Aus dem Adolf-Butenandt-Institut der
Ludwig-Maximilians-Universität München
Lehrstuhl: Molekularbiologie
Direktor: Prof. Dr. Peter B. Becker
Arbeitsgruppe: Prof. Dr. Gunnar Schotta

Regulation of heterochromatic gene silencing in mouse

Dissertation zum Erwerb des Doktorgrades der
Naturwissenschaften (Dr. rer. nat.) an der Medizinischen Fakultät
der Ludwig-Maximilians-Universität München

vorgelegt von

Dennis Šadić

aus Daun

München, 2014

Gedruckt mit Genehmigung der Medizinischen Fakultät der Ludwig-Maximilians-Universität München

Betreuer: Prof. Dr. Gunnar Schotta

Zweitgutachter: Prof. Dr. Heiko Hermeking

Dekan: Prof. Dr. Dr. h.c. Maximilian Reiser, FACR, FRCR

Tag der mündlichen Prüfung: 19.09.2014

Eidesstattliche Versicherung

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Dissertation mit dem Thema

“Regulation of heterochromatic gene silencing in mouse”

selbständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

Ort, Datum

Unterschrift Dennis Šadić

The work of my thesis is under preparation for a publication in a peer-reviewed journal.
During my PhD thesis I collaborated with others in the following scientific projects:

Dambacher, S., Deng, W., Hahn, M., **Sadic, D**., Frohlich, J., Nuber, A., Hoischen, C., Diekmann, S., Leonhardt, H., and Schotta, G. (2012). CENP-C facilitates the recruitment of M18BP1 to centromeric chromatin. Nucleus 3, 101-110.

Hahn, M., Dambacher, S., Dulev, S., Kuznetsova, A.Y., Eck, S., Worz, S., **Sadic, D**., Schulte, M., Mallm, J.P., Maiser, A., et al. (2013). Suv4-20h2 mediates chromatin compaction and is important for cohesin recruitment to heterochromatin. Genes & development 27, 859-872.

Terrados, G., Finkernagel, F., Stielow, B., **Sadic, D**., Neubert, J., Herdt, O., Krause, M., Scharfe, M., Jarek, M., and Suske, G. (2012). Genome-wide localization and expression profiling establish Sp2 as a sequence-specific transcription factor regulating vitally important genes. Nucleic acids research 40, 7844-7857.

I. Table of contents

II. Abstract

Heterochromatin covers a large fraction of the mammalian genome and defects in heterochromatin formation in mice result in severe developmental defects, genome instability, cancer and cell death (Dodge et al., 2004; Fodor et al., 2010; Hahn et al., 2010; Peters et al., 2001; Schotta et al., 2008). However, to which extent heterochromatin proteins are involved in transcriptional gene regulation in mammals and which heterochromatin proteins participate in this process is largely unknown.

In this thesis I utilized the Intracisternal A-type Particle (IAP) retrotransposon of mice as a model system for heterochromatin formation. I identified a novel DNA sequence element of IAP retrotransposons that autonomously recruits heterochromatin formation and silences even strong constitutively active promoters in cis in mouse embryonic stem cells. I generated an EGFP-based silencing assay utilizing this IAP sequence and combined it with RNAi experiments to test genes for their involvement in heterochromatic gene silencing. Reporter gene silencing requires the H3K9 histone methyltransferase Setdb1 and the transcriptional corepressor protein Trim28, which are both known regulators of endogenous IAP retrotransposon silencing (Matsui et al., 2010; Rowe et al., 2010). Having established a reporter system that phenocopies endogenous IAP silencing, I sought to identify additional proteins involved in heterochromatin regulation by genome-wide RNAi-screening.

I found that the SNF2-type ATPase Atrx and the histone H3.3 chaperone Daxx are required for rapid heterochromatic gene silencing. Notably, Atrx is also enriched at endogenous IAP elements. Cells that completely lack Atrx or Daxx show a severe delay of reporter silencing but are not completely deficient in heterochromatin formation. Consistently, endogenous IAP elements are not derepressed in *Atrx* or *Daxx* knockout ES cells. However, when heterochromatin status is challenged by depletion of *Setdb1* or *Trim28*, *Atrx* knockout cells show a stronger derepression of endogenous IAP elements. This suggests that the Atrx/Daxx pathway suppresses heterochromatin plasticity at these elements. Consistently, recruitment of strong transcriptional activators to a heterochromatinized reporter gene only results in an activation of reporter expression when cells lack Atrx.

Surprisingly, rescue experiments in Daxx knockout cells reveal that reporter silencing is independent of H3.3 association with Daxx. This is consistent with the finding that cells lacking H3.3 show no defect in reporter gene silencing.

In summary, this thesis unravels a novel, H3.3-independent role of Atrx and Daxx in catalyzing heterochromatin formation.

III. Zusammenfassung

Ein großer Teil des Säugetiergenoms ist in Heterochromatin verpackt. In *Knockout*-Studien mit Mäusen konnte gezeigt werden, dass Heterochromatin im Säugetier essentiell für das Überleben ist. So führen Defekte in der Etablierung oder Aufrechterhaltung von Heterochromatin zu Entwicklungsdefekten, Instabilität des Genoms, Krebs und Zelltod (Dodge et al., 2004; Fodor et al., 2010; Hahn et al., 2010; Peters et al., 2001; Schotta et al., 2008). Ob Proteine, die bei der Bildung von Heterochromatin beteiligt sind, auch eine direkte Auswirkung auf die Regulation der Genexpression in Säugetieren haben, ist aber weitgehend unbekannt.

In dieser Doktorarbeit, habe ich zunächst eine DNA-Sequenz im Genom der Maus entdeckt, die in der Lage ist, die Bildung von Heterochromatin zu veranlassen, wenn man diese Sequenz neu ins Genom von embryonalen Mausstammzellen integriert. Die Sequenz stammt aus sogenannten IAP-Elementen, einer Unterfamilie von LTR-Retrotransposons die natürlicherweise hundertfach im Mausgenom vorkommt. Das identifizierte Sequenzelement kann nicht nur die Bildung von Heterochromatin veranlassen, sondern auch die Genexpression von benachbarten aktiven Genen in Mausstammzellen abschalten. Ich etablierte ein System bei dem das Sequenzelement ein grün-fluoreszierendes Reportergen (EGFP) ausschaltet, wenn es ins Genom von Mausstammzellen integriert wird. Durch Analyse von Zellen, in denen Heterochromatinkomponenten depletiert wurden, konnte gezeigt werden, dass die Histonmethyltransferase Setdb1 und das transkriptionelle Korepressorprotein Trim28 essentiell an der Abschaltung des EGFP-Reportergens durch das identifizierte Sequenzelement beteiligt sind. Für Trim28 und Setdb1 wurde bereits beschrieben, dass sie die Expression von endogenen IAP-Retrotransposon-Sequenzen in Mausstammzellen unterdrücken (Matsui et al., 2010; Rowe et al., 2010). Da das generierte EGFP-Reportersystem klare Eigenschaften widerspiegelt, die auch bei der Abschaltung von endogenen Retrotransposons eine Rolle spielen, wurde es benutzt um auch andere Gene zu finden, die bei der Stilllegung von Retrotransposons eine Rolle spielen. Dazu kombinierte ich die EGFP-Reportergenstudien mit einem RNA-Interferenz-Screen. Ich fand heraus, dass das EGFP-Reportergen nur noch sehr stark verlangsamt ausgeschaltet werden kann, wenn das Histon H3.3 Chaperon-Protein Daxx und die ATPase Atrx fehlen. Interessanterweise bindet das Atrx-Protein an endogene IAP-Retrotransposon-Sequenzen. Durch Re-Expression von Daxx in *Daxx Knockout*-Zellen konnte ich zeigen, dass die Bindung von Histon H3.3 durch Daxx nicht notwendig ist für die Funktion von Daxx bei der untersuchten Genabschaltung. Dieses Ergebnis ist mit der Beobachtung konsistent, dass Zellen in denen Histon H3.3 depletiert wurde das EGFP-Reportergen immer noch genauso gut ausschalten können wie wild-typische Zellen. Obwohl diese Beobachtungen eine direkte Rolle von Atrx und Daxx bei

der Abschaltung von endogenen IAP-Retrotransposons nahelegen, sind endogene IAP-Retrotransposon weder in *Atrx*-defizienten Zellen noch in *Daxx*-defizienten Zellen dereprimiert. Wird allerdings das Heterochromatin durch eine verminderte Expression von Trim28 oder Setdb1 beeinträchtigt, so reagieren Zellen denen Atrx fehlt mit einer stärkeren Derepression von IAP-Retrotransposons. Außerdem lassen sich Gene die von Heterochromatin umgeben sind, deutlich leichter wieder aktivieren, wenn Atrx in den Zellen ausgeschaltet wurde.

Atrx und Daxx sind somit für die Repression von endogenen Retrotransposons nicht erforderlich, sorgen aber für eine geringe Plastizität des Heterochromatins und beschleunigen und stabilisieren somit dessen Entstehung an repetitiven Sequenzen im Genom.

1. Introduction

1.1. Heterochromatin regulation

1.1.1. DNA, chromatin and a definition of heterochromatin

Nuclear DNA of eukaryotic cells is spooled around octameric protein complexes consisting of basic histone proteins. The formed structures are called nucleosomes and consist of roughly 146 bp of DNA wrapped in 1.7 turns around a core of eight histone proteins. Two copies of each highly evolutionary conserved canonical histone protein H2A, H2B, H3 and H4 form the protein core of a nucleosome (Kornberg, 1974; Luger et al., 1997). In addition, specialized variants of different histones have evolved during evolution to exhibit specialized functions (Talbert and Henikoff, 2010). The association of DNA with histone proteins and non-histone proteins is called chromatin, named by Walther Flemming, who was among the first researchers describing cellular structures during cell division (Flemming, 1882).

Since most somatic cells in an organism share the same DNA content, but every different cell type needs to express a different set of genes, gene expression has to be tightly regulated. The regulation of gene expression happens mainly on the level of transcription. Research in the last decades has unraveled that chromatin regulation is a key process that facilitates and maintains these transcriptional programs (Allis et al., 2007). Chromatin is constantly covalently and non-covalently modified by proteins and RNA to regulate transcription and other nuclear processes such as DNA replication or DNA repair. Many ways of chromatin regulation have been described including chemical modifications of cytosine DNA nucleotides, chemical modifications of histone proteins, sliding, exchange and replacement of histone proteins by chromatin remodeling enzymes and interactions of regulatory RNA with chromatin (Allis et al., 2007; Becker and Horz, 2002; Bernstein and Allis, 2005; Jones, 2012; Kouzarides, 2007). For a more detailed introduction into general chromatin biology the reader is referred to reviews and books that can give a broader overview (Allis et al., 2007; Kouzarides, 2007).

In the early 20th century, optically dense chromatin regions inside the cell nucleus have been discovered and termed heterochromatin and specialized genomic features of these chromatin regions have been proposed (Heitz, 1928). Even though this descriptive definition of heterochromatin holds true for most cells where repetitive, non-coding DNA sequences like centromeres and pericentric regions are constitutively heterochromatinized, already observations from the first half of the 20th century indicated that normally euchromatic regions can also be heterochromatinized under certain conditions (Brown, 1966). These findings prepared the ground for a modern definition of heterochromatin, which is based on the

functional state of the chromatin domain rather than its chromosomal localization, the underlying DNA sequence or the optical density.

However, the parameters that determine the functional state of a chromatin domain are not defined precisely. Genome-wide analysis of histone modifications, binding of chromatin proteins, nuclease accessibility and genome transcription can define different types of chromatin domains (Consortium, 2004; Consortium et al., 2012; Filion et al., 2010; Kharchenko et al., 2011). When considering these datasets for a definition of heterochromatin, a heterochromatic domain is often defined as having the functional molecular features normally present at known constitutive heterochromatic regions like pericentric heterochromatin. Such molecular properties have already been used for a heterochromatin definition earlier (Huisinga et al., 2006). These would include: Trimethylation of lysine 9 in histone H3 (H3K9me3), Trimethylation of lysine 20 in histone H4 (H4K20me3), cytosine methylation of DNA, chromatin-association of Hp1 proteins and DNAseI inaccessibility (Huisinga et al., 2006).

By now, a lot of experimental evidence underlines the fundamental importance of heterochromatin for basic cellular processes, mammalian development and consequently many human diseases (Dambacher et al., 2010; Hahn et al., 2010). However, the functional role and the regulation of heterochromatin are still incompletely understood.

1.1.2. Regulation pathways at pericentric heterochromatin in mouse

Most of our knowledge about heterochromatin in mammals comes from analyses of mouse pericentric heterochromatin that occludes highly repetitive major satellite DNA close to the centromeres. Pericentric heterochromatin is established in a sequential pathway (Figure 1.1). How heterochromatin-modifying enzymes are recruited to major satellite sequences in the first place is not understood, but binding of transcription factors Pax3 and Pax9 are involved in priming major satellite sequences for heterochromatin formation (Bulut-Karslioglu et al., 2012). Next, different methyltransferases catalyze H3K9me1 at pericentric heterochromatin (Loyola et al., 2009; Pinheiro et al., 2012). Yet, it is not fully clear how these enzymes are recruited to major satellites and whether Prdm3, Prdm16 and Setdb1 are partially redundant in this process. Because Suv39h enzymes catalyze H3K9me3 at pericentric heterochromatin (Peters et al., 2001), and knockout of *Suv39h* results in an accumulation of H3K9me1 at pericentric heterochromatin (Peters et al., 2003), generation of H3K9me1 has to be an early and important step in heterochromatin formation. In addition, many H3K9 methyltransferases also directly interact, assuming that H3K9me1 could also be established by collaborative action of different enzymes (Fritsch et al., 2010). Furthermore, H3K9me1 and H3K9me2 are not limited to heterochromatic regions in general and are also abundant at other chromatin

regions making it complicated to unravel their specific deposition at heterochromatin (Lienert et al., 2011a).

Upon generation of H3K9me3 by Suv39h enzymes Hp1 proteins can directly bind this modification via their chromodomains (Bannister et al., 2001; Lachner et al., 2001). Hp1 proteins in turn recruit Suv4-20h enzymes that catalyze the formation of H4K20me3 and stabilize cohesin proteins at pericentric heterochromatin (Hahn et al., 2013; Schotta et al., 2004; Schotta et al., 2008). Importantly, H3K9me3 is also responsible for recruitment of DNA methylation at heterochromatic regions (Lehnertz et al., 2003).

In addition to this sequential establishment of heterochromatin marks and the recruitment of heterochromatic enzymes, a lot of other proteins have been found to bind either Hp1 or H3K9me3 (Dambacher, 2013; Hediger and Gasser, 2006; Kwon and Workman, 2011; Nozawa et al., 2010; Vermeulen et al., 2010). Consequently, many of these proteins have been shown to localize to pericentric heterochromatic foci in mouse cells (Fodor et al., 2010). The detailed mechanisms of how individual heterochromatic proteins act to ensure heterochromatin integrity are mostly unknown.

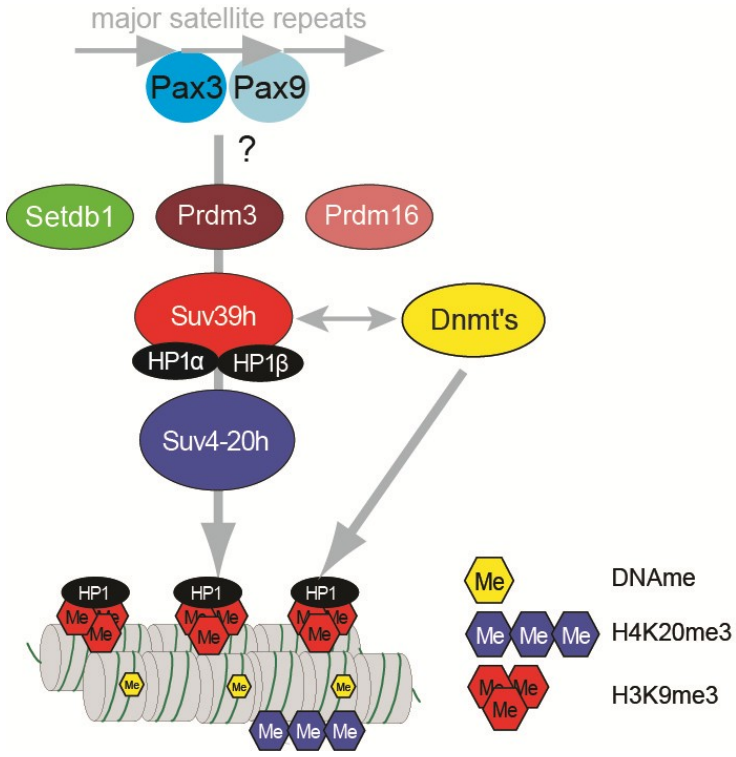


Figure 1.1 A silencing pathway for pericentric heterochromatin
Major satellite repeats are bound by Pax3 and Pax9 and facilitate the recruitment of histone methyltransferases. Prdm3, Prdm16 or Setdb1 generate H3K9me1. Suv39h enzymes can take mono-methylated H3K9 as a substrate to generate H3K9me3. HP1 binding to H3K9me3 recruits Hp1-interacting proteins like Suv4-20h enzymes that catalyze H4K20me3. DNA-methylating enzymes (Dnmts) are recruited by H3K9me3 (modified from (Dambacher et al., 2010; Schotta et al., 2004)).

For Hp1 proteins and the H3K9me3 chromatin mark a spreading mechanism was postulated, that is driven by oligomerization of Hp1 on H3K9me3-modified chromatin fibers (Canzio et al., 2011; Hall et al., 2002; Hathaway et al., 2012; Verschure et al., 2005). Because Hp1 proteins and the H3K9me3-modification can recruit H3K9 methyltransferases like Suv39h and Setdb1, this leads to the formation of additional H3K9me3-marks which again are bound by Hp1 proteins and further extend the heterochromatic domain.

Even though pericentric heterochromatin seems to be a very dense and stable structure, and Suv39h2 and Suv4-20h2 enzymes stably associate with chromatin (Hahn et al., 2013), mobility of Hp1 proteins in the cell is high, showing that some heterochromatin components undergo a rapid turnover (Cheutin et al., 2003). In addition, there is even a higher mobility of heterochromatin proteins in pluripotent cells, indicating that heterochromatin plasticity is influenced in a cell type-specific manner (Meshorer et al., 2006). Therefore, heterochromatin plasticity and protein dynamics present another level of complexity of how heterochromatin is regulated and maintained.

Recent biochemical and genetic studies have uncovered that heterochromatin pathways and proteins also play a role outside of pericentric heterochromatin, were they are involved in the suppression of gene expression, and the control of enhancer elements and transposons (Bulut-Karslioglu et al., 2012; Consortium, 2004; Consortium et al., 2012; Filion et al., 2010; Karimi et al., 2011; Kharchenko et al., 2011; Matsui et al., 2010; Nicetto et al., 2013; Rowe et al., 2013b; Rowe and Trono, 2011). A prominent example is the selective control of olfactory receptor expression in mammalian olfactory neurons (Magklara et al., 2011). In addition, defects in heterochromatin integrity have been linked to cancer and a variety of other human diseases (Ceol et al., 2011; Hahn et al., 2010; Kulis and Esteller, 2010; Ting et al., 2011; Yokoe et al., 2010; Zhu et al., 2011). Due to its essential role in genome regulation a deeper understanding of the molecular mechanisms at heterochromatin and the regulation of heterochromatin plasticity is required. Therefore, I focused on the functional analysis of heterochromatic gene silencing and wanted to uncover the associated molecular mechanisms.

