMARCAD1 binding sites is KAP1 (KRAB-associated protein 1; Krüppel-associated box), a critical factor for the silencing of endogenous retroviral elements (ERVs) in mouse ES cell, the histone methyltransferase SETDB1 (SET Domain Bifurcated Histone Lysine Methyltransferase I), and the histone variant H3.3. Taken together, the discovered binding sites provide new understanding of SMARCAD1 function in ES cells and illustrate that in the open chromatin environment characteristic of the pluripotent state, SMARCAD1 is associated with transcriptional repression.

An unresolved issue is how SMARCAD1 associates with its binding sites without a DNA-binding domain and no known domains for recruitment by histone modifications. Candidate SMARCAD1 targets were investigated and it was discovered that recruitment is dependent on the interaction with KAP1 via the CUE1 (Coupling of Ubiquitin conjugation to ER degradation) domain of SMARCAD1. Sequential ChIP experiments revealed that KAP1 and SMARCAD1 are co-enriched on their shared targets.

Among the discovered binding sites of the remodeller SMARCAD1 are endogenous retroviral elements (ERVs) an abundant type of transposable element derived from viral integrations in the germline. ERV expression is tightly controlled by repressive factors as they pose a threat for genome stability. A series of knockdowns and ChIP-qPCR experiments was performed to understand the cooperation and underlying mechanisms of how these factors shape and control ERV heterochromatin. SMARCAD1 was identified as a crucial component; it is required for the association of the KAP1-SETDB1 silencing machinery over class I and II ERVs, and consequently the maintenance of the histone modifications H3K9me3 and H4K20me3. The histone variant H3.3 has a controversial

role in ERV control and is similarly reduced upon the loss of SMARCAD1, suggesting SMARCAD1 may be involved in the turn-over of this variant.

In summary, heterochromatin organisation is perturbed when SMARCAD1 is not present at ERVs. The presence of H3K9me3 on the other hand is not necessary for SMARCAD1 binding as shown in a SETDB1 knock-down. The assembly of KAP1 and H3K9me3 is rescued by the ectopic expression of wild-type SMARCAD1 but not by an ATPase mutant. Hence, it is the catalytic activity of SMARCAD1 and chromatin remodelling that is required for the silencing of ERVs. The KAP1 interaction mutant had no effect on the association of KAP1 itself but could similarly not restore H3K9me3 emphasizing that SMARCAD1 is required for successful heterochromatin formation on ERVs and identifying chromatin remodelling as a key mechanism of ERV control in mouse ES cells.

Zusammenfassung

Chromatin-Remodeller können Nukleosomen verschieben, zusammenfügen, ausstoßen oder diese restrukturieren, wobei sie die Chromatinstruktur, die Zugänglichkeit der DNA und die Transkriptionsprogramme der Zelle beeinflussen. Der SNF2-like Remodeller SMARCAD1 ist von Hefe bis zu menschlichen Zellen konserviert und wird in embryonalen Stammzellen (ES-Zellen) von Mäusen stark exprimiert. Bei seinem Verlust verlieren Zellen ihren pluripotenten Phänotyp, allerdings ist die Funktion von SMARCAD1 in Stammzellen nicht bekannt.

Um die Rolle von SMARCAD1 in embryonalen Stammzellen zu verstehen, wurde ein robustes ChIP-seq-Protokoll für das getaggte und endogene Protein in Wildtyp- und Knockdown-Zelllinien entwickelt. SMARCAD1-Bindungsstellen liegen überwiegend in intergenen Stellen genomweit und überlappen mit repressiven Histonmodifikationen. Unter den mit SMARCAD1-Bindungsstellen angereicherten Chromatin-gebundenen Proteinen befindet sich KAP1 (KRAB-associated protein 1; Krüppel-associated box), ein kritischer Faktor für die Stummschaltung endogener retroviraler Elemente (ERVs) in Maus-ES-Zellen, die Histonmethyltransferase SETDB1 (SET Domain Bifurcated Histone Lysine Methyltransferase I) und die Histonvariante H3.3. Zusammengenommen liefern die entdeckten Bindungsstellen ein neues Verständnis der SMARCAD1-Funktion in ES-Zellen und veranschaulichen, dass SMARCAD1 in der für den pluripotenten Zustand charakteristischen offenen Chromatin-Umgebung mit Transkriptionsrepression assoziiert ist.

Ein ungelöstes Problem ist, wie SMARCAD1 ohne DNA-Bindungsdomäne und ohne bekannte Domänen für die Rekrutierung durch Histonmodifikationen mit seinen Bindungsstellen assoziiert. Verschiedene Kategorien von SMARCAD1 Bindungsstellen wurden untersucht und es wurde entdeckt, dass die Rekrutierung von der Interaktion mit KAP1 über die CUE1-Domäne (Coupling of Ubiquitin conjugation to ER degradation) von SMARCAD1 abhängt. Sequenzielle ChIP-Experimente zeigten, dass KAP1 und SMARCAD1 an ihren gemeinsamen Bindungsstellen zugleich angereichert sind.

Unter den entdeckten Bindungsstellen des Remodellers SMARCAD1 befinden sich endogene retrovirale Elemente (ERVs), eine häufig vorkommende Art transponierbarer Elemente, die aus viralen Integrationen in der Keimbahn stammen. Die ERV-Expression wird durch repressive Faktoren streng kontrolliert, da sie sonst eine Bedrohung für die Genomstabilität darstellen. Eine Reihe von Knockdown und ChIP-

qPCR-Experimenten wurde durchgeführt, um zu verstehen, wie diese Faktoren ERV-Heterochromatin formen. SMARCAD1 wurde als entscheidende Komponente identifiziert; es ist nötig für die Assoziation der repressiven Faktoren KAP1 und SETDB1 zu ERVs der Klassen I und II und folglich für die Erhaltung von den Histonmodifikationen H3K9me3 und H4K20me3. Die Histonvariante H3.3 spielt eine kontroverse Rolle bei der ERV-Kontrolle und wird beim Verlust von SMARCAD1 reduziert, was darauf hindeutet, dass SMARCAD1 an dem Turnover dieser Variante beteiligt sein könnte.

Zusammenfassend lässt sich sagen, dass die Heterochromatin-Organisation ist gestört, wenn SMARCAD1 nicht auf ERVs vorhanden ist. Das Vorhandensein von H3K9me3 ist andererseits für die SMARCAD1-Bindung nicht erforderlich, wie in einem SETDB1-Knockdown gezeigt. Die Assemblierung von KAP1 und H3K9me3 wird durch die ektopische Expression von Wildtyp-SMARCAD1, jedoch nicht durch eine ATPase-Mutante gerettet. Daher sind es die katalytische Aktivität von SMARCAD1 und das Chromatin-Remodelling, die für die Stummschaltung von ERVs erforderlich sind. Die KAP1-Interaktionsmutante hatte keinen Einfluss auf die Assoziation von KAP1 selbst, konnte jedoch H3K9me3 ebenfalls nicht wiederherstellen, was zeigt, dass SMARCAD1 für eine erfolgreiche Heterochromatinbildung auf ERVs erforderlich ist. Damit ist die Chromatin-Remodellierung als Schlüsselmechanismus der ERV-Kontrolle in Maus-ES-Zellen identifiziert.

References

— Ahmad, K., and Henikoff, S. (2002). The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. Mol Cell *9*, 1191-1200.

— Allis, C.D., Caparros, M.-L., Jenuwein, T., and Reinberg, D. (2015). Chapter 3: Overview and concepts. In Epigenetics, C.D. Allis, M.-L. Caparros, T. Jenuwein, and D. Reinberg, eds. (Cold Spring Harbor, New York: CSH Press, Cold Spring Harbor Laboratory Press), pp. 47-115.

— Allis, C.D., and Jenuwein, T. (2016). The molecular hallmarks of epigenetic control. Nature reviews. Genetics *17*, 487-500.

— Amendola, M., and van Steensel, B. (2014). Mechanisms and dynamics of nuclear lamina-genome interactions. Current opinion in cell biology *28*, 61-68.

— Aranda, P.S., LaJoie, D.M., and Jorcyk, C.L. (2012). Bleach gel: a simple agarose gel for analyzing RNA quality. Electrophoresis *33*, 366-369.

— Arents, G., Burlingame, R.W., Wang, B.C., Love, W.E., and Moudrianakis, E.N. (1991). The nucleosomal core histone octamer at 3.1 A resolution: a tripartite protein assembly and a left-handed superhelix. Proc Natl Acad Sci U S A *88*, 10148-10152.

— Avner, P., and Heard, E. (2001). X-chromosome inactivation: counting, choice and initiation. Nature reviews. Genetics *2*, 59-67.

— Avvakumov, G.V., Walker, J.R., Xue, S., Li, Y., Duan, S., Bronner, C., Arrowsmith, C.H., and Dhe-Paganon, S. (2008). Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. Nature *455*, 822-825.

— Awad, S., Ryan, D., Prochasson, P., Owen-Hughes, T., and Hassan, A.H. (2010). The Snf2 homolog Fun30 acts as a homodimeric ATP-dependent chromatin-remodeling enzyme. The Journal of biological chemistry *285*, 9477-9484.

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

— Barklis, E., Mulligan, R.C., and Jaenisch, R. (1986). Chromosomal position or virus mutation permits retrovirus expression in embryonal carcinoma cells. Cell *47*, 391-399.

— Barlow, D.P., and Bartolomei, M.S. (2014). Genomic imprinting in mammals. Cold Spring Harbor perspectives in biology *6*.

— 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.

— Becker, P.B., and Horz, W. (2002). ATP-dependent nucleosome remodeling. Annual review of biochemistry *71*, 247-273.

— Benit, L., Lallemand, J.B., Casella, J.F., Philippe, H., and Heidmann, T. (1999). ERV-L elements: a family of endogenous retrovirus-like elements active throughout the evolution of mammals. Journal of virology *73*, 3301-3308.

— Bickmore, W.A. (2013). The spatial organization of the human genome. Annual review of genomics and human genetics *14*, 67-84.

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

— Bird, A.P. (1986). CpG-rich islands and the function of DNA methylation. Nature *321*, 209-213.
— Black, J.C., Van Rechem, C., and Whetstine, J.R. (2012). Histone lysine methylation dynamics: establishment, regulation, and biological impact. Mol Cell *48*, 491-507.
— Blahnik, K.R., Dou, L., Echipare, L., Iyengar, S., O'Geen, H., Sanchez, E., Zhao, Y., Marra, M.A., Hirst, M., Costello, J.F., *et al.* (2011). Characterization of the contradictory chromatin signatures at the 3' exons of zinc finger genes. PloS one *6*, e17121.
— Boeke, J.D., Garfinkel, D.J., Styles, C.A., and Fink, G.R. (1985). Ty elements transpose through an RNA intermediate. Cell *40*, 491-500.
— Bostick, M., Kim, J.K., Esteve, P.O., Clark, A., Pradhan, S., and Jacobsen, S.E. (2007). UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science *317*, 1760-1764.
— Bourque, G., Burns, K.H., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M., Imbeault, M., Izsvak, Z., Levin, H.L., Macfarlan, T.S., *et al.* (2018). Ten things you should know about transposable elements. Genome Biol *19*, 199.
— Bourque, G., Leong, B., Vega, V.B., Chen, X., Lee, Y.L., Srinivasan, K.G., Chew, J.L., Ruan, Y., Wei, C.L., Ng, H.H., *et al.* (2008). Evolution of the mammalian transcription factor binding repertoire via transposable elements. Genome Res *18*, 1752-1762.
— Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., *et al.* (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell *122*, 947-956.
— Brehm, A., Tufteland, K.R., Aasland, R., and Becker, P.B. (2004). The many colours of chromodomains. BioEssays : news and reviews in molecular, cellular and developmental biology *26*, 133-140.
— Brockdorff, N., Ashworth, A., Kay, G.F., McCabe, V.M., Norris, D.P., Cooper, P.J., Swift, S., and Rastan, S. (1992). The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. Cell *71*, 515-526.
— Brown, C.J., Ballabio, A., Rupert, J.L., Lafreniere, R.G., Grompe, M., Tonlorenzi, R., and Willard, H.F. (1991). A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. Nature *349*, 38-44.
— Brown, P.O., Bowerman, B., Varmus, H.E., and Bishop, J.M. (1987). Correct integration of retroviral DNA in vitro. Cell *49*, 347-356.
— Brulet, P., Kaghad, M., Xu, Y.S., Croissant, O., and Jacob, F. (1983). Early differential tissue expression of transposon-like repetitive DNA sequences of the mouse. Proc Natl Acad Sci U S A *80*, 5641-5645.
— Bulut-Karslioglu, A., De La Rosa-Velazquez, I.A., Ramirez, F., Barenboim, M., Onishi-Seebacher, M., Arand, J., Galan, C., Winter, G.E., Engist, B., Gerle, B., *et al.* (2014). Suv39h-dependent H3K9me3 marks intact retrotransposons and silences LINE elements in mouse embryonic stem cells. Mol Cell *55*, 277-290.
— Buschbeck, M., and Hake, S.B. (2017). Variants of core histones and their roles in cell fate decisions, development and cancer. Nature reviews. Molecular cell biology *18*, 299-314.
— Byeon, B., Wang, W., Barski, A., Ranallo, R.T., Bao, K., Schones, D.E., Zhao, K., Wu, C., and Wu, W.H. (2013). The ATP-dependent chromatin remodeling

enzyme Fun30 represses transcription by sliding promoter-proximal nucleosomes. The Journal of biological chemistry *288*, 23182-23193.

— Cam, H.P., Sugiyama, T., Chen, E.S., Chen, X., FitzGerald, P.C., and Grewal, S.I. (2005). Comprehensive analysis of heterochromatin- and RNAi-mediated epigenetic control of the fission yeast genome. Nature genetics *37*, 809-819.

— Cammas, F., Herzog, M., Lerouge, T., Chambon, P., and Losson, R. (2004). Association of the transcriptional corepressor TIF1beta with heterochromatin protein 1 (HP1): an essential role for progression through differentiation. Genes Dev *18*, 2147-2160.

— Cammas, F., Mark, M., Dolle, P., Dierich, A., Chambon, P., and Losson, R. (2000). Mice lacking the transcriptional corepressor TIF1beta are defective in early postimplantation development. Development *127*, 2955-2963.

— Cammas, F., Oulad-Abdelghani, M., Vonesch, J.L., Huss-Garcia, Y., Chambon, P., and Losson, R. (2002). Cell differentiation induces TIF1beta association with centromeric heterochromatin via an HP1 interaction. Journal of cell science *115*, 3439-3448.

— Canadas, I., Thummalapalli, R., Kim, J.W., Kitajima, S., Jenkins, R.W., Christensen, C.L., Campisi, M., Kuang, Y., Zhang, Y., Gjini, E., *et al.* (2018). Tumor innate immunity primed by specific interferon-stimulated endogenous retroviruses. Nature medicine *24*, 1143-1150.

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

— Carabana, J., Watanabe, A., Hao, B., and Krangel, M.S. (2011). A barrier-type insulator forms a boundary between active and inactive chromatin at the murine TCRbeta locus. Journal of immunology *186*, 3556-3562.

— Castro-Diaz, N., Ecco, G., Coluccio, A., Kapopoulou, A., Yazdanpanah, B., Friedli, M., Duc, J., Jang, S.M., Turelli, P., and Trono, D. (2014). Evolutionally dynamic L1 regulation in embryonic stem cells. Genes Dev *28*, 1397-1409.