1.2. Retrotransposon silencing

Apart from pericentric heterochromatin, heterochromatin modifications can also be found at many retrotransposon sequences (Consortium, 2004; Consortium et al., 2012; Filion et al., 2010). In this thesis I focused on a subclass of LTR-retrotransposons as a model system to investigate heterochromatin formation at these particular sequences. Therefore, I will give an overview over our current knowledge about chromatin-dependent LTR-retrotransposon silencing in the following chapter.

Understanding the regulation of retrotransposons is important, because around 40 % to 60 % of the human and mouse genome consists of retrotransposons sequences which are implicated into many essential process (de Koning et al., 2011; Lander et al., 2001; Mouse Genome Sequencing et al., 2002). Retrotransposons are autonomous DNA sequences that colonized the genome via a copy and paste mechanism. Upon transcription the retrotransposon RNA gives rise to proteins that facilitate its reverse transcription into cDNA and stably integrate the cDNA into the cellular genome (Coffin et al., 1997). In mammals three different subclasses of retrotransposons exist: LINEs, SINEs and LTR-retrotransposons.

1.2.1. LTR-retrotransposons in mammals

LTR retrotransposons make up around 10 % of the mammalian genome and are present in most animals and plants (Coffin et al., 1997). They received their name from carrying long-terminal-repeat (LTR) sequences at both ends of the transposon sequence (Stocking and Kozak, 2008). In general, LTR retrotransposons encode a structural capsid protein (GAG), a reverse transcriptase (RT), an integrase (IN), an RNAse-helicase and a protease activity (Stocking and Kozak, 2008). Not all of these enzymatic activities are distributed among individual proteins but can be united in multifunctional proteins. LTR retrotransposons most likely originated by a fusion of a DNA-transposon, a non-LTR retrotransposons and host factors early in evolution (Malik and Eickbush, 2001). In mammals all LTR-retrotransposons belong to the endogenous retrovirus (ERV) superfamily, which leads to the fact that in mammals the term LTR-retrotransposon and the term endogenous retrovirus are mostly used synonymously (Stocking and Kozak, 2008). Endogenous retroviruses are closely related to modern exogenous retroviruses and therefore can be grouped according to their sequence similarity with exogenous viruses into three different classes (McCarthy and McDonald, 2004; Stocking and Kozak, 2008) (Figure 1.2A).

Although the abundance of endogenous retroviruses within the mouse and the human genome are quite similar, many endogenous retroviruses are capable of active transposition in mice but not in humans, where only very few copies of the HERV-K family are still active (Belshaw et al., 2005; Medstrand and Mager, 1998). Conversely, around 10 % of *de novo* mutations in inbred mouse strains originate from transposition events of endogenous retroviruses (Maksakova et al., 2006). The most active endogenous retrovirus classes in mice are Intracisternal A-type Particle (IAP) transposons with an estimated number of around 300 functional copies and MusD/ETn retrotransposons with ninefunctional copies (Dewannieux et al., 2004; Kuff and Lueders, 1988; Lueders and Kuff, 1977; Ribet et al., 2004; Ribet et al., 2008). In addition, many mutated copies and subclasses of these

elements exist in the mouse genome, which in total contains around 350 MusD/ETn copies and around 1000 IAP

A

	Class I	Class II	Class III
related exogenous retrovirus genera	Gamma RV, Epsilon RV	Lentivirus, Alpha RV, Beta RV, Delta RV	Spumaviruses
Mouse endogenous retroviruses	MuLV, VL30, MuRRS, MuERV-C, McERV ...	IAPs, MusD/ETn, MMTV	MuERV-L/MaLR
Human endogenous retroviruses	HERV-H, HERV-W	HERV-K	HERV-L

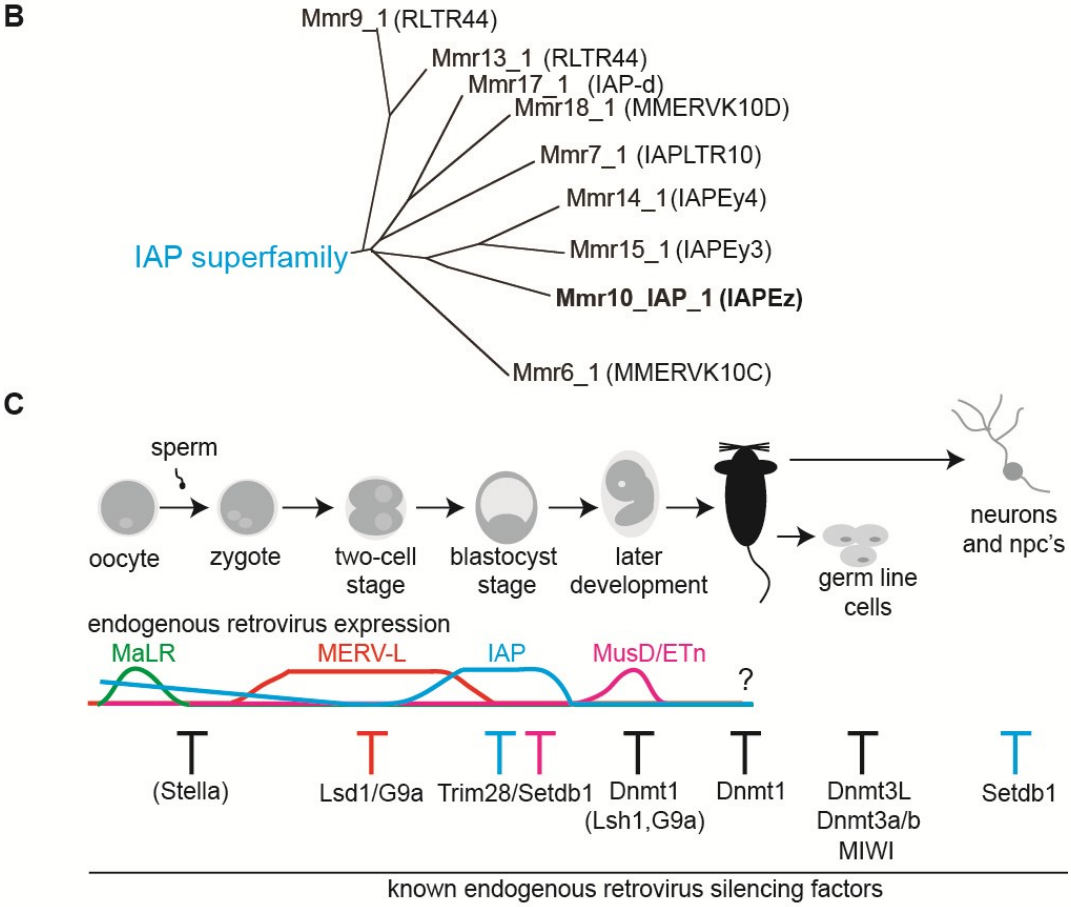


Figure 1.2 Classification of mammalian endogenous retroviruses
A) Endogenous retroviruses present in the mouse and human genome can be grouped according to their similarity with the genera of exogenous retroviruses (modified from (Rowe and Trono, 2011; Stocking and Kozak, 2008)). B) Dendrogram of the IAP superfamily that shows the evolutionary distance between the subfamilies of IAP retrotransposons in the mouse genome. The analysis is based on the sequence homology of the amino acid sequence of the encoded RT gene. The IAPEz family (indicated in bold) is the most active subfamily which was also analyzed in more detail in this thesis. Annotations used by the Repeat Masker database have manually been added behind the automated Mus musculus retrotransposon (Mmr) annotation (modified from (McCarthy and McDonald, 2004)). C) Diagram showing the expression of individual endogenous retroviruses (ERVs) during mouse development in a simplified way (modified from (Rowe and Trono, 2011)). Some of the known ERV silencing factors are listed at the specific time points of development. Factors that cooperate to maintain DNA methylation are listed in brackets.

copies grouped into various subclasses (Ribet et al., 2004; Stocking and Kozak, 2008) (Figure 1.2B). Even though retrotransposition activity of human endogenous retroviruses is almost lost in humans, it has been high 40 million years ago, which explains the high percentage of endogenous retroviral sequences in the human genome (Gifford and Tristem, 2003). However, active autonomous retrotransposons in humans still exist today in terms of non-LTR Line L1 retrotransposons (Beck et al., 2010).

Transposition events of endogenous retroviruses can influence the expression of nearby genes which has been linked to a lot of disease-causing phenotypes in humans and mice (Jern and Coffin, 2008; Kurth and Bannert, 2010; Rowe and Trono, 2011). However, mammals have also profited on endogenous retroviral elements in their evolution by adapting them as regulatory promoter sequences for endogenous gene regulation (van de Lagemaat et al., 2003). The expression of amylase in human saliva, for example, is a result of an insertion of an endogenous retrovirus (Ting et al., 1992). In addition, some essential mammalian genes even evolved directly from retrotransposon sequences, such as the placental syncytin proteins (Blaise et al., 2003; Blaise et al., 2005; Blond et al., 2000; Dupressoir et al., 2009). One of the most important impacts on human evolution might be that retrotransposon sequences are important recombination hotspots in mammalian meiosis. They are specifically targeted by Prdm9 to recruit the meiotic recombination away from genes that might otherwise be disrupted by meiotic recombination more frequently (Baudat et al., 2010; Berg et al., 2010; Brick et al., 2012; Grey et al., 2011; Parvanov et al., 2010).

However, independent of their ability to perform retrotransposition events, endogenous retrovirus sequences contain promoter and enhancer sequences and are expressed at different stages during mouse development (Rowe and Trono, 2011) (Figure 1.2C). Expression of retrotransposons at developmental stages has probably been evolutionary selected, because retrotransposon events are only transmitted to the germ line and consequently to future generations when occuring during the development of the organism or during the generation of its germ cells. In mouse, MaLR retrotransposon transcripts are specifically found in the oocyte (Peaston et al., 2004), while MERV-L retrotransposons can be detected at the 2-cell and are silenced at the blastocyst stage (Evsikov et al., 2004; Kigami et al., 2003; Maksakova et al., 2013; Peaston et al., 2004; Svoboda et al., 2004). IAP element transcripts are found in the oocyte, but peak in their expression at the blastocyst stage before they are efficiently silenced in later stages of development (Poznanski and Calarco, 1991; Svoboda et al., 2004).

Mammals have developed numerous ways to repress the endogenous retrovirus life-cycle in order to maintain evolutionary fitness. The most important pathway is the suppression of

endogenous retrovirus transcription by chromatin regulation. As a consequence a complex system consisting of different silencing mechanisms has evolved in evolution that regulates thousands of endogenous retroviruses and endogenous retrovirus-derived regulatory elements (Gifford et al., 2013; Leung and Lorincz, 2012; Rowe and Trono, 2011). These silencing mechanisms are summarizes in the following paragraphs and have also been adapted for chromatin regulation outside of endogenous retroviral sequences. A lot of the summarized information about retrotransposon silencing presented here originates from three excellent reviews, that are highly recommended to the interested reader (Gifford et al., 2013; Leung and Lorincz, 2012; Rowe and Trono, 2011).

1.2.2. Silencing of endogenous retroviruses by DNA methylation

DNA methylation in mammals involves the methylation of cytosine residues primarily at CpG dinucleotides by Dnmt1, Dnmt3a and Dnmt3b. Dnmt3a and Dnmt3b can *de novo* methylate DNA by collaborating with the non-enzymatic Dnmt3L protein, while Dnmt1 acts together with Np95 (Uhrf1) to maintain DNA methylation during DNA replication (Bird, 2002; Jones and Liang, 2009; Saitou et al., 2012).

Even though *Dnmt1* and *Dnmt3b* knockout mice die at embryonic day 8.5 and 14.5, preimplantation stages in these knockout mice are not compromised (Okano et al., 1999; Walsh et al., 1998). One explanation for this could be that the DNA is globally demethylated at this time point of development and starts to be newly methylated at the blastocyst stage (Saitou et al., 2012).

Analyses of knockout mice revealed that loss of *Dnmt3a* and *Dnmt3b* only leads to a very mild reduction of DNA methylation at IAP elements of mouse ES cells, while loss of *Dnmt1* leads to almost a complete loss of global DNA methylation (Arand et al., 2012; Okano et al., 1999). Dnmt1 knockout embryos show 100-fold upregulation of IAP retrotransposons at embryonic day 8.5 indicating that DNA-methylation by Dnmt1 is responsible for the repression of IAP elements at this embryonic stage (Walsh et al., 1998). Consistently, DNA-methylation inhibitor treatment leads to a strong derepression of IAP elements in MEF cells (Rowe et al., 2013a). In contrast to these observations knockout of *Dnmt1*, *Dnmt3a*, *Dnmt3L* or *Dnmt3b* in embryonic stem cells does not lead to defects in endogenous and exogenous retrovirus silencing suggesting that endogenous retrovirus silencing at preimplantation stages is independent of DNA methylation (Hutnick et al., 2010; Pannell et al., 2000; Quenneville et al., 2012). However, Dnmt3L is essential for IAP repression in the male germ line where it facilitates the *de novo* DNA methylation of IAP elements by Dnmt3a and 3b (Bourc'his and Bestor, 2004; Kato et al., 2007). Interestingly, depletion of Dnmt3 family proteins in somatic cells has only a very mild effect on endogenous retrovirus silencing (Okano et al., 1999). This could be explained by the fact that *de novo* DNA methylation of

endogenous retrovirus elements is not required in preimplantation embryos because DNA methylation at endogenous retroviral elements is not completely lost during global DNA demethylation (Gaudet et al., 2004).

Apart from Np95 and Dnmt3L other proteins are involved in maintaining and facilitating DNA methylation at endogenous retroviral sequences by Dnmt enzymes. The SNF2 chromatin-remodeling factor Lsh1 (*Hells*) binds endogenous retroviral elements and is important for the maintenance of DNA-methylation and heterochromatic marks at these regions (Dennis et al., 2001; Yan et al., 2003). *Lsh1* knockout mice have a strong reduction of DNA methylation at endogenous retroviral elements and consequently show elevated expression of IAP retrotransposons at later stages of development (Huang et al., 2004). Similarly the Stella protein (also known as Pgc7, Dppa3) is critical to maintain DNA methylation at IAP elements and imprinting regions specifically at the maternal genome after fertilization (Nakamura et al., 2007). Moreover, the H3K9 methyltransferase G9a is involved in the maintenance of DNA methylation at endogenous retroviruses and the establishment of DNA methylation at newly introduced MLV proviruses (Dong et al., 2008; Leung et al., 2011; Tachibana et al., 2008). However, the catalytic activity of G9a that facilitates methylation of H3K9 is dispensable for the maintenance of DNA methylation, and H3K9me3 marks at endogenous retroviruses are unaltered in *G9a* knockout ES cells indicating that G9a does not catalyze H3K9 methylation at these regions (Dong et al., 2008).

In summary, DNA methylation seems to be the major pathway that restricts endogenous retrovirus expression in the germline and in postimplantation embryos. Factors like G9a, Stella and Lsh1 seem to cooperate with the DNA methylation machinery to regulate DNA methylation at specitic stages of development. Interestingly, it is suggested that DNA methylation was originally developed in evolution as a host defense mechanism to suppress retrotransposition (Yoder et al., 1997). However, in preimplantation embryos silencing of endogenous retroviruses is independent of DNA-methylation and mouse ES cells silence newly introduced IAP elements and MLV retroviruses (Pannell et al., 2000; Rowe et al., 2010; Teich et al., 1977). This indicates that additional mechanisms of endogenous retrotransposon silencing occur at preimplantation stages.

1.2.3. Repression of (endogenous) retroviruses by H3K9me3, Setdb1 and the corepressor Trim28

The finding that DNA methylation is dispensable for endogenous retrovirus silencing in preimplantation embryos and embryonic stem cells suggests that other mechanisms ensure the repression of endogenous retroviruses during this time window (Figure 1.2C). In embryonic stem cells the heterochromatic modifications H3K9me3 and H4K20me3 are strongly enriched at many endogenous retroviruses including MusD/ETn and IAP elements

indicating that histone methylation is involved in that process (Mikkelsen et al., 2007). Even though Suv39h enzymes are responsible for catalyzing H3K9me3 at pericentric heterochromatin, H3K9me3 is reduced but not lost at endogenous retroviral sequences in *Suv39h* knockout cells (Bulut-Karslioglu et al., 2012).

Instead, the histone methyltransferase Setdb1 (Eset) is the responsible enzyme for catalyzing H3K9 methylation at endogenous retroviruses (Matsui et al., 2010). Depletion of *Setdb1* in mouse ES cells not only leads to a loss of H3K9me3 methylation but also results in a strong upregulation of many endogenous retroviruses classes (Karimi et al., 2011; Matsui et al., 2010). Interestingly, the loss of Suv4-20h enzymes and H4K20me3 does not result in a derepression of endogenous retrotransposons indicating that this chromatin mark is not strongly involved in transcriptional repression (Matsui et al., 2010).

Similar to *Setdb1* knockout cells, a severe derepression of endogenous retrotransposons was observed when the Setdb1-interacting protein Trim28 (Kap1) was depleted from mouse ES cells (Rowe et al., 2010). Depletion of *Trim28* reduces H3K9me3 at endogenous retroviral elements indicating that Trim28 might act upstream or in concert with Setdb1 (Rowe et al., 2013b). Interestingly, loss of *Trim28* or *Setdb1* in MEF cells does not lead to a depression of endogenous retrotransposons, indicating that both proteins are dispensable for retrotransposon silencing at later stages of development (Matsui et al., 2010; Rowe et al., 2010).

Therefore, the Trim28/Setdb1/H3K9me3 pathway is essential to repress endogenous retroviruses during phases of development when silencing is DNA methylation-independent. Strikingly, neuronal progenitor cells that lack Setdb1 are also modestly impaired in IAP silencing (Tan et al., 2012). Therefore, endogenous retroviruses might also be controlled by a Setdb1-dependent and DNA methylation-independent pathway in the neuronal lineage. Recently, active retrotransposition events of non-LTR retrotransposons have been observed in neurons of humans and *Drosophila,* indicating that a specific retrotransposon control in the brain is evolutionary conserved and might contribute to specific functions of neurons (Coufal et al., 2009; Garcia-Perez et al., 2010; Muotri et al., 2005; Perrat et al., 2013) .

However, how the Trim28/Setdb1 complex targets endogenous retrotransposons is unknown. Because Trim28 is recruited by KRAB domain containing zinc finger proteins to the DNA it was suggested that sequence specific KRAB zinc finger proteins distinguish the different retrotransposon classes and recruit the Trim28/Setdb1 complex to its genomic targets (Rowe et al., 2010) (also see introduction section 1.3.1). This is consistent to the observation that there is a high coevolution between KRAB zinc finger proteins and retrotransposons (Thomas and Schneider, 2011). Interestingly, introducing a human

chromosomal fragment into transgenic mice, results in a derepression of human endogenous retrovirus sequences, indicating that mice might not have the recognition system to silence human endogenous retroviruses (Ward et al., 2013). Another piece of evidence for the hypothesis that endogenous retroviruses are targeted by sequence specific factors was the finding that exogenous M-MLV retroviruses are silenced by Zfp809, a mouse-specific KRAB zinc finger protein that recruits Trim28 to facilitate H3K9me3-dependent silencing of the M-MLV provirus (Wolf and Goff, 2007, 2009). Incidentally, M-MLV silencing is also modulated by *Yy1* and *Ebp1* although these proteins only enhance the silencing efficiency by Trim28 and are not strictly required for M-MLV silencing (Schlesinger et al., 2013; Wang et al., 2014a).

Future experiments have to show if KRAB domain-containing zinc fingers recruit the Trim28/Setdb1 silencing complex to endogenous retroviruses. This would imply that endogenous retroviruses contain specific DNA binding sites for KRAB domain zinc finger proteins.

Furthermore, how Setdb1- and Trim28-mediated silencing of endogenous retroviruses occurs mechanistically is unclear. In this thesis I identify a critical role of the proteins Atrx and Daxx in this process. Chapter 1.3 of this introduction gives a broader overview about the current knowledge of Trim28- and Setdb1-dependent silencing and Atrx and Daxx.

1.2.4. Repression of endogenous retroviruses independent of DNA methylation and the Trim28/Setdb1/H3K9me3-pathway

Expression of MERV-L retrotransposons has been linked to the two cell stage (Figure 1.2C), and is restricted by Lsd1, HDACs and G9a at the exit of totipotency (Macfarlan et al., 2011; Maksakova et al., 2013). Interestingly, MERV-L retrotransposon silencing in mouse ES cells is independent of *Setdb1* and only slightly affected upon knockout of *Trim28* underlining that they are silenced by an independent mechanism (Macfarlan et al., 2011; Maksakova et al., 2013). Mechanistically, it was proposed that Lsd1 silences MERV-L elements by demethylating the active chromatin mark H3K4me3 on MERV-L promoters, while G9a is depositing a repressive H3K9me2 mark at these regions (Macfarlan et al., 2011; Maksakova et al., 2013). LTR fragments derived from MERV-L retrotransposons are also used as functional promoter sequences for several cellular genes expressed at the two-cell stage, indicating that mice have exploited the silencing mechanism for MERV-L retrotransposons to establish a time-point specific gene expression mechanism (Macfarlan et al., 2011; Macfarlan et al., 2012).

In the germline of *Drosophila* a complex retrotransposon repression system that utilizes piRNAs has evolved. piRNAs are utilized by heterochromatin-associated RNAi proteins like

Piwi to detect and cleave retrotransposon RNAs (Ross et al., 2014). Heterochromatin formation and the establishment of H3K9me3 at repetitive regions is also dependent on RNA interference and small RNAs in plants and fission yeast, suggesting that these systems were very successful during evolution and might also exist in mammals (Castel and Martienssen, 2013). Indeed, mice express Piwi-like proteins in their germ-line called Miwi, Mili and Miwi2. The depletion of these proteins leads to sterility and results in a slight upregulation of retrotransposon sequences and defects in *de novo* methylation of DNA (Rowe and Trono, 2011). This indicates that small-RNA-dependent pathways might also be implicated in retrotransposon suppression in mammals.

Furthermore, in somatic cells specific corepressor complexes might collaborate with DNA-methylation pathways to repress particular endogenous retrotransposons. Recently the Trim24/Trim33 corepressor complex has been found to suppress VL30-retrotransposons in the liver (Herquel et al., 2013). These retrotransposons get activated by retinoic acid receptors and failure of retrotransposon silencing results in ectopic activation of nearby genes and an inflammatory cell response in hepatocytes (Herquel et al., 2013).

Failure of endogenous retrotransposon silencing has also been implicated in autoimmunity. In mice deficient for the exonuclease Trex1 endogenous retrotransposon transcripts accumulate in the cell and upon reverse transcription trigger a DNA induced interferon response that leads to autoantibody production (Stetson et al., 2008). Thus, cells that upregulate endogenous retroviruses under physiological conditions are harmful to the organism and are wiped out by the immune system due to the presentation of ERV antigens. A finding that supports this model is that antibody-deficient mice cannot remove cells that overexpress endogenous retroviruses (Young et al., 2012). In turn, a constant overexpression of endogenous retroviruses leads to recombination events that form functional retroviruses with oncogenic potential (Young et al., 2012).

Interestingly, it was also postulated that expression of non-functional endogenous retroviruses in early mouse development protects mice form exogenous retroviruses (Best et al., 1996). The most striking example of this theory is given by the Fv1 protein that is derived from a MERV-L Gag protein and blocks the integration of exogenous MLV (Best et al., 1996).

1.2.5. Concluding remarks about retrotransposon silencing pathways

In summary, retrotransposons and the mechanism that repress their activity are important for pluripotency, gene regulation, genome stability, meiotic recombination, development and many fundamental processes in mammals. Investigation of mammalian retrotransposon biology provides important information how mammalian genomes are shaped and evolve during evolution and how gene activity is regulated during development.

In this thesis I have focused on the regulation of Intracisternal A-type-Particle (IAP) retrotransposons, a very active subclass of LTR retrotransposons in mice. This transposon class is specifically silenced by the Trim28/Setdb1/H3K9me3 pathway in mouse ES cells. Therefore, it is a suitable model system to analyze H3K9me3-dependent chromatin regulation and repression of gene-expression outside of highly repetitive pericentric heterochromatin. In addition, IAP retrotransposons are an interesting example for studying how endogenous retrotransposons are specifically recognized and targeted during early mouse development.

1.3. Proteins involved in heterochromatic gene silencing in mouse

In this thesis I screened for proteins involved in heterochromatic silencing induced by small sequence elements of IAP retrotransposons in mouse ES cells. I identified the SNF2-type ATPase Atrx, the Histone H3.3 chaperone Daxx, the corepressor protein Trim28 and the H3K9 methyltransferase Setdb1 to be important regulators of heterochromatic gene silencing at these sequences. Thus, I will introduce these proteins in more detail.

1.3.1. KRAB zinc finger-dependent repression by the master regulator Trim28

Zinc finger transcription factors harboring a Krüppel-associated box (KRAB) are tetrapod-specific and are the largest family of zinc finger transcription factors encoded in the human and mouse genome (Corsinotti et al., 2013; Emerson and Thomas, 2009; Tadepally et al., 2008; Thomas and Emerson, 2009). The KRAB domain is highly conserved among KRAB domain proteins and mediates strong transcriptional repression (Bellefroid et al., 1991; Margolin et al., 1994). Trim28 (also known as Kap1, Tif1β) is the only adapter protein known that binds KRAB domain proteins and it is the master regulator of KRAB domain-dependent silencing processes (Friedman et al., 1996; Kim et al., 1996; Le Douarin et al., 1996; Moosmann et al., 1996). Although Trim28 seems to be the only protein that binds KRAB domains, it is part of the Transcription-Intermediary-Factor protein family that also comprises the proteins Trim24, Trim33 and Trim66 (Iyengar and Farnham, 2011; Peng et al., 2002). Trim28 contains an N-terminal RING-B box-coiled-coil (RBCC) domain, also called tripartite-motif (TRIM) that mediates binding and homotrimerization of three Trim28 proteins with a single KRAB domain (Peng et al., 2000a; Peng et al., 2000b) (Figure 1.3). In addition, Trim28 harbors a HP1-box containing a canonical PxVxL motif that is responsible for direct binding of Trim28 to Hp1 proteins (Brasher et al., 2000; Lechner et al., 2000; Smothers and Henikoff, 2000) (Figure1.3). At its C-terminus Trim28 has a PHD and bromodomain that does not bind acetylated lysine residues *in vitro*. Instead, the PHD domain has an intramolecular E3-Sumo ligase activity that sumoylates the neighboring bromodomain to facilitate Sumo-dependent interaction with Setdb1 and Chd3 (Ivanov et al., 2007; Schultz et al., 2002; Schultz et al.,

2001; Zeng et al., 2008). Importantly, transcriptional silencing by Trim28 depends on its Hp1-interaction motif, its auto-sumoylation-mediated interaction with Setdb1 and is associated with H3K9me3, which is catalyzed by Setdb1 (Ayyanathan et al., 2003; Sripathy et al., 2006; Wolf et al., 2008).

As mentioned, Trim28 is recruited to chromatin through direct interactions with DNA-binding zinc finger proteins that contain a KRAB domain. In this regard Trim28 has been shown to be required for genomic imprinting via binding to Zfp57 (Li et al., 2008; Quenneville et al., 2011), silencing of retroviral M-MLV elements via recruitment by Zfp809 (Wolf and Goff, 2007, 2009), morphogenesis of extraembryonic tissues via binding to Zfp568 (Shibata et al., 2011), and regulation of sex-specific liver-gene expression via binding to Rsl1 and Rsl2 (Krebs et al., 2009; Krebs et al., 2003). Additionally, the KRAB/Trim28 association has been implicated into neuronal differentiation, hematopoesis, mitophagy and maintenance of genome stability via suppression of endogenous retroviral elements (Barde et al., 2013; Jakobsson et al., 2008; Rowe et al., 2010; Rowe et al., 2013b; Santoni de Sio et al., 2012a; Santoni de Sio et al., 2012b). Interestingly, Trim28 can be recruited via KRAB domain proteins that lack a DNA binding domain, but are associated with other DNA-binding proteins. This mechanism was proposed for the DNA-binding proteins SRY and VHL (Li et al., 2003; Peng et al., 2009). Aside of the KRAB domain associated genomic recruitment of Trim28, there are also examples for KRAB-independent recruitment of Trim28 to DNA by interaction with other transcription factors (Iyengar and Farnham, 2011; Iyengar et al., 2011).

In addition to its role in transcriptional repression, Trim28 is also involved in DNA damage regulation. Here phosphorylation of Trim28 by DNA damage induced protein kinases is thought to prevent its sumoylation and consequently results in relief of transcriptional repression at the DNA damage site (Li et al., 2007; White et al., 2006).

Being implicated in so many cellular and physiological processes, Trim28 is essential for early mouse development and pluripotency of embryonic stem cells (Cammas et al., 2000; Fazzio et al., 2008; Hu et al., 2009; Wolf and Goff, 2007). Even Trim28 deletion from the female germ line is sufficient to severely impair mouse development, maybe due to imprinting defects or impaired epigenetic inheritance of important chromatin marks (Messerschmidt et al., 2012). Apart from its function in germ cells and normal development, Trim28 is also implicated into gastric cancer (Silva et al., 2006; Yokoe et al., 2010).

Although Trim28 function is well characterized using *in vitro* experiments where Trim28 is recruited by KRAB domain proteins (Groner et al., 2010; Groner et al., 2012; Quenneville et al., 2012; Rowe et al., 2013a; Sripathy et al., 2006), Trim28 DNA-binding *in vivo* is poorly understood because a lot of Trim28 binding sites do not colocalize with H3K9me3 or Setdb1

and are not located at genes that are deregulated upon Trim28 knockdown (Iyengar et al., 2011; Rowe et al., 2013b). An explanation for this finding could be that Trim28 mainly regulates transcription via the repression of distal retrotransposon-based enhancers, rather than by direct promoter regulation (Rowe et al., 2013b).

Since there is a high coevolution of KRAB zinc finger genes and retrotransposon sequences, the species-specific control of retrotransposons by Trim28 and KRAB zinc fingers links heterochromatic gene silencing to speciation and will be an interesting topic for future analyses (Thomas and Schneider, 2011). In addition, how Trim28-recruitment actually leads to a shutdown of transcription is still not fully understood. Hp1 and Setdb1 seem to be important players in this process. The role of Setdb1 is reviewed in the next chapter.

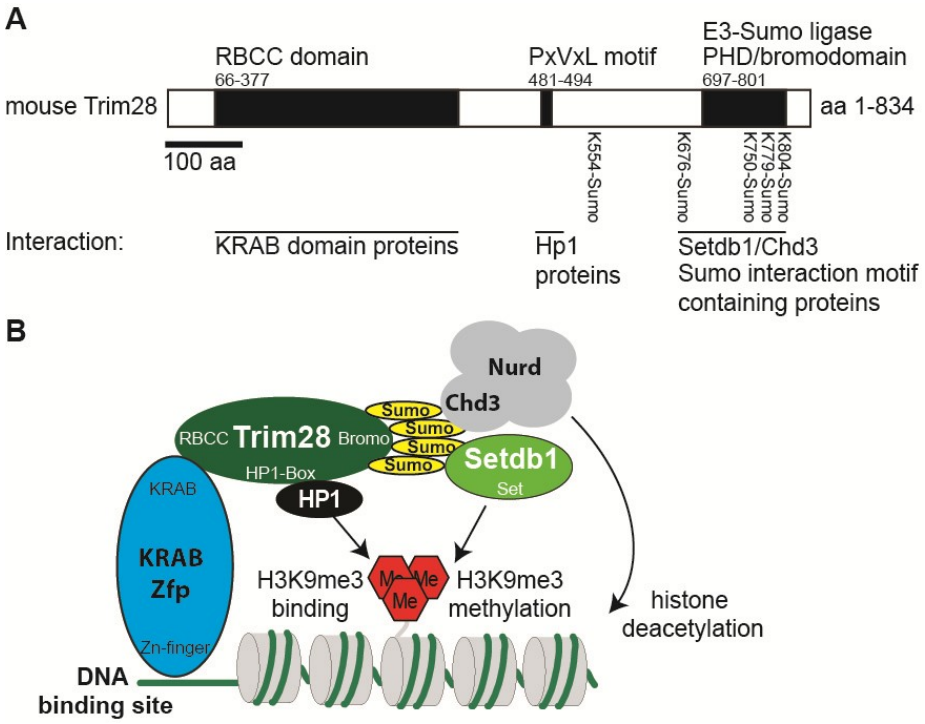


Figure 1.3 The corepressor Trim28
A) Scheme of mouse Trim28 containing an N-Terminal RBCC domain for binding KRAB domain proteins, an HP1-box for association with Hp1 proteins and a C-terminal PHD and Bromodomain. The PHD and Bromodomain is responsible for auto-sumoylation-dependent recruitment of Setdb1 and Chd3. Sumoylated residues of Trim28 are indicated. Numbers below protein domain names indicate the amino acid position. The protein is shown with the N-terminus on the left and the C-terminus on the right. B) Illustration how the repressive Trim28 complex forms. A KRAB domain protein recruits Trim28 to its binding site, which in turn recruits Hp1 proteins, Chd3 and Setdb1. Please note that *in vitro* assays have shown, that zinc finger proteins frequently bind DNA as dimers and one KRAB domain recruits three Trim28 molecules, underlining that there is a high degree of amplification in the binding cascade (modified from (Iyengar and Farnham, 2011; Rowe and Trono, 2011)).

1.3.2. The histone H3 lysine 9 methyltransferase Setdb1 (ESET) cooperates with Trim28 in heterochromatin formation and transcriptional repression

Setdb1 (also known as ESET or KMT1E) was originally identified in two yeast two-hybrid screens for proteins associated with the transcription factor ERG and the corepressor protein

Trim28 (Schultz et al., 2002; Yang et al., 2002). Interaction between Setdb1 and Trim28 is mediated by a Sumo-interaction motif (SIM) at the N-terminus of Setdb1 that binds to the sumoylated PHD and bromodomain of Trim28 (Figure 1.4) (Ivanov et al., 2007). Immuno-purified Setdb1 specifically monomethylates and dimethylates histone H3 lysine 9, but can also catalyze the formation of H3K9me3 when it is associated with its binding protein mAM (also known as Atf7ip or Mcaf1) (Schultz et al., 2002; Wang et al., 2003). Setdb1 is required for Trim28/KRAB-dependent repression and consequently also for repression of endogenous retroviruses by Trim28 (Matsui et al., 2010; Rowe et al., 2013a; Sripathy et al., 2006). Setdb1-mediated repression requires the catalytic activity of its SET domain (Matsui et al., 2010). Genetic experiments where Hp1 proteins are tethered to specific sites in the genome also have shown that Setdb1 is important for Hp1-mediated transcriptional repression and is recruited in a Hp1-dependent manner (Hathaway et al., 2012; Verschure et al., 2005).

Apart from its interaction with Trim28 and mAM, Setdb1 was found in a complex with the methyl-DNA binding protein MBD1 and the chromatin assembly factor CAF-1(p150) were it links H3K9-methylation to replication-dependent chromatin assembly (Sarraf and Stancheva, 2004). The interactions between MBD1 and Setdb1 is presumably mediated via MBD1 sumoylation and Sumo-dependent binding of Setdb1 (Lyst et al., 2006), or via the association of MBD1 with the Setdb1-interacting protein mAM (Ichimura et al., 2005). In addition, also the binding of MBD1 and mAM is dependent on MBD1 sumoylation suggesting that sumoylation plays an important role in recruiting the mAM/Setdb1 complex (Sekiyama et al., 2008; Uchimura et al., 2006). The chromatin assembly factor CAF-1(p150) also interacts with Hp1 proteins via a consensus PxVxL motif and its interaction with Setdb1 could be mediated via Hp1-proteins (Loyola et al., 2009).

Setdb1 has also been found to associate with Pml-nuclear bodies in mouse cells, where it was thought to regulate Pml-nuclear body integrity and gene expression (Cho et al., 2011). Since Pml is also a sumoylated protein, it was again hypothesized, that Setdb1 is recruited in a Sumo-dependent manner to Pml-nuclear bodies (Cho et al., 2011). This is consistent with the finding, that sumoylation of a single transcription factor is sufficient to recruit Setdb1 and to repress transcription (Stielow et al., 2008).

Similar to *Trim28*, *Setdb1* is essential for early mouse development and pluripotency (Bilodeau et al., 2009; Dodge et al., 2004). An explanation for this finding is that Setdb1 seems to repress the differentiation of cells of the inner-cell mass into the trophoblast lineage (Lohmann et al., 2010; Yuan et al., 2009). A more general explanation might be that like Trim28, Setdb1 is important to ensure stability of the transcriptional program in pluripotent cells. This is consistent with the finding that *Setdb1* depletion leads to derepression of

endogenous retrotransposons, activation of silenced enhancers and the generation of cryptic transcripts (Karimi et al., 2011; Rowe et al., 2013b).

Recent findings suggest that *Setdb1* is as a potential oncogene in melanoma and lung tumors and facilitates cell transformation *in vitro* (Ceol et al., 2011; Macgregor et al., 2011; Rodriguez-Paredes et al., 2013; Watanabe et al., 2008). Besides its function in transcriptional regulation, Setdb1 interacts with the DNA-methylation pathway, which could explain some of the Setdb1-dependent effects in cancer cells (Li et al., 2006).

In summary, Setdb1 is an essential H3K9 methyltransferase that is involved in developmental processes, crucial for the repressive activity of Trim28 and the maintenance of transcriptional programs.

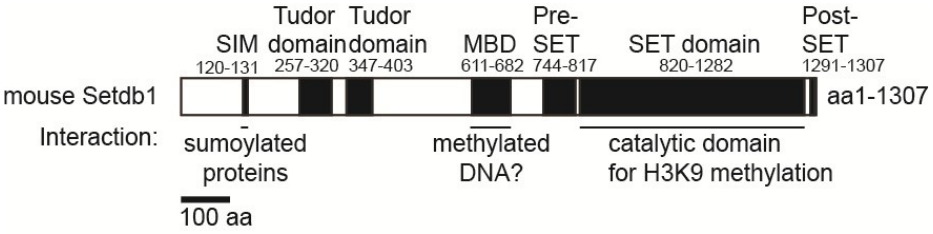


Figure 1.4 The histone H3 lysine 9 methyltransferase Setdb1
The illustration shows the domain architecture of the Setdb1 protein. The positions of known interaction surfaces of the Setdb1 protein are indicated. Numbers below protein domain names indicate the amino acid position. The protein is shown with the N-terminus on the left and the C-terminus on the right. The Sumo-interaction motif (SIM) was mapped in detail by Tanaka and Saitoh (Tanaka and Saitoh, 2010).