— Cedar, H., and Bergman, Y. (2009). Linking DNA methylation and histone modification: patterns and paradigms. Nature reviews. Genetics *10*, 295-304.

— Chen, X., Cui, D., Papusha, A., Zhang, X., Chu, C.D., Tang, J., Chen, K., Pan, X., and Ira, G. (2012). The Fun30 nucleosome remodeller promotes resection of DNA double-strand break ends. Nature *489*, 576-580.

— Cheng, B., Ren, X., and Kerppola, T.K. (2014a). KAP1 represses differentiation-inducible genes in embryonic stem cells through cooperative binding with PRC1 and derepresses pluripotency-associated genes. Molecular and cellular biology *34*, 2075-2091.

— Cheng, C.T., Kuo, C.Y., and Ann, D.K. (2014b). KAPtain in charge of multiple missions: Emerging roles of KAP1. World J Biol Chem *5*, 308-320.

— Chow, J.C., Ciaudo, C., Fazzari, M.J., Mise, N., Servant, N., Glass, J.L., Attreed, M., Avner, P., Wutz, A., Barillot, E., *et al.* (2010). LINE-1 activity in facultative heterochromatin formation during X chromosome inactivation. Cell *141*, 956-969.

— Chuong, E.B., Elde, N.C., and Feschotte, C. (2016). Regulatory evolution of innate immunity through co-option of endogenous retroviruses. Science *351*, 1083-1087.

— Clapier, C.R., and Cairns, B.R. (2009). The biology of chromatin remodeling complexes. Annual review of biochemistry *78*, 273-304.

— Clapier, C.R., Iwasa, J., Cairns, B.R., and Peterson, C.L. (2017a). Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. Nature reviews. Molecular cell biology *18*, 407-422.
— Clapier, C.R., Iwasa, J., Cairns, B.R., and Peterson, C.L. (2017b). Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. Nat Rev Mol Cell Bio *18*, 407-422.
— Cohen, C.J., Lock, W.M., and Mager, D.L. (2009). Endogenous retroviral LTRs as promoters for human genes: a critical assessment. Gene *448*, 105-114.
— Collins, R.E., Tachibana, M., Tamaru, H., Smith, K.M., Jia, D., Zhang, X., Selker, E.U., Shinkai, Y., and Cheng, X. (2005). In vitro and in vivo analyses of a Phe/Tyr switch controlling product specificity of histone lysine methyltransferases. The Journal of biological chemistry *280*, 5563-5570.
— Costas, J. (2003). Molecular characterization of the recent intragenomic spread of the murine endogenous retrovirus MuERV-L. Journal of molecular evolution *56*, 181-186.
— Costelloe, T., Louge, R., Tomimatsu, N., Mukherjee, B., Martini, E., Khadaroo, B., Dubois, K., Wiegant, W.W., Thierry, A., Burma, S., *et al.* (2012). The yeast Fun30 and human SMARCAD1 chromatin remodellers promote DNA end resection. Nature *489*, 581-584.
— Cuellar, T.L., Herzner, A.M., Zhang, X., Goyal, Y., Watanabe, C., Friedman, B.A., Janakiraman, V., Durinck, S., Stinson, J., Arnott, D., *et al.* (2017). Silencing of retrotransposons by SETDB1 inhibits the interferon response in acute myeloid leukemia. The Journal of cell biology *216*, 3535-3549.
— Davey, C.A., Sargent, D.F., Luger, K., Maeder, A.W., and Richmond, T.J. (2002). Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution. Journal of molecular biology *319*, 1097-1113.
— Densham, R.M., Garvin, A.J., Stone, H.R., Strachan, J., Baldock, R.A., Daza-Martin, M., Fletcher, A., Blair-Reid, S., Beesley, J., Johal, B., *et al.* (2016). Human BRCA1-BARD1 ubiquitin ligase activity counteracts chromatin barriers to DNA resection. Nature structural & molecular biology *23*, 647-655.
— Densham, R.M., and Morris, J.R. (2017). The BRCA1 Ubiquitin ligase function sets a new trend for remodelling in DNA repair. Nucleus *8*, 116-125.
— Deuring, R., Fanti, L., Armstrong, J.A., Sarte, M., Papoulas, O., Prestel, M., Daubresse, G., Verardo, M., Moseley, S.L., Berloco, M., *et al.* (2000). The ISWI chromatin-remodeling protein is required for gene expression and the maintenance of higher order chromatin structure in vivo. Mol Cell *5*, 355-365.
— Ding, D., Bergmaier, P., Sachs, P., Klangwart, M., Ruckert, T., Bartels, N., Demmers, J., Dekker, M., Poot, R.A., and Mermoud, J.E. (2018). The CUE1 domain of the SNF2-like chromatin remodeler SMARCAD1 mediates its association with KRAB-associated protein 1 (KAP1) and KAP1 target genes. The Journal of biological chemistry *293*, 2711-2724.
— Dodge, J.E., Kang, Y.K., Beppu, H., Lei, H., and Li, E. (2004). Histone H3-K9 methyltransferase ESET is essential for early development. Molecular and cellular biology *24*, 2478-2486.
— Du, J., Johnson, L.M., Jacobsen, S.E., and Patel, D.J. (2015). DNA methylation pathways and their crosstalk with histone methylation. Nature reviews. Molecular cell biology *16*, 519-532.
— Durand-Dubief, M., Will, W.R., Petrini, E., Theodorou, D., Harris, R.R., Crawford, M.R., Paszkiewicz, K., Krueger, F., Correra, R.M., Vetter, A.T., *et al.*

(2012). SWI/SNF-like chromatin remodeling factor Fun30 supports point centromere function in S. cerevisiae. PLoS genetics *8*, e1002974.

— Eapen, V.V., Sugawara, N., Tsabar, M., Wu, W.H., and Haber, J.E. (2012). The Saccharomyces cerevisiae chromatin remodeler Fun30 regulates DNA end resection and checkpoint deactivation. Molecular and cellular biology *32*, 4727-4740.

— Ecco, G., Cassano, M., Kauzlaric, A., Duc, J., Coluccio, A., Offner, S., Imbeault, M., Rowe, H.M., Turelli, P., and Trono, D. (2016). Transposable Elements and Their KRAB-ZFP Controllers Regulate Gene Expression in Adult Tissues. Developmental cell *36*, 611-623.

— Ecco, G., Imbeault, M., and Trono, D. (2017). KRAB zinc finger proteins. Development *144*, 2719-2729.

— Efroni, S., Duttagupta, R., Cheng, J., Dehghani, H., Hoeppner, D.J., Dash, C., Bazett-Jones, D.P., Le Grice, S., McKay, R.D., Buetow, K.H., *et al.* (2008). Global transcription in pluripotent embryonic stem cells. Cell stem cell *2*, 437-447.

— Elsasser, S.J., Ernst, R.J., Walker, O.S., and Chin, J.W. (2016). Genetic code expansion in stable cell lines enables encoded chromatin modification. Nature methods *13*, 158-164.

— Elsässer, S.J., Noh, K.-M., Diaz, N., Allis, C.D., and Banaszynski, L.A. (2017). Elsässer et al. reply. Nature *548*, E7-E9.

— Elsasser, S.J., Noh, K.M., Diaz, N., Allis, C.D., and Banaszynski, L.A. (2015). Histone H3.3 is required for endogenous retroviral element silencing in embryonic stem cells. Nature *522*, 240-244.

— Emerson, R.O., and Thomas, J.H. (2009). Adaptive evolution in zinc finger transcription factors. PLoS genetics *5*, e1000325.

— Epsztejn-Litman, S., Feldman, N., Abu-Remaileh, M., Shufaro, Y., Gerson, A., Ueda, J., Deplus, R., Fuks, F., Shinkai, Y., Cedar, H., *et al.* (2008). De novo DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes. Nature structural & molecular biology *15*, 1176-1183.