1.3.3. The chromatin-associated SNF2-type ATPase Atrx

In this thesis I found an important function of the SNF2-type ATPase Atrx in heterochromatic gene regulation. Atrx was originally identified as the gene mutated in Alpha-Thalassemia X-Linked Mental Retardation (ATRX) Syndrome, a genetic disease causing an altered facial morphogenesis, genital abnormalities, mental retardation and reduced expression of α-globin in red blood cells (alpha-thalassemia) (Gibbons, 2006; Gibbons et al., 1995). The *Atrx* gene gives rise to different splice isoforms which predominantly encode a full-length Atrx protein and a truncated Atrx protein which is usually expressed to a lesser extent (Garrick et al., 2004; Garrick et al., 2006) (Figure 1.5).

Disease causing mutations in the *Atrx* gene specifically cluster in the evolutionary highly conserved regions of the protein, which are the N-terminal ADD-domain and the seven C-terminal helicase motifs forming a large SNF2-type ATPase domain (Gibbons et al., 2008). Most mutations occurring in ATRX-syndrome patients either reduce the ATPase activity of the helicase domain or impair the capability of the ADD domain to bind unmethylated H3K4 together with di- or trimethylated H3K9 (Dhayalan et al., 2011; Eustermann et al., 2011; Iwase et al., 2011; Tang et al., 2004). Interestingly, the C-terminal helicase domain is missing

in the truncated protein isoforms of Atrx (Figure 1.5), assuming that the truncated isoforms might take over a specialized function. The function of the C-terminal SNF2-type helicase of Atrx is unknown. However, it has DNAse translocase activity, but shows only very weak nucleosome remodeling activity *in vitro*, which suggests that Atrx alone might not act as a classical nucleosome remodeler on chromatin (Xue et al., 2003).

Atrx localizes to pericentric heterochromatin via interactions of its ADD-domain with pericentric H3K9me3 and via interactions of a consensus PxVxL motif with Hp1 (Dhayalan et al., 2011; Eustermann et al., 2011; Iwase et al., 2011; Kourmouli et al., 2005; McDowell et al., 1999). In addition, Atrx localization to pericentric heterochromatin is lost in neurons upon depletion of the DNA-methyl-binding protein MeCP2 (Nan et al., 2007) indicating a possible connection between Atrx and DNA-methylation. Consistently, cells from ATRX-syndrome patients or *Atrx* knockout mouse embryos show changes of DNA-methylation at rDNA repeats and some other parts of the genome (Garrick et al., 2006; Gibbons et al., 2000). In addition, the N-terminal ADD domain of Atrx is also strongly similar to the ADD domain present in Dnmt proteins underlining a structural relationship between Atrx and other proteins involved in gene repression (Dhayalan et al., 2011; Eustermann et al., 2011; Iwase et al., 2011).

Atrx forms a complex with Daxx, which also recruits Atrx to Pml-nuclear bodies in a phosphorylation-dependent manner (Ishov et al., 2004; Tang et al., 2004; Xue et al., 2003). Daxx is a histone H3.3 specific-chaperone that catalyzes the incorporation of H3.3/H4 tetramers into repressive chromatin (Drane et al., 2010; Elsasser et al., 2012; Lewis et al., 2010; Liu et al., 2012). Apart from their colocalization at Pml nuclear bodies, Atrx colocalizes with Daxx at pericentric heterochromatin and is thought to facilitate incorporation of H3.3 into pericentric heterochromatin (Drane et al., 2010). H3.3 incorporation at pericentric heterochromatin correlates with major satellite expression. However, the specific function of histone H3.3 incorporation at pericentric heterochromatin remains elusive.

Heterochromatic chromatin modifications like H3K9me3, H4K20me3 and DNA methylation and many heterochromatin proteins are also present at telomeres, where they ensure telomere integrity (Benetti et al., 2007; Garcia-Cao et al., 2004; Gonzalo et al., 2006). Atrx is also found at telomeres where it colocalizes with the histone variant H3.3 (Wong et al., 2010), and where Atrx might cooperate with Daxx to incorporate histone H3.3 into telomeres (Goldberg et al., 2010; Lewis et al., 2010). However, Atrx is not strictly required for the Daxx-dependent assembly of H3.3 containing nucleosomes *in vitro*, but is important to recruit Daxx to telomeres in mouse ES cells (Lewis et al., 2010). Even though it has not been directly shown that Daxx incorporates histone H3.3 at telomeres, H3.3 incorporation is reduced at telomeres in *Daxx* or *Atrx* depleted cells, suggesting that the Atrx/Daxx complex

plays an important role of in this process (Goldberg et al., 2010; Lewis et al., 2010). Interestingly, inhibition of H3.3 incorporation at telomeres by depletion of *Atrx* only results in a very mild derepression of telomeric TERRA transcripts (Goldberg et al., 2010). Therefore, it is unclear, if the major molecular function of Atrx at telomeres is only the incorporation of histone H3.3.

Genome-wide binding profiles of Atrx indicate an enrichment of Atrx at tandem repeats, telomeres, pericentric regions and G-rich sequences (Law et al., 2010). Very importantly, H3.3 incorporation is not changed at Atrx target loci when Atrx is depleted indicating that the main function of Atrx outside telomeres is not to incorporate H3.3 (Law et al., 2010). Instead, it was postulated that secondary structures of the DNA generated by tandem repeat and G-rich sequences are resolved by Atrx (Law et al., 2010). These inhibitory DNA structures might interfere with transcription when *Atrx* is depleted (Law et al., 2010). Consistently, the length of tandem repeats at the *α-globin* locus correlates with the reduction of *α-globin* expression in ATRX-syndrome patients (Law et al., 2010). Functional analysis at the *α-globin* locus upon Atrx depletion indicates an elevated binding of the repressive histone variant macroH2A, suggesting an activating function of Atrx at the *α-globin* locus by inhibiting macro-H2A incorporation (Ratnakumar et al., 2012). However, how macro-H2A influences nearby transcription and whether Atrx directly influences macro-H2A incorporation is unknown.

Functional analyses of ATRX-syndrome mutations indicate that they are hypomorphic and result in a residual protein function (Gibbons et al., 2008). This is consistent with the finding that *Atrx* knockout mouse embryos die before embryonic day 9.5 in development (Garrick et al., 2006). Furthermore, Atrx has been implicated into loading of the chromatin insulator CTCF and the methyl-binding protein MeCP2 at imprinted genes, suggesting that misregulation of imprinted genes could be a cause for the mental retardation phenotype observed in ATRX-syndrome (Kernohan et al., 2010).

Atrx and *Daxx* are commonly mutated in pancreatic neuroendocrine tumors, glioblastoma and pediatric neuroendocrine tumors, suggesting that both genes act as strong tumor suppressors (Cheung et al., 2012; Heaphy et al., 2011; Jiao et al., 2012; Jiao et al., 2011; Molenaar et al., 2012; Schwartzentruber et al., 2012). Interestingly a lot of these cancers activate an alternative lengthening of telomeres (ALT) pathway to ensure telomere elongation in the absence of telomerase expression and ALT activation correlates with mutations in *Atrx* and *Daxx* (Lovejoy et al., 2012). Associated mutations in the genes encoding histone H3.3 in pediatric glioblastoma have also been suggested to be involved in Atrx- and Daxx-dependent processes in these tumors (Elsasser et al., 2011). Recently it has been shown that these H3.3 mutations directly affect gene expression by globally inhibiting

histone H3K27-methylation, suggesting that H3.3 mutations might not be directly implicated into the ALT phenotype and Atrx-dependent processes (Lewis et al., 2013).

Despite the indications that Atrx might be a tumor-suppressor, ATRX-syndrome patients do not have an increased risk for malignancies. This suggests that either residual Atrx activity is sufficient for its tumor suppressor function or *Atrx* loss is not a driver for tumor development. *Atrx* loss has also been correlated with genomic instability and chromatid cohesion defects, which might also contribute to tumor development (De La Fuente et al., 2004; Ritchie et al., 2008). However, other groups have not observed major genetic instability in *Atrx* depleted cells (Garrick et al., 2006; Wong et al., 2010).

In summary, Atrx is an essential chromatin regulator in development and tumor suppression with still enigmatic functions. In this thesis I found Atrx to be involved in mammalian retrotransposon repression and established a quantitative assay to measure Atrx function during heterochromatic gene silencing.

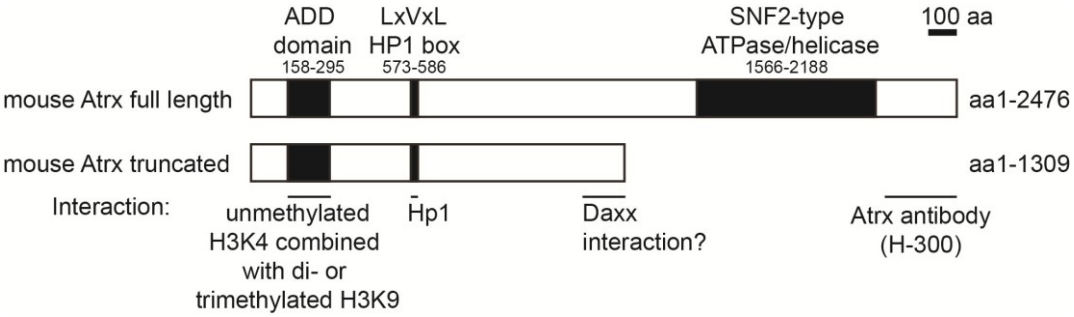


Figure 1.5 Atrx protein domain structure
The scheme shows the domain architecture of the Atrx protein. The two most prominent splice isoforms are shown. The positions of known interaction surfaces are indicated. Numbers below protein domain names indicate the amino acid position (modified from (Garrick et al., 2006; Ratnakumar and Bernstein, 2013)).

1.3.4. The role of Daxx in transcriptional repression and histone H3.3 incorporation

Apart from identifying Atrx as an important factor in retrotransposon silencing I also found that the Atrx-associated protein Daxx is involved in heterochromatic repression of retrotransposons.

Daxx was originally discovered as a protein that connects Fas receptor signaling at the plasma membrane to activation of the JNK signaling pathway (Yang et al., 1997). However, these initial findings were mainly based on yeast two-hybrid experiments and further reports showed that Daxx has a nuclear localization and Daxx-dependent activation of JNK signaling might be due to transcriptional effects (Lindsay et al., 2009; Torii et al., 1999). In the nucleus, Daxx is found finely dispersed in the nucleoplasm, but accumulates at specific subnuclear domains.

The most prominent nuclear structures where Daxx accumulates are Pml nuclear bodies to which Daxx is recruited by binding sumoylated Pml protein with its C-terminal Sumo-interaction motif (Ishov et al., 1999; Li et al., 2000a; Lin et al., 2003; Zhong et al., 2000) (Figure 1.6A). Pml nuclear bodies (also called nuclear dots) are dynamic matrix-associated structures consisting of large protein complexes that are either naturally present in the cell or can form under a variety of cellular stresses (Gamell et al., 2014). In addition, Daxx can itself recruit Atrx to Pml nuclear bodies or colocalize with Atrx at pericentric heterochromatin (Ishov et al., 2004; Tang et al., 2004; Xue et al., 2003). Daxx is also recruited to the nucleolus by Msp58 (Lin and Shih, 2002). Besides its localization to these distinct nuclear structures, Daxx is also extensively recruited to the DNA by many transcription factors to repress transcription, where it is mostly recruited in a Sumo-dependent manner (Chang et al., 2005; Croxton et al., 2006; Hollenbach et al., 1999; Kim et al., 2003; Kuo et al., 2005; Lehembre et al., 2001; Li et al., 2000b; Lin et al., 2004; Lin et al., 2006; Lin et al., 2003; Obradovic et al., 2004; Park et al., 2007; Shih et al., 2007; Zhao et al., 2004).

Interestingly, the activity of Daxx as a transcriptional corepressor is dependent on the concentration of the Pml protein. Upon overexpression of Pml, Daxx is predominantly recruited to Pml nuclear bodies by binding sumoylated Pml with its C-terminal SIM. This reduces the recruitment of Daxx to sumoylated transcription factors and consequently Daxx-mediated transcriptional repression (Lehembre et al., 2001; Lin et al., 2006; Lin et al., 2003). When cells are depleted of Pml, Daxx is strongly enriched at heterochromatic regions together with Atrx (Ishov et al., 1999). Importantly, also the recruitment of Daxx to heterochromatin is dependent on its Sumo-interaction motif (Lin et al., 2006), indicating that Sumo-SIM interactions are the predominant ways of recruiting Daxx to its genomic targets. Consequently, recruitment of Daxx to sumoylated Pml can globally regulate transcriptional repression by Daxx or its association to sumoylated heterochromatin (Ishov et al., 1999; Lin et al., 2006; Shih et al., 2007).

How Daxx mediates transcriptional repression is not precisely understood. In addition to its interaction partners mentioned above, Daxx has been found to associate with HDAC1, HDAC2, the Dnmt1-associated protein (DNAP1) and the Dek oncoprotein (Hollenbach et al., 2002; Li et al., 2000a; Muromoto et al., 2004). It can be hypothesized, that Daxx-mediated transcriptional repression could be due to the recruitment of additional heterochromatic proteins that are associated with Daxx. A prominent candidate for this could be Setdb1, which also colocalizes with Daxx at PML-bodies and pericentric heterochromatin and can be recruited by sumoylated transcription factors (Cho et al., 2011; Stielow et al., 2008).

In addition to repressing endogenous gene expression, Daxx, Pml and Atrx have been found to repress viral replication of herpesviruses, adenoviruses, papillomaviruses and retroviruses

(Schreiner and Wodrich, 2013). How the Pml, Daxx and Atrx pathway mechanistically restricts viruses replication is hardly understood, but probably involves transcriptional repression of the viral genome. Interestingly, virion proteins that inactivate Daxx do not share any structural similarity, but can efficiently bind endogenous Daxx, suggesting that also these protein interactions might be facilitated by posttranslational modifications like sumoylation (Schreiner and Wodrich, 2013).

Apart from its function as a transcriptional regulator, Daxx is a histone H3.3 chaperone that incorporates H3.3/H4 tetramers at heterochromatic regions (Drane et al., 2010; Elsasser et al., 2012; Goldberg et al., 2010; Lewis et al., 2010; Liu et al., 2012). Apart from centromeric histone H3, mammalian cells express three highly similar core histone proteins in somatic cells, H3.1, H3.2 and H3.3 (Hake and Allis, 2006). Histone H3.1 and H3.2 only differ in one amino acid and are incorporated into chromatin in a replication-coupled manner (Burgess and Zhang, 2013). In contrast, H3.3 differs in four to five amino acids from the other variants and is incorporated into chromatin in a replication-independent manner (Ahmad and Henikoff, 2002). Despite only minor differences in their amino acid composition, H3.3 is incorporated by a distinct set of histone chaperones, which leads to an enrichment of H3.3 at specific region in the genome (Burgess and Zhang, 2013; Goldberg et al., 2010). At active and repressed genes the histone chaperone Hira incorporates H3.3 (Adam et al., 2013; Banaszynski et al., 2013; Goldberg et al., 2010). Hira-dependent incorporation is specifically important to recruit the repressive H3K27me3 chromatin mark at some developmentally regulated promoters and to recover a functional transcriptional state upon DNA damage (Adam et al., 2013; Banaszynski et al., 2013). However, Hira does not catalyze H3.3 incorporation at heterochromatic regions and at some transcription factor binding sites (Goldberg et al., 2010). Instead, H3.3 incorporation at telomeres and pericentric regions has been shown to rely on Daxx (Drane et al., 2010; Goldberg et al., 2010; Lewis et al., 2010), and *in vitro* assays confirm that Daxx binds H3.3/H4 tetramers with great specificity (Elsasser et al., 2012; Liu et al., 2012). Yet, the function of H3.3 incorporation at heterochromatic regions is unclear. It was suggested that H3.3 incorporation results in transcription of heterochromatic repeats (Drane et al., 2010). Interestingly, expression of heterochromatic repeats is important for heterochromatin formation in yeast (Grewal and Jia, 2007), but H3.3 has so far not been implicated into heterochromatin formation. Another possibility is that H3.3/Daxx complexes form a nuclear reservoir of H3.3, and H3.3 incorporation into heterochromatin is a secondary effect of the accumulation of H3.3/Daxx complexes at these regions. In addition, the *Drosophila* homolog of the Daxx-associated protein Dek also shows H3.3 chaperon activity (Sawatsubashi et al., 2010), suggesting that the mechanism of Daxx-mediated H3.3 deposition could be even more complex. Upon Atrx or Daxx depletion, H3.3 incorporation at telomeres is reduced, leading to a slight upregulation of telomeric

TERRA transcripts (Goldberg et al., 2010). It is unclear, if H3.3 incorporation has an important function in telomere maintenance.

Consistent with its important cellular functions *Daxx* knockout mice die at embryonic day 9.5 of development (Michaelson et al., 1999). Extensive apoptosis and cell death precedes the death of *Daxx*-deficient embryos suggesting that either Daxx suppresses apoptotic pathways *in vivo* or Daxx deficiency leads to the accumulation of cellular defects (Michaelson et al., 1999). Interestingly, the time point of embryonic lethality coincides with that of *Atrx*-deficient embryos implying that the observed association of Atrx and Daxx *in vitro* might also reflect a functional cooperativity *in vivo* (Garrick et al., 2006). In addition, to its function in normal development, Daxx seems to be an important tumor suppressor gene (already reviewed in the introduction chapter about Atrx).

In summary, the histone H3.3 chaperone Daxx is an essential protein in development that is targeted to most of its genomic targets by sumoylation where it represses transcription. However, the relationship between its catalytic activity to incorporate H3.3 and its ability to repress transcription are not understood.

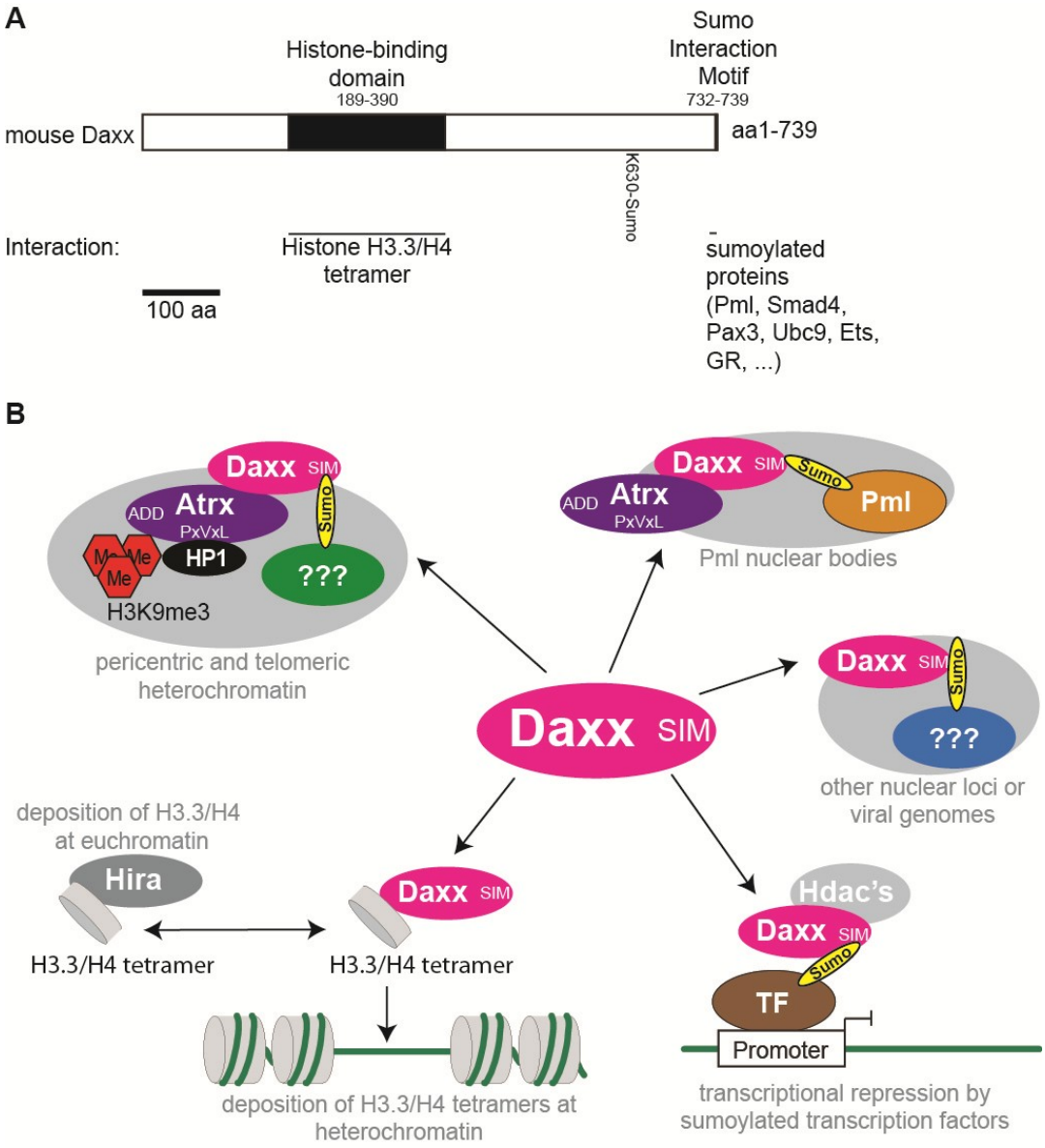


Figure 1.6 Daxx protein functions and interactions
A) The scheme shows the domain architecture of the Daxx protein. The positions of known interaction surfaces are indicated. Numbers below protein domain names indicate the amino acid. The protein is shown with the N-terminus on the left. Daxx is also sumoylated, but this is neither important for its repressive activity nor its localization (Lin et al., 2006). B) The illustration gives an overview over the most prominent nuclear localizations and functions of Daxx. Daxx is recruited to Pml nuclear bodies via binding to sumoylated Pml, which could be a reservoir of nuclear Daxx. Similarly, Daxx is recruited by sumoylated transcription factors to repress transcription. In addition, it can also be found at heterochromatin, where it colocalizes with Atrx. Please note that Daxx recruitment to heterochromatin is also dependent on its Sumo-interaction motif. However, which sumoylated protein is bound by Daxx at heterochromatin is not known. At heterochromatin Daxx has been shown to be a H3.3/H4 chaperone, while in euchromatic regions H3.3 is incorporated by the histone chaperone Hira (modified from (Shih et al., 2007)).

1.4. Aim of the thesis

Most of the heterochromatin proteins in mouse have been investigated due their orthology with heterochromatin factors identified in *Drosophila* by genetic reporter screens for position-effect-variegation (Fodor et al., 2010). In addition, genetic screens for heterochromatin proteins have also been performed in mouse in low throughput (Ashe et al., 2008; Blewitt et al., 2006; Chong et al., 2007).

Although mammalian species share a high number of orthologous genes with other model organisms, the higher complexities of the mammalian genome and mammalian gene regulation request a deeper functional understanding of heterochromatin regulation in mammals. Consequently, proteomic studies revealed important insights and characteristics of mammalian heterochromatin (Bartke et al., 2010; Nozawa et al., 2010; Vermeulen et al., 2010). However, proteomic studies provide no information about the functionality of these heterochromatin players and their importance in heterochromatic gene silencing. Moreover, transient and short-term interactions of proteins often remain undetected by proteomic analysis.

Therefore, we wanted to establish a genetic screening platform for heterochromatic gene silencing in mammals that is fast, easy and reliable.

The first aim of my thesis was the establishment of a robust cellular heterochromatin reporter system. I first focused on genetic regions that are modified by heterochromatic marks and that are regulated by heterochromatin proteins.

Secondly, having established such a screening system, the additional aim was to identify molecular mechanisms and genes involved in mammalian heterochromatin formation using unbiased RNAi experiments.

After identification of primary hits from the screen, these hits should be validated by genetic and biochemical experiments to learn more about the underlying mechanisms.

2. Results

2.1. Investigation of a heterochromatic reporter gene locus. The pseudoautosomal exons of the *midline1* gene are derepressed in *Suv39h* double knockout and *Suv4-20h* double knockout mouse embryonic stem cells

Since the aim of this study was to establish a heterochromatic screening system for the functional identification of heterochromatic proteins, the initial idea was to use endogenous gene loci that are regulated by heterochromatic pathways as a heterochromatin reporter locus.

A large expression dataset of different mouse ES cell lines during embryoid body differentiation was previously generated in our lab by Dr. Matthias Hahn. This dataset was analyzed to find derepressed genes in cell lines mutated for the histone methyltransferases Suv39h1 and Suv39h2 or Suv4-20h1 and Suv4-20h2, which act together in a consecutive pathway in heterochromatin formation (introduction chapter 1.1.2) (Schotta et al., 2004; Schotta et al., 2008). The only gene that was consistently derepressed through all stages of embryoid body differentiation in *Suv4-20h* double knockout and *Suv39h* double knockout mES cells was *midline1* (data not shown).

Semi-quantitative RT-PCR was performed from RNA isolated from mES cells deficient for *Suv39h* or *Suv4-20h* enzymes to validate that *midline1* is deregulated in these cells (Figure 2.1A). Strong amplification of *midline1* cDNA only in mES cells deficient for *Suv39h* or *Suv4-20h* suggested that *midline1* transcription is normally repressed by Suv39h and Suv4-20h enzymes (Figure 2.1A).

The *midline1* gene lies on the X-chromosome but flanks the pseudoautosomal boundary in mice. This means the 5' end of the *midline1* gene lies exclusively on the X-chromosome (exons 1-4), while its 3' exons lie on the pseudoautosomal region (PAR) which is present on both, the X-chromosome and the Y- chromosome (Figure 2.1B). Published ChIP-sequencing profiles of the heterochromatic marks H4K20me3 and H3K9me3 indicate an enrichment of these histone marks on the pseudoautosomal part of the gene (Mikkelsen et al., 2007) (Figure 2.1B). Because these repressive histone marks are only present on the pseudoautosomal part of the *midline1* gene, I performed exon-specific RT-qPCR to investigate whether the observed derepression of the *midline1* gene in *Suv39h* and *Suv4-20h* mutant cells originates from the pseudoautosomal or the X-chromosome-specific part of the gene. Interestingly, the derepression of the *midline1* gene is exclusively limited to its pseudoautosomal exons (Figure 2.1B). In summary, this data suggests that *Suv39h* and *Suv4-20h* regulate specific *midline1* transcripts from the pseudoautosomal region in mice.

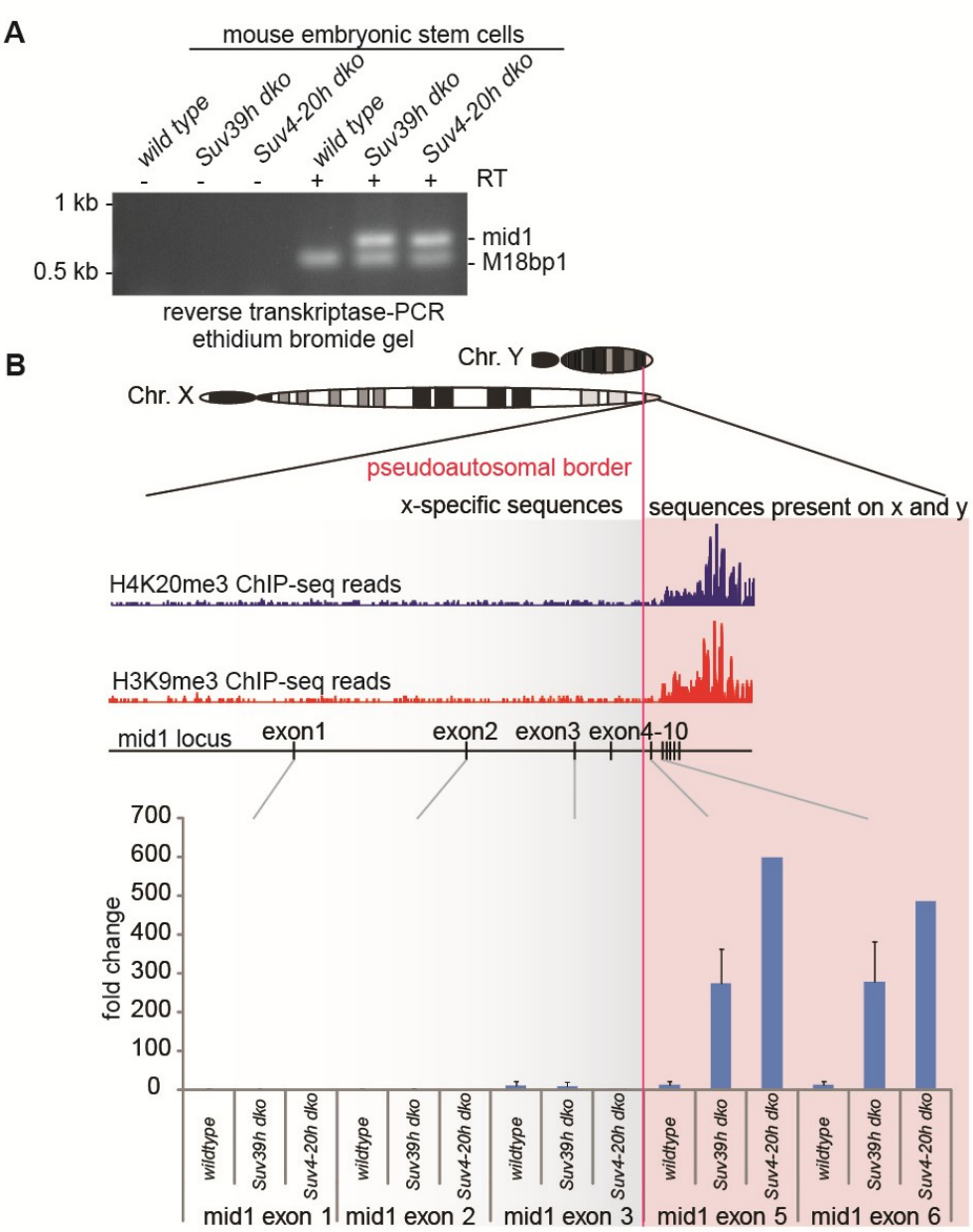


Figure 2.1 The pseudoautosomal exons of the *midline1* gene are derepressed in *Suv39h* and *Suv4-20h* double knockout mES cells
A) The *midline1* transcript is ectopically expressed in *Suv39h* and *Suv4-20h* dko mES cells as shown by semi-quantitative reverse transcriptase PCR. The PCR amplification of the *midline1* transcript is dependent on the addition of reverse transcriptase during reverse transcription. The *M18bp1* transcript serves as an internal positive control. B) The *midline1* gene flanks the pseudoautosomal boundary in mice. The 5' end of the midline1 gene lies exclusively on the X-chromosome (exons 1-4), while its 3' exons are situated on the pseudoautosomal region which is present on both, the X-chromosome and the Y-chromosome. Publicly available ChIP profiles point to an enrichment of H3K9me3 and H4K20me3 on the pseudoautosomal region (Mikkelsen et al., 2007). Exon-specific RT-qPCR shows that the derepression of *midline1* in *Suv39h* and *Suv4-20h* dko mES cells is limited to the pseudoautosomal exons of the gene. Number of different cell lines analyzed: wild type n=2, *Suv39h* dko n=2, *Suv4-20h* dko n=1. Error bars indicate the standard deviation.

2.2. The pseudoautosomal *midline1* transcript potentially originates from the Y-chromosome, is orientated in sense direction and originates from a complex genetic environment

Because the derepression of the *midline1* gene in *Suv39h* and *Suv4-20h* knockout cells is limited to its pseudoautosomal part, it was unclear, whether these *midline1* transcripts originated from the X- or the Y-chromosome. Northern blot analysis of female and male mES cells revealed that the *midline1* gene is specifically expressed in male cells deficient for *Suv39h* or *Suv4-20h* (Figure 2.2A). This suggests that these transcripts arise from the Y-chromosome. In addition, the derepressed *midline1* transcript seems to be of a specific size on the Northern blot, indicating that it originates from a specific transcriptional start site. To check for any directionality of the transcript, direction-dependent RT-PCR was performed. Because amplification of the derepressed *midline1* transcript could only be detected when an antisense primer was used for reverse transcription, the derepressed *midline1* transcript is originated in sense direction (Figure 2.2B). This represents the same directionality as the full-length X-chromosomal *midline1* transcript.

Having identified a specific *midline1* transcript that is probably arising from the Y-chromosome, I wanted to make use of it as a heterochromatic reporter locus for further genetic screening approaches. However, there is still no reliable source of sequence information of this part of the mouse Y-chromosome. In addition, several papers suggested that the mouse pseudoautosomal region is short, highly polymorphic and a source of many *de novo* mutations (Kipling et al., 1996a; Kipling et al., 1996b; Palmer et al., 1997; Perry et al., 2001). Since using the Y-chromosomal *midline1* locus as a reporter locus would involve gene-targeting in that particular region in mES cells, I tested the heterogeneity and copy number of the deregulated *midline1* exons in different cell types (Figure 2.2C). In contrast to the copy number of autosomal gene loci, the copy number of the derepressed *midline1* exons was highly polymorphic and different in every cell line. In summary, the pseudoautosomal *midline1* transcript might originate from a highly polymorphic locus on the Y-chromosome, a region which is difficult to modify by standard gene-targeting approaches. In addition, there is only poor sequence information available. These findings encouraged me to develop an alternative genetic screening approach.

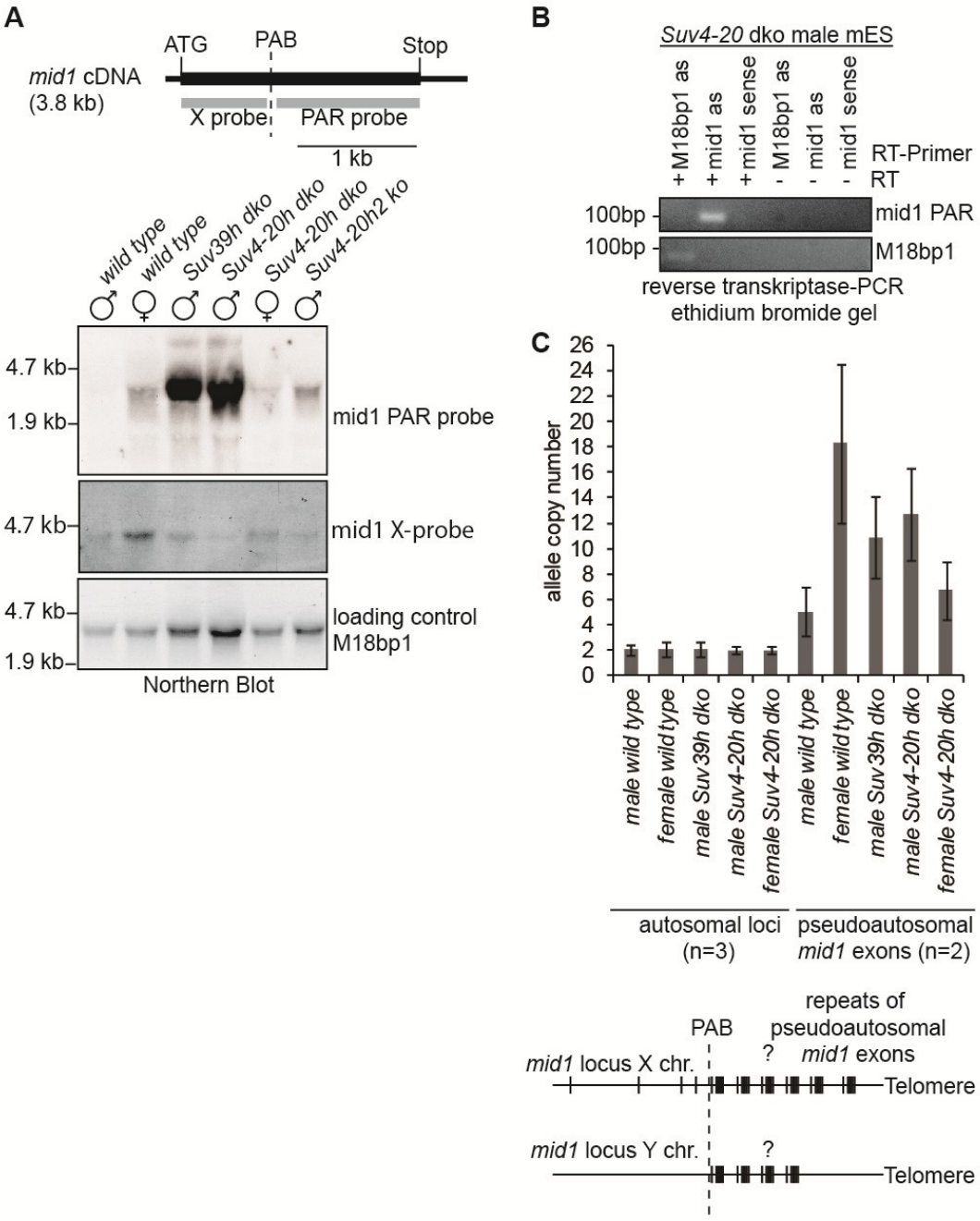


Figure 2.2 The pseudoautosomal *midline1* transcript potentially arises from the Y-chromosome, is orientated in sense direction and originates from a complex genetic surrounding.
A) Male *Suv39h* and *Suv4-20h* dko, but not female *Suv4-20* dko cells show a robust upregulation of pseudoautosomal *midline1* transcripts in northern blot analysis. The position of the northern probes upstream or downstream of the pseudoautosomal boundary is indicated in the cartoon. M18bp1 and the X-chromosomal probes serve as loading controls. B) Only the antisense Primer used for reverse transcription gives rise to an RT-PCR product, revealing that the derepressed *midline1* transcript is oriented in sense direction. The *M18bp1* transcript serves as a positive control. C) Genomic DNA of different mutant cell lines was isolated and allele copy number was analyzed using qPCR. The number of autosomal alleles is identical between different mutant cell lines, while the number of pseudoautosomal *midline1* exons varies among different cell lines. The barplot represents two biological replicates. The average number of autosomal loci was determined by analyzing three independent autosomal loci. The number of pseudoautosomal *midline1* exons was determined using two independent primer pairs. Error bars indicate the standard deviation.