— Fadloun, A., Le Gras, S., Jost, B., Ziegler-Birling, C., Takahashi, H., Gorab, E., Carninci, P., and Torres-Padilla, M.E. (2013). Chromatin signatures and retrotransposon profiling in mouse embryos reveal regulation of LINE-1 by RNA. Nature structural & molecular biology *20*, 332-338.

— Fang, J., Feng, Q., Ketel, C.S., Wang, H., Cao, R., Xia, L., Erdjument-Bromage, H., Tempst, P., Simon, J.A., and Zhang, Y. (2002). Purification and functional characterization of SET8, a nucleosomal histone H4-lysine 20-specific methyltransferase. Curr Biol *12*, 1086-1099.

— Fasching, L., Kapopoulou, A., Sachdeva, R., Petri, R., Jonsson, M.E., Manne, C., Turelli, P., Jern, P., Cammas, F., Trono, D., *et al.* (2015). TRIM28 represses transcription of endogenous retroviruses in neural progenitor cells. Cell reports *10*, 20-28.

— Feldman, N., Gerson, A., Fang, J., Li, E., Zhang, Y., Shinkai, Y., Cedar, H., and Bergman, Y. (2006). G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. Nature cell biology *8*, 188-194.

— Feng, Q., and Zhang, Y. (2001). The MeCP1 complex represses transcription through preferential binding, remodeling, and deacetylating methylated nucleosomes. Genes Dev *15*, 827-832.

— Feschotte, C., and Gilbert, C. (2012). Endogenous viruses: insights into viral evolution and impact on host biology. Nature reviews. Genetics *13*, 283-296.

— Ficz, G., Branco, M.R., Seisenberger, S., Santos, F., Krueger, F., Hore, T.A., Marques, C.J., Andrews, S., and Reik, W. (2011). Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. Nature *473*, 398-402.
— Fischle, W., Wang, Y., and Allis, C.D. (2003). Binary switches and modification cassettes in histone biology and beyond. Nature *425*, 475-479.
— Flaus, A., Martin, D.M., Barton, G.J., and Owen-Hughes, T. (2006). Identification of multiple distinct Snf2 subfamilies with conserved structural motifs. Nucleic Acids Res *34*, 2887-2905.
— Flaus, A., and Owen-Hughes, T. (2011). Mechanisms for ATP-dependent chromatin remodelling: the means to the end. The FEBS journal *278*, 3579-3595.
— Fort, A., Hashimoto, K., Yamada, D., Salimullah, M., Keya, C.A., Saxena, A., Bonetti, A., Voineagu, I., Bertin, N., Kratz, A., *et al.* (2014). Deep transcriptome profiling of mammalian stem cells supports a regulatory role for retrotransposons in pluripotency maintenance. Nature genetics *46*, 558-566.
— Frietze, S., O'Geen, H., Blahnik, K.R., Jin, V.X., and Farnham, P.J. (2010). ZNF274 recruits the histone methyltransferase SETDB1 to the 3' ends of ZNF genes. PloS one *5*, e15082.
— 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.
— Gehani, S.S., Agrawal-Singh, S., Dietrich, N., Christophersen, N.S., Helin, K., and Hansen, K. (2010). Polycomb group protein displacement and gene activation through MSK-dependent H3K27me3S28 phosphorylation. Mol Cell *39*, 886-900.
— Gelbart, M.E., Bachman, N., Delrow, J., Boeke, J.D., and Tsukiyama, T. (2005). Genome-wide identification of Isw2 chromatin-remodeling targets by localization of a catalytically inactive mutant. Genes Dev *19*, 942-954.
— Gemmell, P., Hein, J., and Katzourakis, A. (2015). Orthologous endogenous retroviruses exhibit directional selection since the chimp-human split. Retrovirology *12*, 52.
— Gendrel, A.V., and Heard, E. (2014). Noncoding RNAs and epigenetic mechanisms during X-chromosome inactivation. Annual review of cell and developmental biology *30*, 561-580.
— Gerdes, P., Richardson, S.R., Mager, D.L., and Faulkner, G.J. (2016). Transposable elements in the mammalian embryo: pioneers surviving through stealth and service. Genome Biol *17*.
— Gifford, W.D., Pfaff, S.L., and Macfarlan, T.S. (2013). Transposable elements as genetic regulatory substrates in early development. Trends Cell Biol *23*, 218-226.
— Goke, J., and Ng, H.H. (2016). CTRL+INSERT: retrotransposons and their contribution to regulation and innovation of the transcriptome. Embo Rep *17*, 1131-1144.
— Goldberg, A.D., Allis, C.D., and Bernstein, E. (2007). Epigenetics: a landscape takes shape. Cell *128*, 635-638.
— 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.
— Grabundzija, I., Messing, S.A., Thomas, J., Cosby, R.L., Bilic, I., Miskey, C., Gogol-Doring, A., Kapitonov, V., Diem, T., Dalda, A., *et al.* (2016). A Helitron

transposon reconstructed from bats reveals a novel mechanism of genome shuffling in eukaryotes. Nature communications *7*, 10716.
— Greenberg, M.V.C., and Bourc'his, D. (2019). The diverse roles of DNA methylation in mammalian development and disease. Nature reviews. Molecular cell biology *20*, 590-607.
— Grewal, S.I., and Moazed, D. (2003). Heterochromatin and epigenetic control of gene expression. Science *301*, 798-802.
— Groh, S., and Schotta, G. (2017). Silencing of endogenous retroviruses by heterochromatin. Cellular and molecular life sciences : CMLS *74*, 2055-2065.
— Hackett, J.A., and Surani, M.A. (2014). Regulatory principles of pluripotency: from the ground state up. Cell stem cell *15*, 416-430.
— Hainer, S.J., Boskovic, A., McCannell, K.N., Rando, O.J., and Fazzio, T.G. (2019). Profiling of Pluripotency Factors in Single Cells and Early Embryos. Cell *177*, 1319-1329 e1311.
— Hall, I.M., Shankaranarayana, G.D., Noma, K., Ayoub, N., Cohen, A., and Grewal, S.I. (2002). Establishment and maintenance of a heterochromatin domain. Science *297*, 2232-2237.
— Harada, F., Peters, G.G., and Dahlberg, J.E. (1979). The primer tRNA for Moloney murine leukemia virus DNA synthesis. Nucleotide sequence and aminoacylation of tRNAPro. The Journal of biological chemistry *254*, 10979-10985.
— Harr, J.C., Gonzalez-Sandoval, A., and Gasser, S.M. (2016). Histones and histone modifications in perinuclear chromatin anchoring: from yeast to man. Embo Rep *17*, 139-155.
— Hayashi, M., Maehara, K., Harada, A., Semba, Y., Kudo, K., Takahashi, H., Oki, S., Meno, C., Ichiyanagi, K., Akashi, K., *et al.* (2016). Chd5 Regulates MuERV-L/MERVL Expression in Mouse Embryonic Stem Cells Via H3K27me3 Modification and Histone H3.1/H3.2. Journal of cellular biochemistry *117*, 780-792.
— He, Q.Y., Kim, H., Huang, R., Lu, W.S., Tang, M.F., Shi, F.T., Yang, D., Zhang, X.Y., Huang, J.J., Liu, D., *et al.* (2015). The Daxx/Atrx Complex Protects Tandem Repetitive Elements during DNA Hypomethylation by Promoting H3K9 Trimethylation. Cell stem cell *17*, 273-286.
— Hirano, T. (2012). Condensins: universal organizers of chromosomes with diverse functions. Genes Dev *26*, 1659-1678.
— Hoelper, D., Huang, H., Jain, A.Y., Patel, D.J., and Lewis, P.W. (2017). Structural and mechanistic insights into ATRX-dependent and -independent functions of the histone chaperone DAXX. Nature communications *8*, 1193.
— Hota, S.K., and Bruneau, B.G. (2016). ATP-dependent chromatin remodeling during mammalian development. Development *143*, 2882-2897.
— Hu, G., Kim, J., Xu, Q., Leng, Y., Orkin, S.H., and Elledge, S.J. (2009). A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal. Genes Dev *23*, 837-848.
— Huntley, S., Baggott, D.M., Hamilton, A.T., Tran-Gyamfi, M., Yang, S., Kim, J., Gordon, L., Branscomb, E., and Stubbs, L. (2006). A comprehensive catalog of human KRAB-associated zinc finger genes: insights into the evolutionary history of a large family of transcriptional repressors. Genome Res *16*, 669-677.
— Hurley, J.H., Lee, S., and Prag, G. (2006). Ubiquitin-binding domains. Biochem J *399*, 361-372.