2.3. A candidate luciferase reporter screen for heterochromatic gene regulators identifies Setdb1, Trim28, Prdm6 and Hp1 proteins as transcriptional repressors

Apart from *midline1* no other strongly deregulated gene could be identified in our expression dataset generated from *Suv39h* and *Suv4-20h* knockout mES cells. Because depletion of *Suv39h* and *Suv4-20h* did not result in consistent transcriptional misregulations of specific genes, I wondered whether other heterochromatin-associated proteins directly influence gene expression. Gal4-fusion proteins of various heterochromatin-associated proteins were cloned and tested on their potential to silence a *luciferase* reporter gene. A candidate list of heterochromatin-associated proteins was taken from a mass-spectrometry-based screen for proteins associated with H3K9me3-peptides (Dambacher, 2013). In addition, other proteins were chosen, which were implicated into trimethylation of H3K9 (Falandry et al., 2010; Schultz et al., 2002).

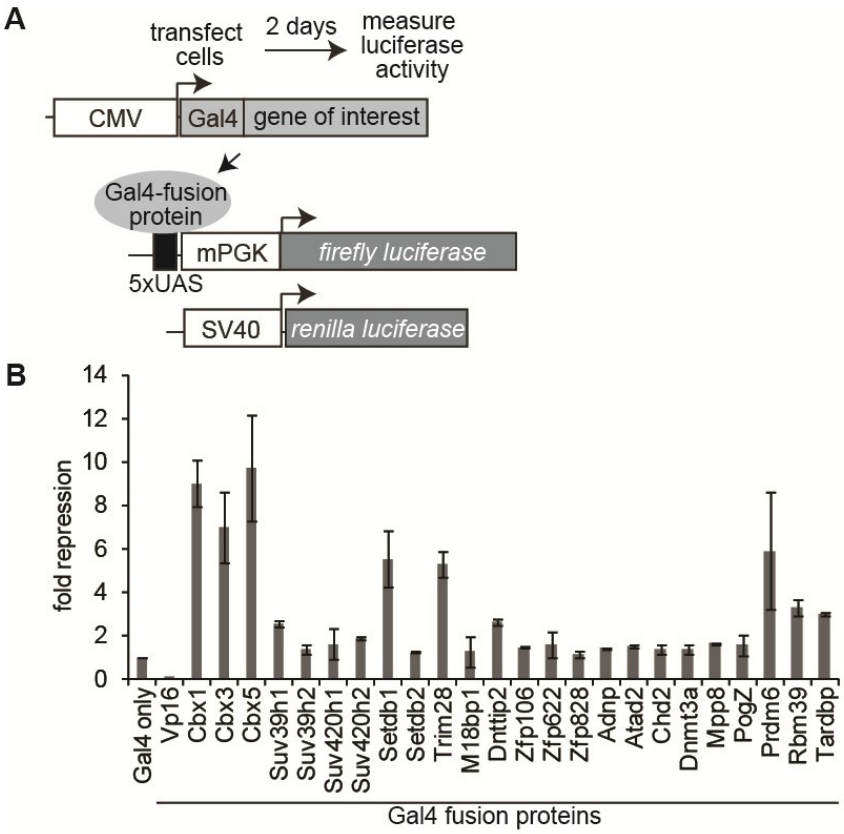


Figure 2.3 Setdb1, Trim28, Prdm6 and Hp1 Gal4-fusion proteins silence a *luciferase* reporter gene
A) A candidate screen checking the repression capabilities of H3K9me3-associated chromatin proteins. Candidate genes were fused to the Gal4-DNA binding domain and co-transfected with a constitutively active *firefly luciferase* vector containing five UAS binding sites for the Gal4 protein. A *renilla luciferase* control vector was co-transfected to normalize for unspecific effects on transcription and gene expression. B) Gal4-fusions of Hp1 proteins (Cbx1, Cbx3 and Cbx5), Setdb1, Trim28 and Prdm6 repress the *firefly luciferase* reporter vector. A representative experiment performed in triplicates is shown. The transactivator Vp16 is used as a negative control. Error bars represent the standard deviation.

A mouse phosphoglycerate kinase (mPGK) promoter was cloned between five Gal4-binding sites (5xUAS) and the *firefly luciferase* gene to ensure a high basic level of transcription of the *firefly luciferase* reporter gene (Figure 2.3A). A *renilla luciferase* reference plasmid was used as an internal control for normalization of global and unspecific transcriptional effects. Setdb1, Trim28, Prdm6 and Hp1 proteins (Cbx1, Cbx3 and Cbx5) were identified to repress the mouse PGK promoter (Figure 2.3B). In contrast, no strong repression was observed by targeting Suv39h and Suv4-20h enzymes to the *firefly luciferase* gene promoter.

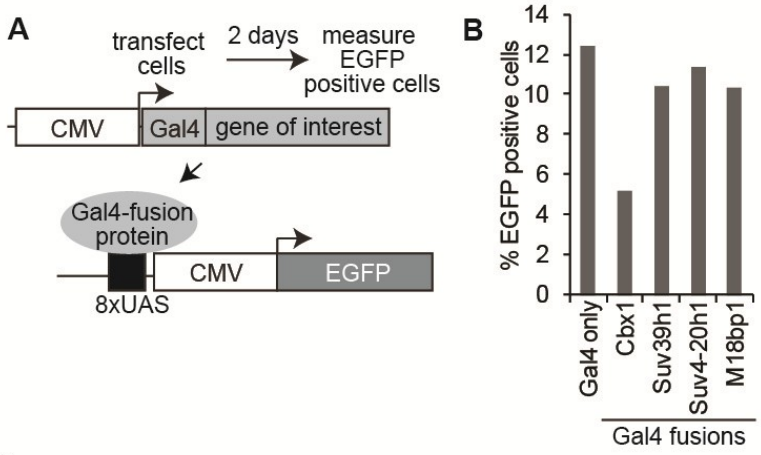
In summary, proteins involved in heterochromatic gene silencing can be identified by using reporter gene assays and Setdb1, Trim28, Hp1 proteins and Prdm6 are sufficient to induce gene silencing. These proteins are suitable positive controls for establishing a novel screening platform for heterochromatic gene silencing.

2.4. EGFP is a suitable reporter gene for heterochromatic gene silencing

Heterochromatin repression assays using *luciferase* as a reporter gene can detect transrepressive effects (Figure 2.3). However, measurement of luciferase activity always occurs after lysis of a cell pool and the reporter expression of individual cells cannot be detected. Because I wanted to combine a heterochromatic reporter assay with pooled shRNA-screening, a reporter assay was required that allows the isolation of functional shRNA sequences from single cells that bypass heterochromatic silencing. Thus, a reporter gene was needed that can be analyzed in a single cell-based manner. Enhanced green fluorescent protein (EGFP) was cloned behind a strong CMV promoter and Gal4-binding sites to test, whether EGFP-detection using fluorescence-activated cell sorting (FACS) could be used to detect reporter gene repression (Figure 2.4A). Gal4-fusion proteins of Cbx1, Suv39h1, Suv4-20h1 and M18bp1 were transiently co-transfected with the 8xUAS-CMV-EGFP reporter vector into HeLa cells and EGFP expression was monitored after two days. Similar to the data obtained by the luciferase assay in Figure 2.3 only Cbx1 could strongly reduce reporter gene expression (Figure 2.4B). Please note that an internal control gene similar to the *renilla luciferase* plasmid could not be used, due to a lack of an excitation-laser for other fluorescent proteins in FACS. However, since the EGFP-based silencing assay phenocopied the outcome of the internally controlled luciferase assay, I concluded that EGFP expression is usable as a read-out for single-cell-based repression assays.

Because transient transfection experiments of the EGFP reporter results in a high variability of EGFP expression, it is difficult to evaluate whether a single cell that expresses EGFP only on a low level has just taken up a low amount of plasmid or whether reporter gene silencing has occurred. Therefore, a very homogenous EGFP reporter expression was required. In addition the EGFP reporter should be incorporated into the genome of the cell, to ensure that

also chromatin-dependent events can be detected, when combining the assay with shRNA-based screening. MEFs and HeLa cells were transfected with the 8xUAS-CMV-EGFP reporter vectors to generate stable cell lines and single cell derived clones were generated using a FACS sorter. Even though transfection of the 8xUAS-CMV-EGFP reporter resulted in the generation of some EGFP-positive cell lines, a homogenous EGFP expression could not be obtained with this construct (Figure 2.4C). Because of the inhomogeneous expression of the CMV reporter constructs the CMVpromoter of the reporter vector was replaced with the *mPGK* promoter. This modification resulted in several clonal reporter cell lines that homogenously expressed EGFP. However, when transfecting these cells with Gal4-Cbx1 or Gal4-Vp16 no repression or further activation of EGFP expression could be seen (Figure 2.4C). In summary, EGFP is a suitable reporter gene for heterochromatic gene silencing, but the generation of reporter cell lines that are responsive to heterochromatic silencing was not successful.



C

plasmid for stable transfection	clonal cell lines created	EGFP expression detectable	uniform EGFP expression	EGFP repressed by Gal4-Cbx1
8xUAS-CMV-EGFP	31	5/31	0/5	n/a
8xUAS-PGK-EGFP	18	18/18	16/18	0/6

Figure 2.4 EGFP is a suitable reporter gene for investigating heterochromatic repression activities
A) The principle of the luciferase experiment of Figure 2.3 was translated into a single-cell-based assay by replacing the luciferase cassette with EGFP. B) Proof-of-principle experiment showing that Gal4-Cbx1 represses the 8xUAS-CMV-EGFP transgene. HeLa cells were transiently transfected with different Gal4-fusion proteins and the 8xUAS-CMV-EGFP reporter plasmid. Two days after transfection the number of EGFP-positive cells was measured using FACS. C) Generation and testing of stable EGFP reporter cell lines. Two independent EGFP reporter plasmids carrying either the CMV or the PGK promoter were transfected into MEF and HeLa cells and selected using Zeocin™ as a selection marker. After several days post transfection individual EGFP-positive cells were sorted into a 96-Well plate giving rise to clonal cell lines. The 8xUAS-CMV-EGFP transgene is silenced after several days and EGFP expression is lost within these clones. In contrast the PGK-EGFP transgene mainly results in clonal cell lines with homogenous EGFP expression. However, transfection of these cells with Gal4-Cbx1 leads to no repression of EGFP signals.

2.5. Generation and optimization of a second generation lentiviral packaging system pseudotyped with ecotropic envelope protein of M-MLV

Incorporating reporter cassettes into the cellular genome and into a chromatinized genomic environment is difficult and time consuming using standard plasmid transfection technologies (Figure 2.4C). For generating a reliable method to deliver reporter genes into cells, lentiviral transduction was optimized for using it under biosafety level 1 condition. This was necessary, because modifications of a lentiviral vector used under biosafety level 2 would in each case require a notification to governmental authorities. This organizational bottleneck was avoided by restricting the lentiviral system to exclusively transduce mouse and rat cells resulting in a biosafety level 1 transduction system. Restriction of lentivirus to rodent cells has been performed earlier (Schambach et al., 2006). However, a more reproducible and more efficient second generation lentiviral packaging system has not been established.

For restricting lentiviral transduction to mouse or rat cells, the commonly used VSVg envelope protein was replaced by the ecotropic envelope protein of the Moloney-Murine Leukemia Virus (M-MLV) (Figure 2.5A). This envelope protein uses the mouse-specific cationic amino acid transporter Slc7a1 (mCAT1) as an entry receptor (Albritton et al., 1989). Second generation lentiviral packaging works via co-transfection of a plasmid encoding the required proteins of HIV-1 (gag, pol, tet, rev), a plasmid encoding an envelope protein determining viral tropism and a lentiviral transfer vector into HEK293T cells (Naldini et al., 1996; Zufferey et al., 1997) (Figure 2.5A). The lentiviral transfer vector will give rise to an RNA, which after transduction and reverse transcription inside the host cell, will integrate at active sites of the genome (Schroder et al., 2002). After testing different plasmid backbones for the ecotropic envelope plasmid, the highest lentiviral titers were obtained when the plasmid carrying the ecotropic envelope gene was lacking an SV-40 origin of replication and contained a CMV promoter including an intron of the *β-globin* gene (Figure 2.5A and data not shown).

After optimizing transfection efficiency and transfection method (Figure 2.5B), the ratio of the three packaging vectors used for co-transfection of HEK293T cells was optimized with an EGFP containing transfer vector using MEF cells (Figure 2.5C). Optimal titers were achieved by using equal amounts of packaging vectors (Figure 2.5C). Next, transduction efficiencies of differently treated viral particles were determined. While storage of the viral supernatants at 4°C or freezing hardy influences the virus titer, applying g-forces of 1000 x g during transduction of target cells increased the number of functional transduction particles (Figure 2.5D). Careful optimization of ecotropic pseudotyped lentivirus provided a reliable tool for using reporter vectors that stably and quantitatively integrate into the genome of mouse cells.

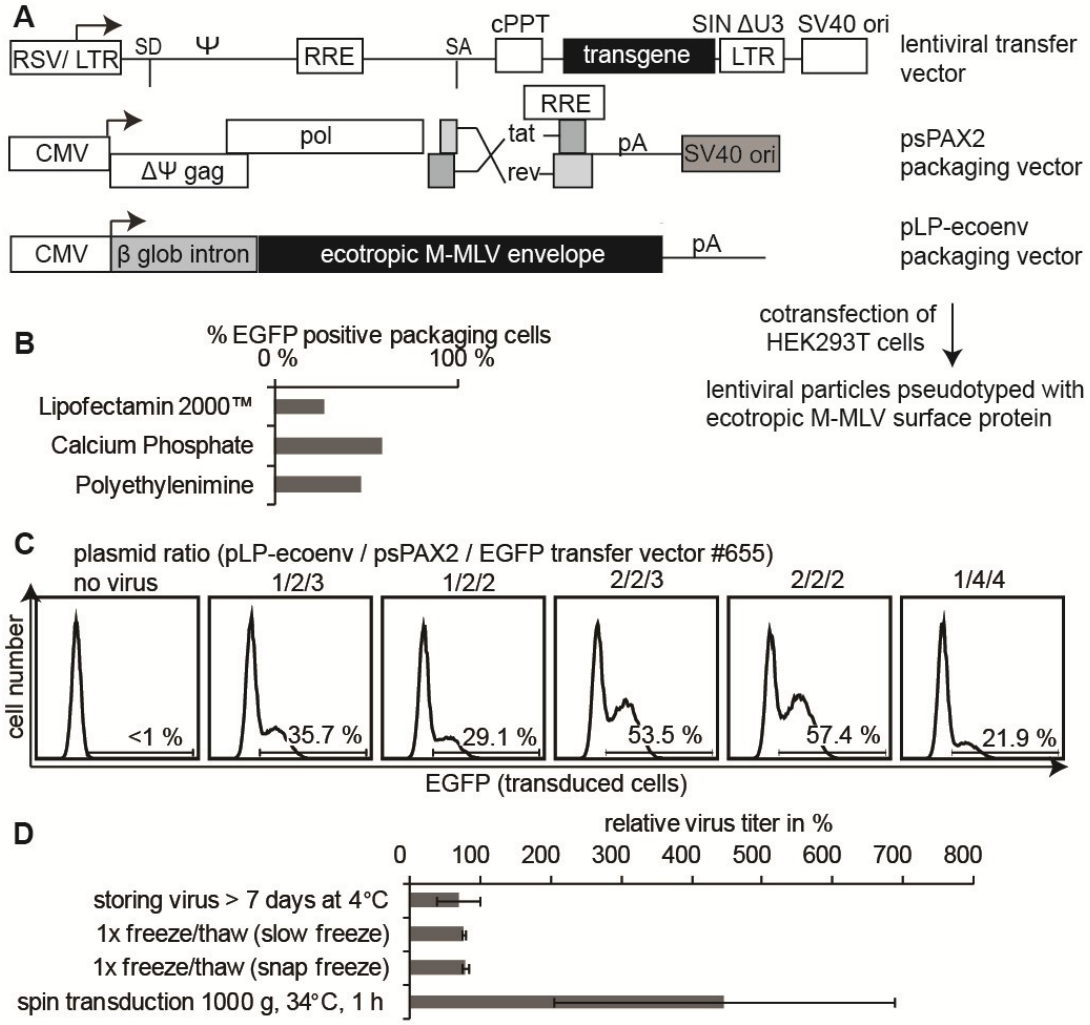


Figure 2.5 Generation and optimization of a second generation lentiviral packaging system pseudotyped with ecotropic envelope protein of M-MLV
A) The packaging system is based on a co-transfection of three plasmids into HEK293T cells that will produce viral transduction particles. The packaging vector provides the four required proteins of the HIV-1 virus (gag,pol,tet,rev) and was described earlier (Zufferey et al., 1997). These four proteins work with most commercially available lentiviral transfer vectors. The use of the ecotropic M-MLV envelope protein replaces the traditionally used VSVg protein and leads to the generation of ecotropically pseudotyped virus, which can only transduce mouse or rat cells. B) Optimization of transfection efficiency in HEK293T cells. HEK293T cells were transfected with an EGFP transfer vector and the ratio of EGFP-positive cells was measured in FACS. Calcium phosphate also outperforms the other transfection methods in terms of cell viability (data not shown). C) Optimization of plasmid ratios during lentiviral packaging to increase viral titers. Different ratios of the three packaging plasmids are used for viral packaging while maintaining the total amount of plasmid DNA constant. In contrast to many VSVg-based packaging systems, the viral supernatant has the highest number of infectious particles when equal ratios of the plasmids are used. D) Differently treated viral supernatants packaged from an EGFP transfer vector were titrated on MEF and T37 HeLa cells to determine effects on virus titer. Storage of the virus at 4°C and freezing can slightly impair virus titer (n=2). Spinning the cells at 1000 x g for one hour during viral transduction strongly increases the relative virus titer (n=6). Error bars indicate the standard deviation.

2.6. Transrepressor binding sites cloned into lentiviral EGFP expression vectors are tools for monitoring transcriptional silencing kinetics

Because reporter gene assays can be used for measuring the silencing activities of heterochromatin proteins (Figure 2.3) and because EGFP can be used as a reporter gene to monitor single-cell reporter expression (Figure 2.4), the next aim was to establish a silencing assay based on EGFP expression in individual cells. Because standard transfection technologies did not obtain stable EGFP reporter cell lines which can be silenced by heterochromatin proteins (Figure 2.4), a lentiviral reporter system was established. Similar lentiviral assays have previously been described for analyzing silencing activities of the M-MLV LTR promoter or the detection of repressive DNA sequences in transposable elements (Haas et al., 2003; Rowe et al., 2010). For reducing the complexity of the system, repressor binding sites of an endogenous repressor protein were directly cloned upstream of an EGFP reporter cassette inside a lentiviral vector. This “one-component” approach did not require the transduction of additional transgenic repressor proteins like Gal4-fusions constructs (Figure 2.6A).