— Hutnick, L.K., Huang, X., Loo, T.C., Ma, Z., and Fan, G. (2010). Repression of retrotransposal elements in mouse embryonic stem cells is primarily mediated by a DNA methylation-independent mechanism. The Journal of biological chemistry *285*, 21082-21091.
— Hyun, K., Jeon, J., Park, K., and Kim, J. (2017). Writing, erasing and reading histone lysine methylations. Experimental & molecular medicine *49*, e324.
— Iyengar, S., and Farnham, P.J. (2011). KAP1 protein: an enigmatic master regulator of the genome. The Journal of biological chemistry *286*, 26267-26276.
— Jacobs, F.M., Greenberg, D., Nguyen, N., Haeussler, M., Ewing, A.D., Katzman, S., Paten, B., Salama, S.R., and Haussler, D. (2014). An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. Nature *516*, 242-245.
— Jang, C.W., Shibata, Y., Starmer, J., Yee, D., and Magnuson, T. (2015). Histone H3.3 maintains genome integrity during mammalian development. Genes Dev *29*, 1377-1392.
— Jang, S.M., Kauzlaric, A., Quivy, J.P., Pontis, J., Rauwel, B., Coluccio, A., Offner, S., Duc, J., Turelli, P., Almouzni, G., *et al.* (2018). KAP1 facilitates reinstatement of heterochromatin after DNA replication. Nucleic Acids Res.
— Jern, P., Sperber, G.O., and Blomberg, J. (2005). Use of Endogenous Retroviral Sequences (ERVs) and structural markers for retroviral phylogenetic inference and taxonomy. Retrovirology *2*.
— Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nature reviews. Genetics *13*, 484-492.
— Jonkers, I., Barakat, T.S., Achame, E.M., Monkhorst, K., Kenter, A., Rentmeester, E., Grosveld, F., Grootegoed, J.A., and Gribnau, J. (2009). RNF12 is an X-Encoded dose-dependent activator of X chromosome inactivation. Cell *139*, 999-1011.
— Jurka, J., and Smith, T. (1988). A fundamental division in the Alu family of repeated sequences. Proc Natl Acad Sci U S A *85*, 4775-4778.
— Karimi, M.M., Goyal, P., Maksakova, I.A., Bilenky, M., Leung, D., Tang, J.X., Shinkai, Y., Mager, D.L., Jones, S., Hirst, M., *et al.* (2011). DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mESCs. Cell stem cell *8*, 676-687.
— Karnowski, A., Cao, C., Matthias, G., Carotta, S., Corcoran, L.M., Martensson, I.L., Skok, J.A., and Matthias, P. (2008). Silencing and nuclear repositioning of the lambda5 gene locus at the pre-B cell stage requires Aiolos and OBF-1. PloS one *3*, e3568.
— Kassiotis, G. (2014). Endogenous retroviruses and the development of cancer. Journal of immunology *192*, 1343-1349.
— Kikyo, N., Wade, P.A., Guschin, D., Ge, H., and Wolffe, A.P. (2000). Active remodeling of somatic nuclei in egg cytoplasm by the nucleosomal ATPase ISWI. Science *289*, 2360-2362.
— Klemm, S.L., Shipony, Z., and Greenleaf, W.J. (2019). Chromatin accessibility and the regulatory epigenome. Nature reviews. Genetics *20*, 207-220.
— Knoepfler, P.S., and Eisenman, R.N. (1999). Sin meets NuRD and other tails of repression. Cell *99*, 447-450.
— Kornberg, R.D. (1974). Chromatin structure: a repeating unit of histones and DNA. Science *184*, 868-871.
— Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693-705.

— Kuff, E.L., and Lueders, K.K. (1988). The intracisternal A-particle gene family: structure and functional aspects. Advances in cancer research *51*, 183-276.
— Kunarso, G., Chia, N.Y., Jeyakani, J., Hwang, C., Lu, X., Chan, Y.S., Ng, H.H., and Bourque, G. (2010). Transposable elements have rewired the core regulatory network of human embryonic stem cells. Nature genetics *42*, 631-634.
— 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.
— Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nature methods *9*, 357-359.
— Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25.
— Lechner, M.S., Schultz, D.C., Negorev, D., Maul, G.G., and Rauscher, F.J., 3rd (2005). The mammalian heterochromatin protein 1 binds diverse nuclear proteins through a common motif that targets the chromoshadow domain. Biochemical and biophysical research communications *331*, 929-937.
— Lee, J., Choi, E.S., Seo, H.D., Kang, K., Gilmore, J.M., Florens, L., Washburn, M.P., Choe, J., Workman, J.L., and Lee, D. (2017). Chromatin remodeller Fun30Fft3 induces nucleosome disassembly to facilitate RNA polymerase II elongation. Nature communications *8*, 14527.
— Lee, J.T., and Lu, N. (1999). Targeted mutagenesis of Tsix leads to nonrandom X inactivation. Cell *99*, 47-57.
— Lehnertz, B., Ueda, Y., Derijck, A.A.H.A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T.P., 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.
— Leung, D.C., Dong, K.B., Maksakova, I.A., Goyal, P., Appanah, R., Lee, S., Tachibana, M., Shinkai, Y., Lehnertz, B., Mager, D.L., *et al.* (2011). Lysine methyltransferase G9a is required for de novo DNA methylation and the establishment, but not the maintenance, of proviral silencing. Proc Natl Acad Sci U S A *108*, 5718-5723.
— Li, X., Ito, M., Zhou, F., Youngson, N., Zuo, X., Leder, P., and Ferguson-Smith, A.C. (2008). A maternal-zygotic effect gene, Zfp57, maintains both maternal and paternal imprints. Developmental cell *15*, 547-557.
— Liu, S., Brind'Amour, J., Karimi, M.M., Shirane, K., Bogutz, A., Lefebyre, L., Sasaki, H., Shinkai, Y., and Lorinczi, M.C. (2014). Setdb1 is required for germline development and silencing of H3K9me3-marked endogenous retroviruses in primordial germ cells. Gene Dev *28*, 2041-2055.
— Loh, Y.H., Wu, Q., Chew, J.L., Vega, V.B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J., *et al.* (2006). The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nature genetics *38*, 431-440.
— Loyola, A., Tagami, H., Bonaldi, T., Roche, D., Quivy, J.P., Imhof, A., Nakatani, Y., Dent, S.Y., and Almouzni, G. (2009). The HP1alpha-CAF1-SetDB1-containing complex provides H3K9me1 for Suv39-mediated K9me3 in pericentric heterochromatin. Embo Rep *10*, 769-775.
— Luger, K., Mader, 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.