I chose three binding sites of the mouse-specific KRAB-box protein Zfp809. A binding site of Zfp809 is present downstream of the LTR region of the M-MLV provirus and is a well described repressor binding-site in mouse cells (Barklis et al., 1986; Feuer et al., 1989; Loh et al., 1987; Petersen et al., 1991; Wolf and Goff, 2009; Yamauchi et al., 1995). In addition, Zfp809-dependent silencing has been shown to be dependent on H3K9me3, Trim28 and Hp1 proteins, suggesting that it reflects a form of heterochromatic gene silencing (Wolf et al., 2008; Wolf and Goff, 2007).

Twelve hours after transduction of the lentiviral reporter vector, weak EGFP expression can be observed inside a MEF cell population (Figure 2.6B). When the Zfp809 binding sites are present upstream of the EGFP expression cassette, EGFP expression is strongly repressed in the majority of the cells (Figure 2.6B, right panel). However, in the absence of a repressor binding site EGFP expression increased over time (Figure 2.6B, middle panel).

In summary, combining lentiviral transduction technology with a repressive silencing sequence resulted in the establishment of an EGFP-based silencing assay that allows detection of reporter silencing in single cells.

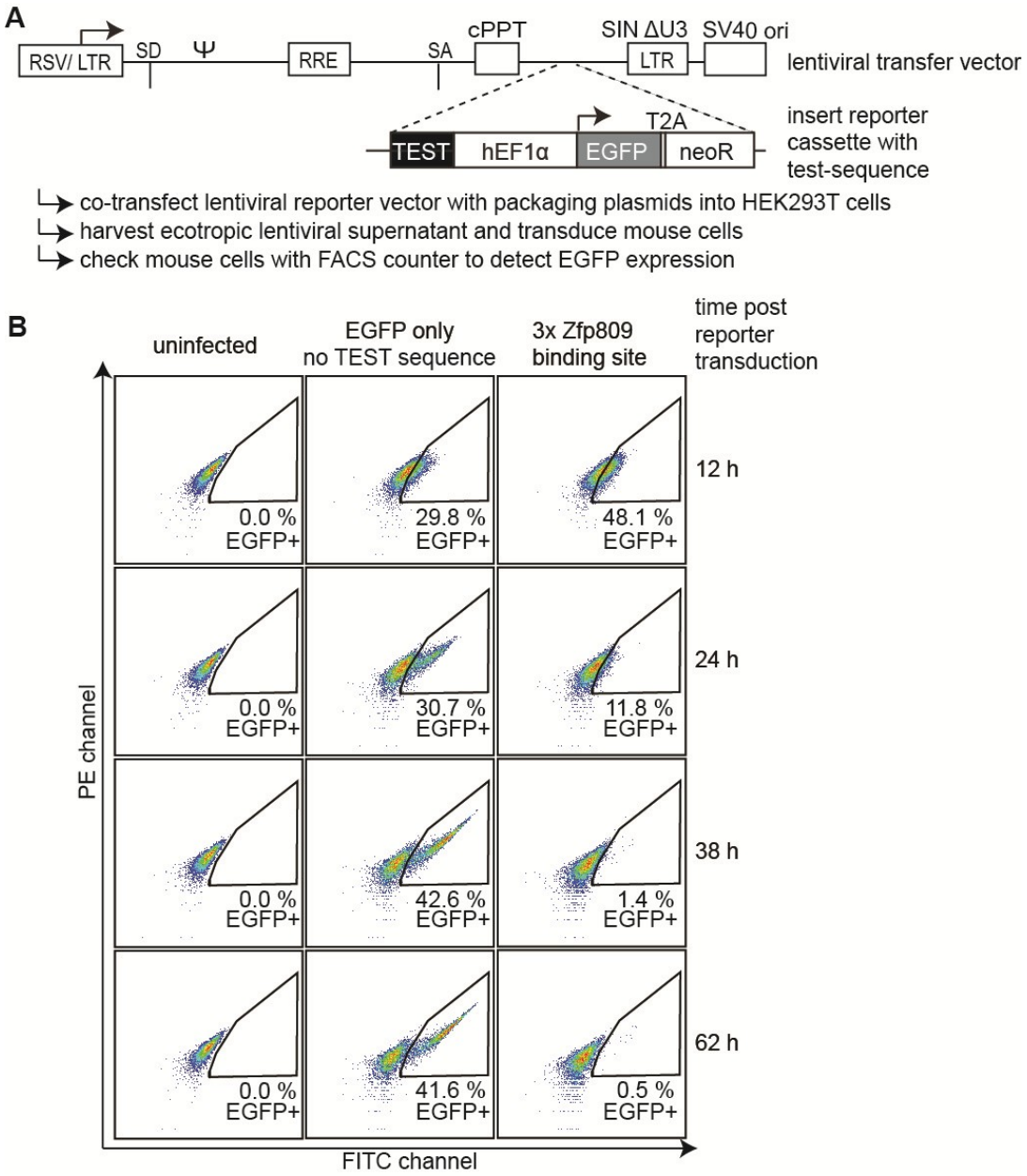


Figure 2.6 Establishment of a quantitative EGFP-based silencing assay
A) A reporter cassette consisting of an EGFP-T2A-neoR reporter gene driven by a constitutively active EF1-alpha promoter was inserted into a lentiviral vector backbone. Different test sequences can be inserted into a cloning site upstream of the reporter cassette to monitor their influence on reporter gene expression using FACS. A lentiviral system was chosen, because the reporter stably integrates into the genome of the transduced cell. B) FACS plots showing a representative experiment how a sequence element silences reporter gene expression. Three consecutive binding sites for the KRAB zinc finger protein 809 where cloned upstream of the reporter cassette. Zfp809 is a known repressor in MEF cells (Barklis et al., 1986; Feuer et al., 1989; Loh et al., 1987; Petersen et al., 1991; Wolf and Goff, 2009; Yamauchi et al., 1995), and Zfp809 binding leads to a repression of EGFP expression shortly after transduction. The FITC-channel detecting green fluorescence is plotted on the x-axis and the PE-channel detecting yellow-fluorescence is plotted on the y-axis. This arrangement allows a good discrimination between weak EGFP-fluorescence and autofluorescence.

2.7. HeLa cells stably expressing Slc7a1 (mCAT-1) are susceptible to ecotropic M-MLV pseudotyped lentivirus transduction and can control for mouse- and cell-type-specific silencing activities

Having established a lentiviral system to analyze repressive silencing sequences using EGFP reporters, a quantitative comparison of lentiviral titers carrying silencing or control sequences is needed (Figure 2.6). This is because the number of functional particles resulting in genomic integrations has to be known to quantify the strength of repression of different sequence elements inside a lentiviral reporter. This can only be achieved, when the virus titers of different constructs can be accurately measured. One way of comparing the virus titer of different reporter viruses is testing them on a cell line where no silencing occurs. Such a cell line, which does not contain repressors acting on mouse-specific repressor sequences (like Zfp809), was generated by making a human HeLa cell line susceptible to ecotropic lentiviral transduction (Figure 2.7A).

This was done by stably expressing the *Slc7a1* (*mCAT-1*) gene in HeLa cells. A similar approach was used for characterizing Slc7a1 as the ecotropic entry receptor in the first place (Albritton et al., 1989). The HeLa cell line that was generated was named T37. By titrating different vectors on T37 cells, mouse-specific silencing pathways did not influence the expression strength of the EGFP transgene in different vectors. In this way the number of functional viral particles could be quantified and compared between different reporter viruses (Figure 2.7B). This allowed a relative quantification of silencing capabilities between different reporter viruses.

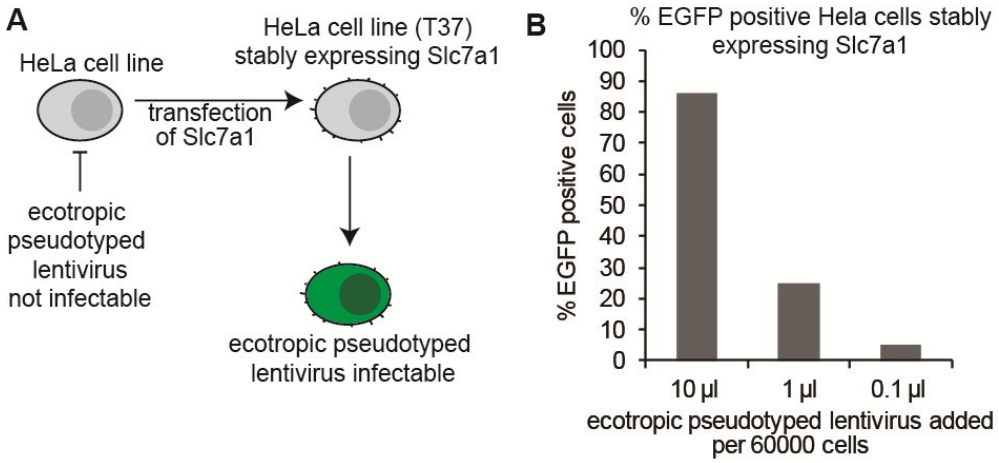


Figure 2.7 Generation of HeLa cells susceptible to transduction with ecotropic lentivirus

A) Expression of the mouse Slc7a1 protein makes HeLa cells susceptible to transduction with lentivirus pseudotyped with M-MLV ecotropic envelope protein. B) Proof of concept experiment showing that the generated Hela cell line (T37) is easily transduced with a lentivirus encoding an EGFP transgene cassette. The concentration of functional lentiviral particles can be determined by titrating different amounts of viral supernatant on the cells.

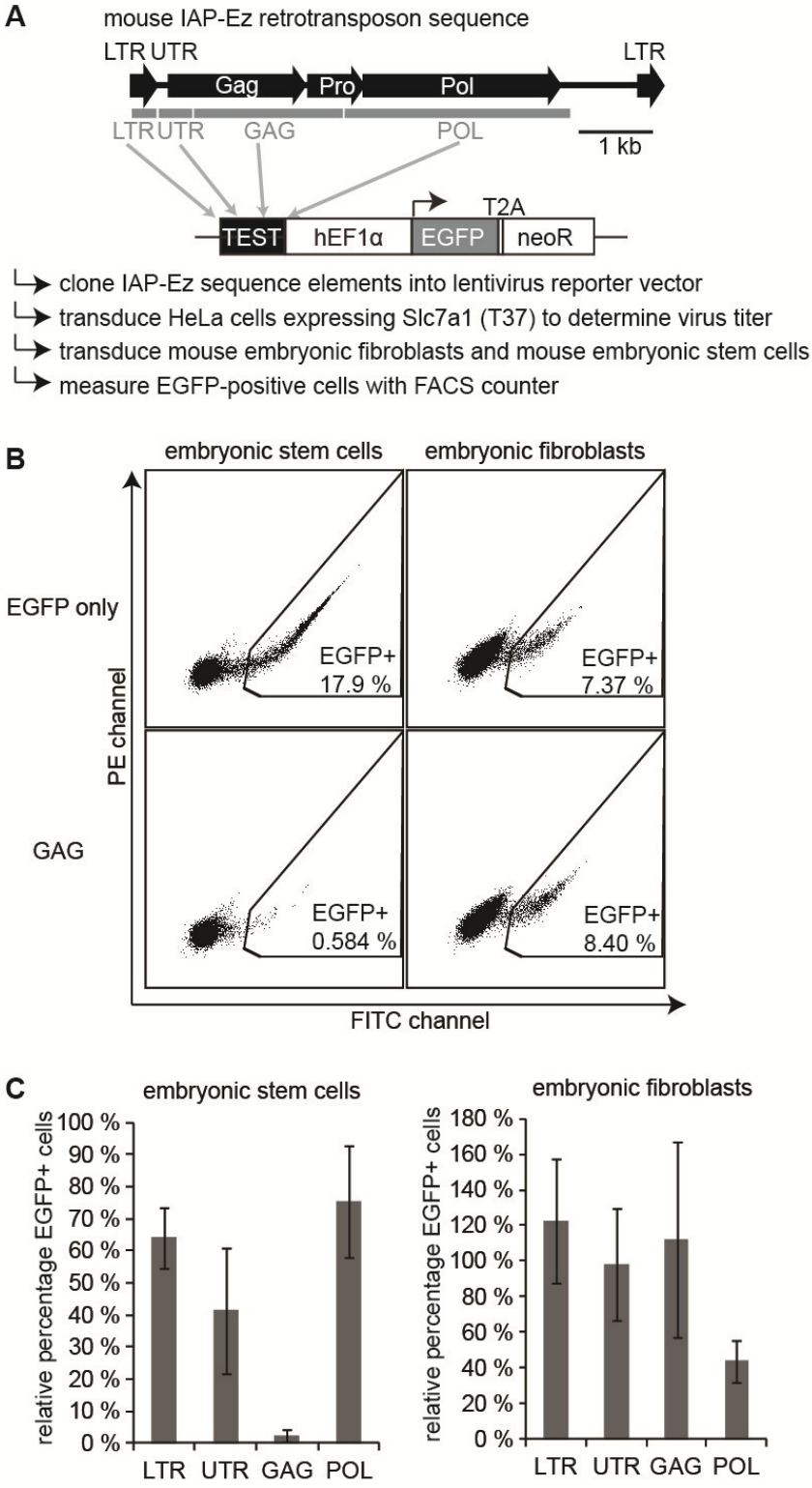
During titrations the number of EGFP-positive cells was tried to be kept below 15 %, to ensure that the number of EGFP-positive cells linearly correlates with the virus titer. In summary, the establishment of a human control cell line for normalizing ecotropic virus titers can control for mouse and cell-type specific silencing activities of DNA elements.

2.8. The GAG-region of the mouse IAP-Ez retrotransposon subclass contains a mouse embryonic stem cell specific silencing sequence

Cloning transrepressor binding sites into lentiviral EGFP reporter vectors is a powerful system not only to characterize events at known transrepressor binding sites, but also to identify new sequence elements that recruit reporter silencing. This is a way of genetic screening for the DNA-elements that recruit silencing factors. A similar assay has been performed by others to check for silencing recruiting sequences at the UTR region of IAP retrotransposons (Rowe et al., 2010). I tried to verify the data obtained on IAP retrotransposons to check whether my generated vector system can detect the presence of transrepressor binding sites. Fragments of IAP-Ez retrotransposons were cloned from mouse DNA into the lentiviral EGFP reporter vector as shown in Figure 2.8A. Lentivirus was packaged, titrated and the number of EGFP-positive MEF and mES cells was determined by FACS (Figure 2.8A). Consistent with published data, a mild transrepressive activity could be measured for the UTR-region of IAP elements, when normalized on the empty EGFP control vector and the virus titer in T37 HeLa cells (Rowe et al., 2010) (Figure 2.8C). Surprisingly, the GAG sequence element of the IAP-Ez retrotransposon leads to almost a full repression of the EGFP reporter in mouse ES cells (Figure 2.8B and C). This repression was specific for mouse ES cells, since it could not be observed in MEF cells (Figure 2.8B right panel).

In summary, the established lentiviral reporter system can detect transrepressor binding sites in cloned genomic DNA fragments. The GAG-sequence element of IAP retrotransposons contains a novel and strong silencing element that recruits reporter gene silencing in mES cells. MEF cells probably do not contain the recognition machinery to silence the GAG sequence.

Figure 2.8 The GAG region of IAP-Ez retrotransposons contains a strong heterochromatic silencing element
A) General description of the workflow for identifying heterochromatin recruiting sequence elements. Four fragments (LTR, UTR, GAG, POL) of IAP-Ez retrotransposons were cloned from mouse genomic DNA upstream of the EGFP reporter cassette. Titers of the ecotropically pseudotyped lentiviruses were determined by titration on T37 HeLa cells and EGFP-FACS. After transduction of MEF and mES cells the number of EGFP-positive cells was determined using FACS. B) FACS plots of a representative dataset showing that the GAG silencing element represses EGFP reporter gene expression in mES cells but not in MEF cells. C) Quantification of the individual reporter gene silencing capabilities of the different IAP-Ez sequence elements. The GAG sequence element strongly silences the EGFP reporter gene in mES cells. The relative percentage of EGFP-positive cells was calculated by normalizing the percentage of EGFP-positive cells to the percentage of EGFP-positive cells transduced with an empty reporter vector carrying only EGFP but



having an identical virus titer. The plot represents three individual experiments of each reporter vector. Error bars indicate the standard deviation. The raw data contain some data points acquired by experiments by Katharina Schmidt who I supervised during her medical PhD project (Schmidt, 2014).

2.9. The minimal silencing element inside the IAP-GAG sequence is 160 bp long and requires the histone methyltransferases Setdb1 for its repressive activity.

Please note that the detailed mapping data in this passage and the experiments in different knockout mES cells were obtained and performed by Katharina Schmidt, who I supervised during her medical PhD thesis in our lab. Detailed descriptions of these experiments can be found in her doctoral thesis, once it is finished (Schmidt, 2014).

The GAG-sequence element of IAP retrotransposons I identified in Figure 2.8 is roughly 2 kb long. For identification of the core silencing element inside this region, a large number of different subfragments were cloned into the lentiviral reporter vector and their silencing properties in mouse ES cells and MEF cells were determined as described in Figure 2.8. A small subset of the analyzed fragments is shown in Figure 2.9A. First, a larger fragment of 400 bp (GAG2) was identified that contained the majority of the transrepressive silencing activity (Figure 2.9B). Finally this GAG2 element could be narrowed down to a minimal silencing element of 160 bp, named GAG2.22 (Schmidt, 2014) (Figure 2.9). I also cloned the GAG2.22 sequence in front of a different promoter to test whether GAG2.22 silencing behavior differs when the EGFP transgene is driven by mPGK promoter rather than the human EF1α promoter (Figure 2.9B). No obvious difference in silencing properties could be observed (Figure 2.9B). However, since the EF1α promoter-driven EGFP transgene generates a higher fluorescence of the cells, all following experiments were performed using EF1α promoter driven EGFP reporters.

Next, we asked whether the repressive activity of the GAG2.22 sequence element on the EGFP reporter cassette was dependent on the recruitment of known heterochromatin proteins. Therefore, different mouse embryonic stem cell lines deficient for heterochromatin proteins were analyzed for their ability to silence the EGFP reporter dependent on the GAG2.22 sequence (Figure 2.9C). Mouse ES cells deficient for Dnmt1, Dnmt3a and Dnmt3b, Suv39h enzymes, and Suv4-20h enzymes silence the GAG2.22 reporter just like wild type cells (Figure 2.9C).

Yet, mES cells deficient for the histone methyltransferases Setdb1 could no longer silence the lentiviral GAG2.22-EGFP reporter cassette (Figure 2.9C).