— Macfarlan, T.S., Gifford, W.D., Agarwal, S., Driscoll, S., Lettieri, K., Wang, J., Andrews, S.E., Franco, L., Rosenfeld, M.G., Ren, B., *et al.* (2011a). Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A. Genes Dev *25*, 594-607.
— Macfarlan, T.S., Gifford, W.D., Agarwal, S., Driscoll, S., Lettieri, K., Wang, J.X., Andrews, S.E., Franco, L., Rosenfeld, M.G., Ren, B., *et al.* (2011b). Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A. Gene Dev *25*, 594-607.
— Macfarlan, T.S., Gifford, W.D., Driscoll, S., Lettieri, K., Rowe, H.M., Bonanomi, D., Firth, A., Singer, O., Trono, D., and Pfaff, S.L. (2012). Embryonic stem cell potency fluctuates with endogenous retrovirus activity. Nature *487*, 57-63.
— Mager, D.L., and Freeman, J.D. (2000). Novel mouse type D endogenous proviruses and ETn elements share long terminal repeat and internal sequences. Journal of virology *74*, 7221-7229.
— Maksakova, I.A., Romanish, M.T., Gagnier, L., Dunn, C.A., de Lagemaat, L.N.V., and Mager, D.L. (2006). Retroviral elements and their hosts: Insertional mutagenesis in the mouse germ line. PLoS genetics *2*, 1-10.
— 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*, 15.
— Malik, H.S., and Eickbush, T.H. (2001). Phylogenetic analysis of ribonuclease H domains suggests a late, chimeric origin of LTR retrotransposable elements and retroviruses. Genome Res *11*, 1187-1197.
— Marino-Ramirez, L., and Jordan, I.K. (2006). Transposable element derived DNaseI-hypersensitive sites in the human genome. Biology direct *1*, 20.
— Markopoulos, G., Noutsopoulos, D., Mantziou, S., Gerogiannis, D., Thrasyvoulou, S., Vartholomatos, G., Kolettas, E., and Tzavaras, T. (2016). Genomic analysis of mouse VL30 retrotransposons. Mobile DNA *7*, 10.
— Martens, J.H., O'Sullivan, R.J., Braunschweig, U., Opravil, S., Radolf, M., Steinlein, P., and Jenuwein, T. (2005a). The profile of repeat-associated histone lysine methylation states in the mouse epigenome. Embo J *24*, 800-812.
— Martens, J.H.A., O'Sullivan, R.J., Braunschweig, U., Opravil, S., Radolf, M., Steinlein, P., and Jenuwein, T. (2005b). The profile of repeat-associated histone lysine methylation states in the mouse epigenome. Embo J *24*, 800-812.
— 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.
— Maze, I., Noh, K.M., Soshnev, A.A., and Allis, C.D. (2014). Every amino acid matters: essential contributions of histone variants to mammalian development and disease. Nature reviews. Genetics *15*, 259-271.
— Meshorer, E., and Misteli, T. (2006). Chromatin in pluripotent embryonic stem cells and differentiation. Nature reviews. Molecular cell biology *7*, 540-546.
— Micallef, L., and Rodgers, P. (2014). eulerAPE: Drawing Area-Proportional 3-Venn Diagrams Using Ellipses. PloS one *9*.
— 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.

— Mizuguchi, G., Shen, X., Landry, J., Wu, W.H., Sen, S., and Wu, C. (2004). ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. Science *303*, 343-348.
— Morey, C., and Avner, P. (2011). The demoiselle of X-inactivation: 50 years old and as trendy and mesmerising as ever. PLoS genetics *7*, e1002212.
— Mouse Genome Sequencing, C., Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., *et al.* (2002). Initial sequencing and comparative analysis of the mouse genome. Nature *420*, 520-562.
— Muchardt, C., and Yaniv, M. (2001). When the SWI/SNF complex remodels...the cell cycle. Oncogene *20*, 3067-3075.
— Muller, C., and Leutz, A. (2001). Chromatin remodeling in development and differentiation. Current opinion in genetics & development *11*, 167-174.
— Muller, 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.
— Nasmyth, K., and Haering, C.H. (2009). Cohesin: its roles and mechanisms. Annual review of genetics *43*, 525-558.
— Neves-Costa, A., Will, W.R., Vetter, A.T., Miller, J.R., and Varga-Weisz, P. (2009). The SNF2-family member Fun30 promotes gene silencing in heterochromatic loci. PloS one *4*, e8111.
— Nishioka, K., Rice, J.C., Sarma, K., Erdjument-Bromage, H., Werner, J., Wang, Y., Chuikov, S., Valenzuela, P., Tempst, P., Steward, R., *et al.* (2002). PR-Set7 is a nucleosome-specific methyltransferase that modifies lysine 20 of histone H4 and is associated with silent chromatin. Mol Cell *9*, 1201-1213.
— Nowick, K., Hamilton, A.T., Zhang, H., and Stubbs, L. (2010). Rapid sequence and expression divergence suggest selection for novel function in primate-specific KRAB-ZNF genes. Molecular biology and evolution *27*, 2606-2617.
— Papamichos-Chronakis, M., Krebs, J.E., and Peterson, C.L. (2006). Interplay between Ino80 and Swr1 chromatin remodeling enzymes regulates cell cycle checkpoint adaptation in response to DNA damage. Genes Dev *20*, 2437-2449.
— Peters, A.H., O'Carroll, D., Scherthan, H., Mechtler, K., Sauer, S., Schofer, C., Weipoltshammer, K., Pagani, M., Lachner, M., Kohlmaier, A., *et al.* (2001). Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. Cell *107*, 323-337.
— Petersen, R., Kempler, G., and Barklis, E. (1991). A stem cell-specific silencer in the primer-binding site of a retrovirus. Molecular and cellular biology *11*, 1214-1221.
— Pirrotta, V. (1998). Polycombing the genome: PcG, trxG, and chromatin silencing. Cell *93*, 333-336.
— Qin, C., Wang, Z., Shang, J., Bekkari, K., Liu, R., Pacchione, S., McNulty, K.A., Ng, A., Barnum, J.E., and Storer, R.D. (2010). Intracisternal A particle genes: Distribution in the mouse genome, active subtypes, and potential roles as species-specific mediators of susceptibility to cancer. Mol Carcinog *49*, 54-67.
— Quenneville, S., Verde, G., Corsinotti, A., Kapopoulou, A., Jakobsson, J., Offner, S., Baglivo, I., Pedone, P.V., Grimaldi, G., Riccio, A., *et al.* (2011). In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. Mol Cell *44*, 361-372.

— Quina, A.S., Buschbeck, M., and Di Croce, L. (2006). Chromatin structure and epigenetics. Biochemical pharmacology *72*, 1563-1569.
— 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.
— Rebollo, R., Karimi, M.M., Bilenky, M., Gagnier, L., Miceli-Royer, K., Zhang, Y., Goyal, P., Keane, T.M., Jones, S., Hirst, M., *et al.* (2011). Retrotransposon-induced heterochromatin spreading in the mouse revealed by insertional polymorphisms. PLoS genetics *7*, e1002301.
— Reik, W., and Walter, J. (2001). Genomic imprinting: parental influence on the genome. Nature reviews. Genetics *2*, 21-32.
— Ren, J., Briones, V., Barbour, S., Yu, W., Han, Y., Terashima, M., and Muegge, K. (2015). The ATP binding site of the chromatin remodeling homolog Lsh is required for nucleosome density and de novo DNA methylation at repeat sequences. Nucleic Acids Res *43*, 1444-1455.
— Reuss, F.U. (1992). Expression of intracisternal A-particle-related retroviral element-encoded envelope proteins detected in cell lines. Journal of virology *66*, 1915-1923.
— Ribet, D., Harper, F., Dupressoir, A., Dewannieux, M., Pierron, G., and Heidmann, T. (2008). An infectious progenitor for the murine IAP retrotransposon: emergence of an intracellular genetic parasite from an ancient retrovirus. Genome Res *18*, 597-609.
— 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.
— Richmond, E., and Peterson, C.L. (1996). Functional analysis of the DNA-stimulated ATPase domain of yeast SWI2/SNF2. Nucleic Acids Res *24*, 3685-3692.
— Robbez-Masson, L., and Rowe, H.M. (2015). Retrotransposons shape species-specific embryonic stem cell gene expression. Retrovirology *12*, 45.
— Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nature biotechnology *29*, 24-26.
— Robinson, P.J., and Rhodes, D. (2006). Structure of the '30 nm' chromatin fibre: a key role for the linker histone. Current opinion in structural biology *16*, 336-343.
— Rothbart, S.B., Krajewski, K., Nady, N., Tempel, W., Xue, S., Badeaux, A.I., Barsyte-Lovejoy, D., Martinez, J.Y., Bedford, M.T., Fuchs, S.M., *et al.* (2012). Association of UHRF1 with methylated H3K9 directs the maintenance of DNA methylation. Nature structural & molecular biology *19*, 1155-1160.
— Rowbotham, S.P., Barki, L., Neves-Costa, A., Santos, F., Dean, W., Hawkes, N., Choudhary, P., Will, W.R., Webster, J., Oxley, D., *et al.* (2011). Maintenance of silent chromatin through replication requires SWI/SNF-like chromatin remodeler SMARCAD1. Mol Cell *42*, 285-296.
— 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.
— Rowe, H.M., Kapopoulou, A., Corsinotti, A., Fasching, L., Macfarlan, T.S., Tarabay, Y., Viville, S., Jakobsson, J., Pfaff, S.L., and Trono, D. (2013). TRIM28

repression of retrotransposon-based enhancers is necessary to preserve transcriptional dynamics in embryonic stem cells. Genome Res *23*, 452-461.
— Rowe, H.M., and Trono, D. (2011). Dynamic control of endogenous retroviruses during development. Virology *411*, 273-287.
— Sachs, M., Onodera, C., Blaschke, K., Ebata, K.T., Song, J.S., and Ramalho-Santos, M. (2013). Bivalent chromatin marks developmental regulatory genes in the mouse embryonic germline in vivo. Cell reports *3*, 1777-1784.
— Sachs, P., Ding, D., Bergmaier, P., Lamp, B., Schlagheck, C., Finkernagel, F., Nist, A., Stiewe, T., and Mermoud, J.E. (2019). SMARCAD1 ATPase activity is required to silence endogenous retroviruses in embryonic stem cells. Nature communications *10*, 1335.
— Sadic, D., Schmidt, K., Groh, S., Kondofersky, I., Ellwart, J., Fuchs, C., Theis, F.J., and Schotta, G. (2015). Atrx promotes heterochromatin formation at retrotransposons. Embo Rep *16*, 836-850.
— 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.
— Schlesinger, S., and Goff, S.P. (2015). Retroviral transcriptional regulation and embryonic stem cells: war and peace. Molecular and cellular biology *35*, 770-777.
— Schoor, M., Schustergossler, K., and Gossler, A. (1993). The Etl-1 Gene Encodes a Nuclear-Protein Differentially Expressed during Early Mouse Development. Developmental Dynamics *197*, 227-237.
— Schoorlemmer, J., Perez-Palacios, R., Climent, M., Guallar, D., and Muniesa, P. (2014). Regulation of Mouse Retroelement MuERV-L/MERVL Expression by REX1 and Epigenetic Control of Stem Cell Potency. Frontiers in oncology *4*, 14.
— 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.
— Schotta, G., Sengupta, R., Kubicek, S., Malin, S., Kauer, M., Callen, E., Celeste, A., Pagani, M., Opravil, S., De La Rosa-Velazquez, I.A., *et al.* (2008). A chromatin-wide transition to H4K20 monomethylation impairs genome integrity and programmed DNA rearrangements in the mouse. Genes Dev *22*, 2048-2061.
— Schultz, D.C., Ayyanathan, K., Negorev, D., Maul, G.G., and Rauscher, F.J., 3rd (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.
— Sharif, J., Endo, T.A., Nakayama, M., Karimi, M.M., Shimada, M., Katsuyama, K., Goyal, P., Brind'Amour, J., Sun, M.A., Sun, Z.X., *et al.* (2016). Activation of Endogenous Retroviruses in Dnmt1(-/-) ESCs Involves Disruption of SETDB1-Mediated Repression by NP95 Binding to Hemimethylated DNA. Cell stem cell *19*, 81-94.
— 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.
— Smit, A.F. (1999). Interspersed repeats and other mementos of transposable elements in mammalian genomes. Current opinion in genetics & development *9*, 657-663.
— Smith, Z.D., and Meissner, A. (2013). DNA methylation: roles in mammalian development. Nature reviews. Genetics *14*, 204-220.

— Song, J., Rechkoblit, O., Bestor, T.H., and Patel, D.J. (2011). Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. Science *331*, 1036-1040.
— Sorek, R., Ast, G., and Graur, D. (2002). Alu-containing exons are alternatively spliced. Genome Res *12*, 1060-1067.
— Stavenhagen, J.B., and Robins, D.M. (1988). An ancient provirus has imposed androgen regulation on the adjacent mouse sex-limited protein gene. Cell *55*, 247-254.
— Steglich, B., Stralfors, A., Khorosjutina, O., Persson, J., Smialowska, A., Javerzat, J.P., and Ekwall, K. (2015). The Fun30 Chromatin Remodeler Fft3 Controls Nuclear Organization and Chromatin Structure of Insulators and Subtelomeres in Fission Yeast. PLoS genetics *11*.
— Stocking, C., and Kozak, C.A. (2008). Murine endogenous retroviruses. Cellular and molecular life sciences : CMLS *65*, 3383-3398.
— Stralfors, A., Walfridsson, J., Bhuiyan, H., and Ekwall, K. (2011). The FUN30 chromatin remodeler, Fft3, protects centromeric and subtelomeric domains from euchromatin formation. PLoS genetics *7*, e1001334.
— Strogantsev, R., Krueger, F., Yamazawa, K., Shi, H., Gould, P., Goldman-Roberts, M., McEwen, K., Sun, B., Pedersen, R., and Ferguson-Smith, A.C. (2015). Allele-specific binding of ZFP57 in the epigenetic regulation of imprinted and non-imprinted monoallelic expression. Genome Biol *16*, 112.
— 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.
— Tessarz, P., and Kouzarides, T. (2014). Histone core modifications regulating nucleosome structure and dynamics. Nature reviews. Molecular cell biology *15*, 703-708.
— Thompson, P.J., Macfarlan, T.S., and Lorincz, M.C. (2016). Long Terminal Repeats: From Parasitic Elements to Building Blocks of the Transcriptional Regulatory Repertoire. Mol Cell *62*, 766-776.
— Tian, B., Yang, J., and Brasier, A.R. (2012). Two-step cross-linking for analysis of protein-chromatin interactions. Methods in molecular biology *809*, 105-120.
— Tian, D., Sun, S., and Lee, J.T. (2010). The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation. Cell *143*, 390-403.
— Tie, C.H., Fernandes, L., Conde, L., Robbez-Masson, L., Sumner, R.P., Peacock, T., Rodriguez-Plata, M.T., Mickute, G., Gifford, R., Towers, G.J., *et al.* (2018). KAP1 regulates endogenous retroviruses in adult human cells and contributes to innate immune control. Embo Rep *19*.
— Trojer, P., and Reinberg, D. (2007). Facultative heterochromatin: is there a distinctive molecular signature? Mol Cell *28*, 1-13.
— Tsukada, Y., Fang, J., Erdjument-Bromage, H., Warren, M.E., Borchers, C.H., Tempst, P., and Zhang, Y. (2006). Histone demethylation by a family of JmjC domain-containing proteins. Nature *439*, 811-816.
— Urrutia, R. (2003). KRAB-containing zinc-finger repressor proteins. Genome Biol *4*, 231.
— Volkel, S., Stielow, B., Finkernagel, F., Stiewe, T., Nist, A., and Suske, G. (2015). Zinc finger independent genome-wide binding of Sp2 potentiates recruitment of histone-fold protein Nf-y distinguishing it from Sp1 and Sp3. PLoS genetics *11*, e1005102.

— Voon, H.P.J., Hughes, J.R., Rode, C., De La Rosa-Velazquez, I.A., Jenuwein, T., Feil, R., Higgs, D.R., and Gibbons, R.J. (2015). ATRX Plays a Key Role in Maintaining Silencing at Interstitial Heterochromatic Loci and Imprinted Genes. Cell reports *11*, 405-418.
— Wade, P.A., Gegonne, A., Jones, P.L., Ballestar, E., Aubry, F., and Wolffe, A.P. (1999). Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation. Nature genetics *23*, 62-66.
— Wang, G.G., and Allis, C.D. (2009). "Misinterpretation" of a histone mark is linked to aberrant stem cells and cancer development. Cell cycle *8*, 1982-1983.
— Wang, P.J. (2017). Tracking LINE1 retrotransposition in the germline. Proc Natl Acad Sci U S A *114*, 7194-7196.
— Winter, A.G., Sourvinos, G., Allison, S.J., Tosh, K., Scott, P.H., Spandidos, D.A., and White, R.J. (2000). RNA polymerase III transcription factor TFIIIC2 is overexpressed in ovarian tumors. Proc Natl Acad Sci U S A *97*, 12619-12624.
— Wolf, D., and Goff, S.P. (2009). Embryonic stem cells use ZFP809 to silence retroviral DNAs. Nature *458*, 1201-1204.
— Wolf, G., Rebollo, R., Karimi, M.M., Ewing, A.D., Kamada, R., Wu, W., Wu, B., Bachu, M., Ozato, K., Faulkner, G.J., *et al.* (2017). On the role of H3.3 in retroviral silencing. Nature *548*, E1-E3.
— Wolf, G., Yang, P., Fuchtbauer, A.C., Fuchtbauer, E.M., Silva, A.M., Park, C., Wu, W., Nielsen, A.L., Pedersen, F.S., and Macfarlan, T.S. (2015). The KRAB zinc finger protein ZFP809 is required to initiate epigenetic silencing of endogenous retroviruses. Genes Dev *29*, 538-554.
— Wolstenholme, J.T., Rissman, E.F., and Bekiranov, S. (2013). Sexual differentiation in the developing mouse brain: contributions of sex chromosome genes. Genes, brain, and behavior *12*, 166-180.
— Wong, L.H., McGhie, J.D., Sim, M., Anderson, M.A., Ahn, S., Hannan, R.D., George, A.J., Morgan, K.A., Mann, J.R., and Choo, K.H. (2010). ATRX interacts with H3.3 in maintaining telomere structural integrity in pluripotent embryonic stem cells. Genome Res *20*, 351-360.
— Wossidlo, M., Nakamura, T., Lepikhov, K., Marques, C.J., Zakhartchenko, V., Boiani, M., Arand, J., Nakano, T., Reik, W., and Walter, J. (2011). 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. Nature communications *2*, 241.
— Xiao, S., Lu, J., Sridhar, B., Cao, X., Yu, P., Zhao, T., Chen, C.C., McDee, D., Sloofman, L., Wang, Y., *et al.* (2017). SMARCAD1 Contributes to the Regulation of Naive Pluripotency by Interacting with Histone Citrullination. Cell reports *18*, 3117-3128.
— Yamada, T., Fischle, W., Sugiyama, T., Allis, C.D., and Grewal, S.I. (2005). The nucleation and maintenance of heterochromatin by a histone deacetylase in fission yeast. Mol Cell *20*, 173-185.
— Yang, H., Pesavento, J.J., Starnes, T.W., Cryderman, D.E., Wallrath, L.L., Kelleher, N.L., and Mizzen, C.A. (2008). Preferential dimethylation of histone H4 lysine 20 by Suv4-20. The Journal of biological chemistry *283*, 12085-12092.
— Zang, C., Schones, D.E., Zeng, C., Cui, K., Zhao, K., and Peng, W. (2009). A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. Bioinformatics *25*, 1952-1958.
— Zemojtel, T., Penzkofer, T., Schultz, J., Dandekar, T., Badge, R., and Vingron, M. (2007). Exonization of active mouse L1s: a driver of transcriptome evolution? BMC genomics *8*, 392.

— Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., *et al.* (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137.
— Zhang, Y., Ng, H.H., Erdjument-Bromage, H., Tempst, P., Bird, A., and Reinberg, D. (1999). Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. Genes Dev *13*, 1924-1935.
— Zhou, K., Gaullier, G., and Luger, K. (2019). Nucleosome structure and dynamics are coming of age. Nature structural & molecular biology *26*, 3-13.

Appendices

List of academic teachers

At the University of Leicester:

Badge, Dr Richard
Cann, Dr Alan
Cashmore, Professor Annette
Dubrova, Professor Yuri
Eperon, Professor Ian
Gornall, Dr Richard
Grubb, Professor Blair
Hart, Professor Paul
Heaphy, Dr Shaun
Kyriacou, Professor Charalambous
Matheson, Dr Tom
Makarova, Dr Olga
McDearmid, Dr Jonathan
Meacock, Dr Peter
Pringle, Dr J Howard
Rawlings. Dr Bernard
Royle, Dr Nicola
Scott, Professor Jon
Schwarzacher. Dr Trude
Smith, Dr Carl
Tata, Dr Fred
Wells, Dr Christine

At Newcastle University:

Bateson, Professor Melissa
Brennan, Dr Paul
Burn, Professor Sir John
Bushby, Professor Katherine
Cunningham, Professor Mark
Elliott, Professor David
Goodship, Professor Judith
Goodship, Professor Tim
Jackson, Dr Michael
Matthews, Professor John
McAllister Williams, Professor Hamish
Rowe, Professor Candy
Santibanez Koref, Dr Mauro
Smulders, Dr Tom
Strachan, Professor Tom
Wilson, Dr Ian
Young, Professor David

Acknowledgments

I would like to express my outmost gratitude to my supervisor Dr Jacqueline Mermoud for her constant support during my PhD, for her enthusiasm, criticism and knowledge. I have learned so much under her guidance, and the knowledge and skills I gained are useful to me every day. I cannot imagine having achieved this much in my PhD without her supervision.

I would like to thank Dr Boris Lamp, Dr Andrea Nist, and the Core Facility Genomics at the Institute of Molecular Tumour Biology (IMT) in Marburg for the sequencing and bioinformatics analyses, and Dr Bastian Stielow and Prof. Dr Guntram Suske for their help in establishing ChIP and ChIP-seq. Their knowledge and pointers were invaluable.

I also would like to thank all members of the Mermoud lab for creating an amazing work environment and for their contributions to my PhD. I want to particularly thank Katrin Treutwein, Bianca Bamberger, Dr Dong Ding, and Dr Philipp Bergmaier for their help and support. My thank also goes to all members of the IMT and TRR81 for the creative research atmosphere and interesting discussions.

I want to thank Dr Colin Dingwall for his astute advice on presentations, experiments and publications.

Special thanks go to Dr Igor Mačinković for the proofreading, the innumerable amount of cooked dinners, and his annoying, albeit in hindsight incredibly appreciated, constant reminders to keep writing. Thanks for being there.

I would also like to thank my sister Lotta-Lilli Fiedel for the invigorating, beautiful, and tasty work-retreats. Let’s do it again without theses to write.

I would like to thank my friends Anna Hofmann, Sena Kosan, and Kai Gelhard for supporting me emotionally during the PhD and for trying to organise a celebration almost two years too early. Soon.

Finally, I want to thank my closest family; Beate Sachs, Martin Stiefel, and Robert Stemmer for always being proud of me and for everything really. I love you.