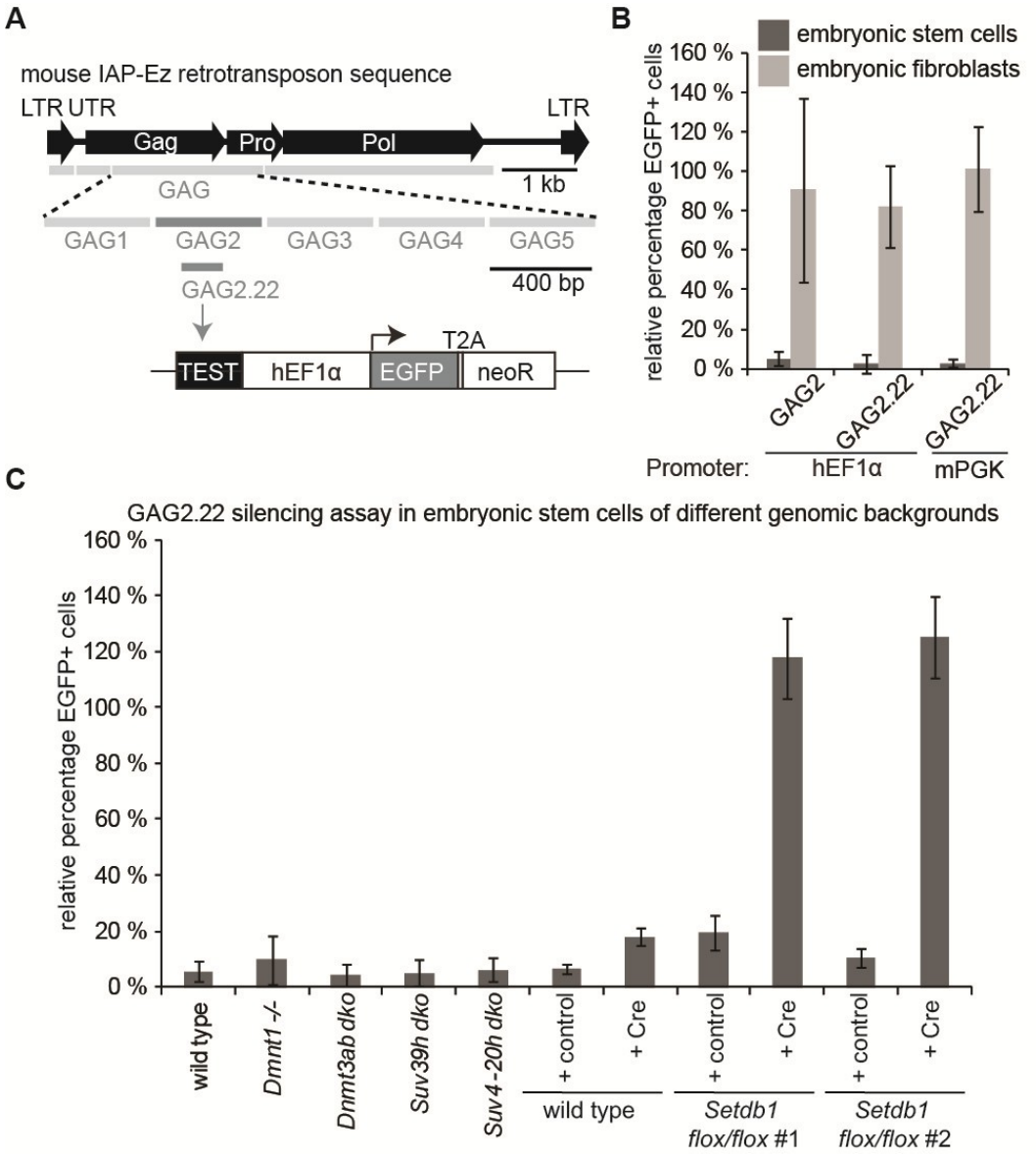


Figure 2.9 The minimal silencing element inside the IAP-GAG sequence is 160 bp long, can silence different promoters and requires the histone methyltransferases Setdb1 for silencing.
A) The cartoon indicates the position of the IAP sequence elements that are responsible for the core silencing activity of the IAP-GAG element. B) The sequence elements in A) were cloned in front of the EGFP expression cassette either driven by the hEF1α promoter or the mPGK promoter into lentiviral reporter vectors. Mouse ES cells and MEF cells were transduced with the according lentivirus. The percentage of EGFP expressing cells was measured in FACS two days after viral transduction. The experiment shows that the GAG2 and GAG2.22 silencing element is sufficient to robustly repress expression of the EGFP transgene-cassette in mES cells but not MEF cells. The GAG2.22 sequence can silence the mouse PGK promoter as well as the human EF1α promoter. The barplot is generated from three independent experiments. Error bars indicate the standard deviation. C) Setdb1 but not Suv39h, Suv4-20h, Dnmt1 or Dnmt3ab is required for silencing of the GAG2.22 sequence. Mouse embryonic stem cell lines with indicated genomic backgrounds were transduced with a lentiviral EGFP reporter vector carrying the GAG2.22 silencing element. In the case of Setdb1, mouse ES cells were used that carry a floxed *Setdb1* allele and knockout was applied two days before reporter transduction using a lentivirus enconding the Cre recombinase. The percentage of EGFP expressing cells was measured in FACS two days after reporter transduction. The relative percentage of EGFP-positive cells was calculated as described in B). The barplot is generated from three independent experiments. Error bars indicate the standard deviation. Raw data of B) and C) originate from Katharina Schmidt (Schmidt, 2014).

Please note that Setdb1 is required for pluripotency of mES cells and cells don't survive in the absence of Setdb1 (Bilodeau et al., 2009; Dodge et al., 2004).

Therefore, mES cells were used which contained floxed Setdb1 gene loci. For analyzing *Setdb1* depleted mES cells, cells were transduced with a lentivirus encoding the *Cre recombinase* two days before GAG2.22-EGFP reporter transduction.

In summary, GAG2.22 a very potent silencing element of 160 bp was identified inside of IAP retrotransposons. In addition, GAG2.22 silences EGFP reporter gene expression in mouse ES cells in a *Setdb1*-dependent manner.

2.10. The GAG2.22 sequence recruits Histone 3 Lysine 9 trimethylation when integrated into chromatin of mES cells

GAG2.22, a 160 bp sequence of mouse IAP retrotransposons, silences strong constitutive reporter genes (Figure 2.9). Silencing occurs in mES cells and requires the histone methyltransferase Setdb1 (Figure 2.9).

Endogenous IAP retrotransposons are modified with H3K9me3 (Mikkelsen et al., 2007). However, whether the short GAG2.22 sequence element is really a nucleation site of IAP retrotransposon silencing and is sufficient to recruit H3K9me3 independently of the surrounding IAP sequence, is unknown. Therefore, a single copy of the GAG2.22 sequence was integrated into the genome of mES cells using a recombinase-mediated cassette exchange system (Baubec et al., 2013; Lienert et al., 2011b) (Figure 2.10A). For this purpose, the GAG2.22 sequence was directly cloned in a plasmid downstream of an EGFP-T2A-ZeoR reporter gene under the control of an inactive Tet-inducible promoter (TRE). The full cassette was flanked with inverted loxP sites to enable a cassette exchange with a floxed and stably integrated locus harboring a HygroR-HSV-TK marker gene (Figure 2.10). After co-transfection of the transfer plasmid with a plasmid encoding the Cre recombinase successfully recombined clones were devoid of the HygroR-HSV-TK gene and sensitive to ganciclovir selection.

Chromatin immunoprecipitation experiments revealed that the integrated GAG2.22 cassette was strongly enriched for H3K9me3 (Figure 2.10B). H3K9me3 also spread across the entire reporter locus.

In summary, the GAG2.22 sequence element of IAP retrotransposons autonomously recruits H3K9me3 to nearby chromatin.

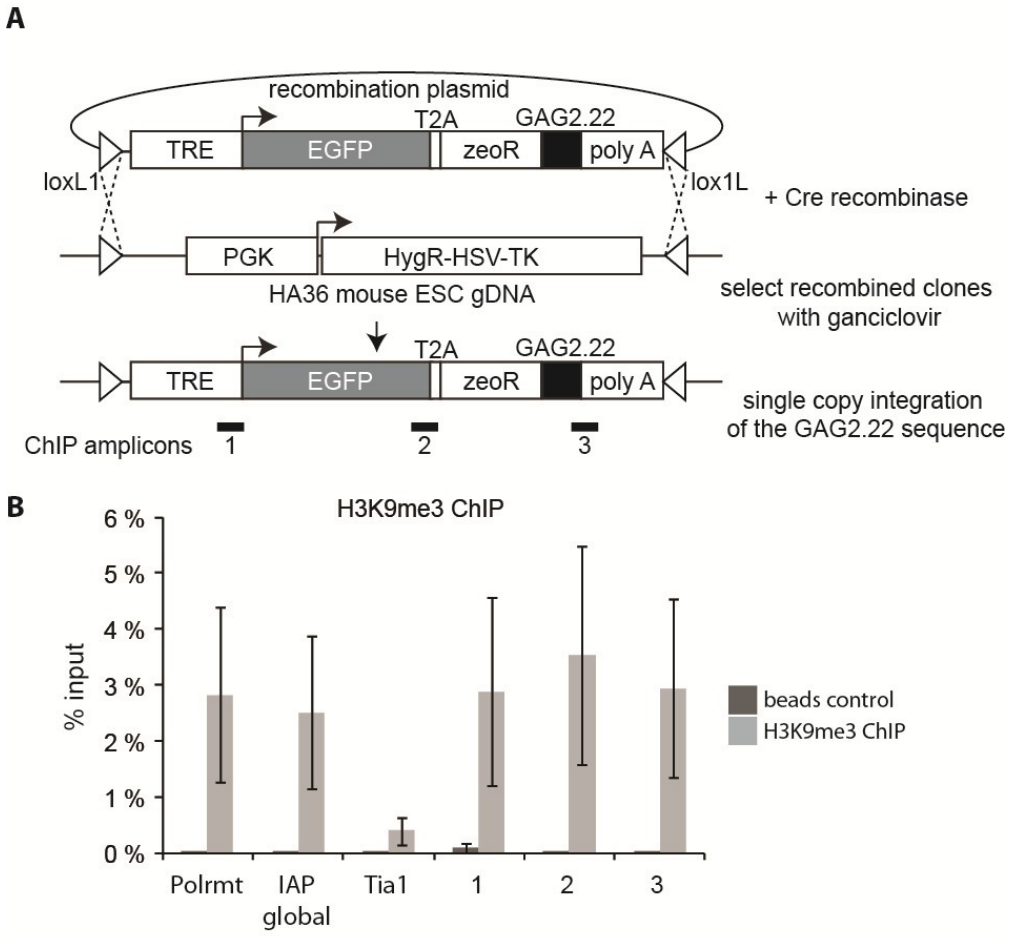


Figure 2.10 The GAG2.22 sequence recruits Histone 3 Lysine 9 trimethylation in mouse ES cells
A) A recombinant cassette containing the GAG2.22 sequence is inserted into HA36 mouse ES cells using recombinase-mediated cassette exchange (RMCE)(Baubec et al., 2013; Lienert et al., 2011b). First, a plasmid flanked by inverted loxP sites is co-transfected into HA36 mouse ES cells together with a plasmid encoding the Cre recombinase. The cell line contains a positive and negative selection marker (HygR-HSV-TK) flanked by inverted loxP sites that is removed after successful RMCE allowing selection of recombined cell clones using ganciclovir. Numbered black bars indicate the position of qPCR amplicons used for ChIP-qPCR in figure B. B) Chromatin immunoprecipitation experiment. H3K9me3 is recruited to a positive control region (Polrmt), to IAP elements (IAP global) and to the GAG2.22 sequence. H3K9me3 spreads over the entire reporter locus. Tia1 serves as a negative control. The barplot represents the average of two biological replicates. Error bars indicate the standard error of the mean.

2.11. The IAP-GAG silencing sequences are silenced in a *Trim28*-dependent manner

Setdb1 has been found to be essential for silencing of the GAG silencing element of IAP retrotransposons in reporter assays (Figure 2.9C). This is consistent with the finding that *Setdb1* knockout mES cells show a strong upregulation of endogenous retroviral elements including IAP retrotransposons (Matsui et al., 2010). Apart from *Setdb1* also *Trim28* has been implicated into silencing of retrotransposons in mES cells (Rowe et al., 2010). Therefore, I wondered, whether *Trim28* is also required for the recognition and silencing of EGFP reporter genes in a GAG2.22-dependent-manner. I depleted *Trim28* from wild type

mES cells using lentiviral short-hairpin RNAs (shRNAs). *Trim28* depletion by four independent shRNAs led to an impairment of GAG2.22-EGFP reporter silencing (Figure 2.11A). One of the Trim28 shRNAs was tested in three different mouse embryonic stem cell lines to rule out cell line-specific effects. In addition, I tested, whether the longer 400 bp parental GAG2-sequence fragment that contains the GAG2.22 sequence is also silenced in a *Trim28*–dependent manner. Indeed, *Trim28* knockdown led to impaired silencing of the GAG2-EGFP reporter in three different cell lines (Figure 2.11B). Because I only analyzed Trim28 knockdown cells and did not investigate *Trim28* knockout cells, I could not judge whether the effect of Setdb1 depletion on reporter silencing was severer than the effect of Trim28 depletion. Notably, the reduction of Trim28 mRNA by the shRNA used in Figure 2.11B was on average only around 60 % (Figure 2.11C).

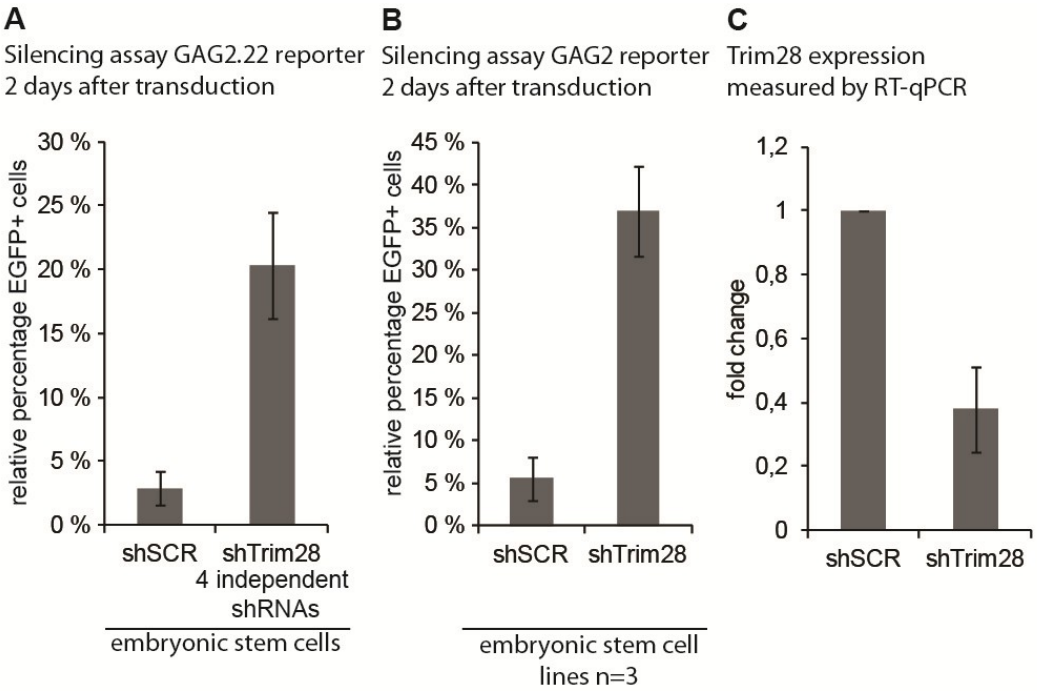


Figure 2.11 Combining the EGFP-based GAG-silencing assay with shRNA knockdowns reveals that the IAP-GAG sequence is silenced in a Trim28-dependent manner.

A) Mouse ES cells were transduced with lentiviral shRNAs against *Trim28* or a scrambled control sequence. Cells were grown for two days and selected with 2 µg/ml puromycin for additional two days. Afterwards cells were transduced again with lentiviral EGFP reporters and the silencing potential of the GAG2.22 sequence was measured after additional two days as described in Figure 2.9. The barplot shows an average of four independent shRNAs targeting Trim28. Error bars represent the standard deviation. B) Different wildtype cell lines silence the GAG2 sequence in a *Trim28*-dependent manner. Wild type mouse embryonic stem cell lines were used and tested for their silencing potential of the GAG2 sequence after knockdown with one Trim28 shRNA. The figure represents an average of three experiments using different wild type cells lines. Error bars represent the standard deviation. C) The average amount of Trim28 knockdown generated by the Trim28 shRNA used in Figure 2.11B is around 60 %. The expression was measured by quantitative RT-PCR and normalized to a scrambled shRNA control. The barplot represents four experiments. Error bars represent the standard deviation.

In summary, the GAG2.22 or the GAG2 sequence of the mouse IAP retrotransposon silenced EGFP reporter genes in a *Trim28*-dependent manner. Combining shRNA depletions of potential candidate genes with lentiviral EGFP reporter assays can identify genes that are involved in GAG-mediated transcriptional repression.

2.12. Optimization of two genome-wide shRNA screening assays for factors involved in reporter gene-silencing

Sh-RNA-mediated depletion of *Trim28* led to impaired repression of the GAG2.22-EGFP reporter (Figure 2.11). Therefore, unbiased RNA-interference (RNAi) screening experiments could potentially identify novel factors involved in reporter gene silencing of this sequence. An easy way of testing a large set of shRNA sequences is using pooled lentiviral shRNA libraries (Figure 2.12) (Moffat et al., 2006; Ngo et al., 2006; Silva et al., 2008). These libraries consist of a mix of thousands of different viral particles of which each encodes a different shRNA sequence targeted against a different mRNA transcript. After transduction of a cell line with a pooled shRNA library at a low multiplicity of infection every individual cell takes up only one shRNA and downregulates the respective mRNA. This heterogenic cell pool can be used to isolate individual cells that bypass the silencing of a lentiviral EGFP silencing reporter. If an shRNA in an individual cell downregulates a factor that is required for EGFP reporter gene silencing, EGFP expression will be maintained in this cell. Because lentiviral shRNA sequences stably integrate into the genome of the transduced cells, shRNA sequences leading to a bypass of silencing can be identified by PCR after sorting EGFP-positive cells using a FACS sorter. An EGFP reporter not bearing a silencing sequence can be used as a negative control and to normalize for unequal representations of shRNA sequences in the shRNA pool (Figure 2.12).

Except *Trim28* and *Setdb1* other factors involved in IAP retrotransposon silencing initiation are unknown. Since the GAG2-EGFP reporter silencing assay mirrors many aspects of endogenous IAP silencing in mouse ES cells like *Setdb1*- and *Trim28*-dependency (Figure 2.11B, Figure 2.9B and C), a genome-wide screening protocol was optimized on the GAG2-EGFP repression assay (Figure 2.12B). The required amount of pooled shRNA virus, the optimal puromycin concentration, virus titers of the lentiviral EGFP reporters, the choice whether the GAG, GAG2 or GAG2.22 elements should be used for screening and the optimal time points for all steps were empirically optimized for a good signal to noise ratio between control shRNAs and shRNAs targeting *Trim28* (data not shown). The final protocol of the genome-wide shRNA screen is shown in Figure 2.12B.

Because IAP-GAG-mediated silencing only occurs in mES cells (Figure 2.8), the shRNA screen had to be performed in these cells. Mouse ES cells are rather sensitive cells that

respond to a variety of stimuli with cell death. Precautionary, a similar screening approach was also established in more robust MEF cells to control, if my screening approach in general results in the identification of functional hits. For this second screening experiment the Zfp809-EGFP reporter silencing assay was chosen, which also works in MEF cells (Figure 2.6). *Zfp809*-dependent silencing constructs were already established (Figure 2.6). Moreover, *Zfp809*-dependent silencing is also dependent on the corepressor protein Trim28, H3K9me3 and Hp1 proteins (Wolf et al., 2008; Wolf and Goff, 2007, 2009). Screening parameters for the Zfp809 control screen were determined accordingly using shRNAs against *Zfp809* and *Trim28* (data not shown and Figure 2.12A).

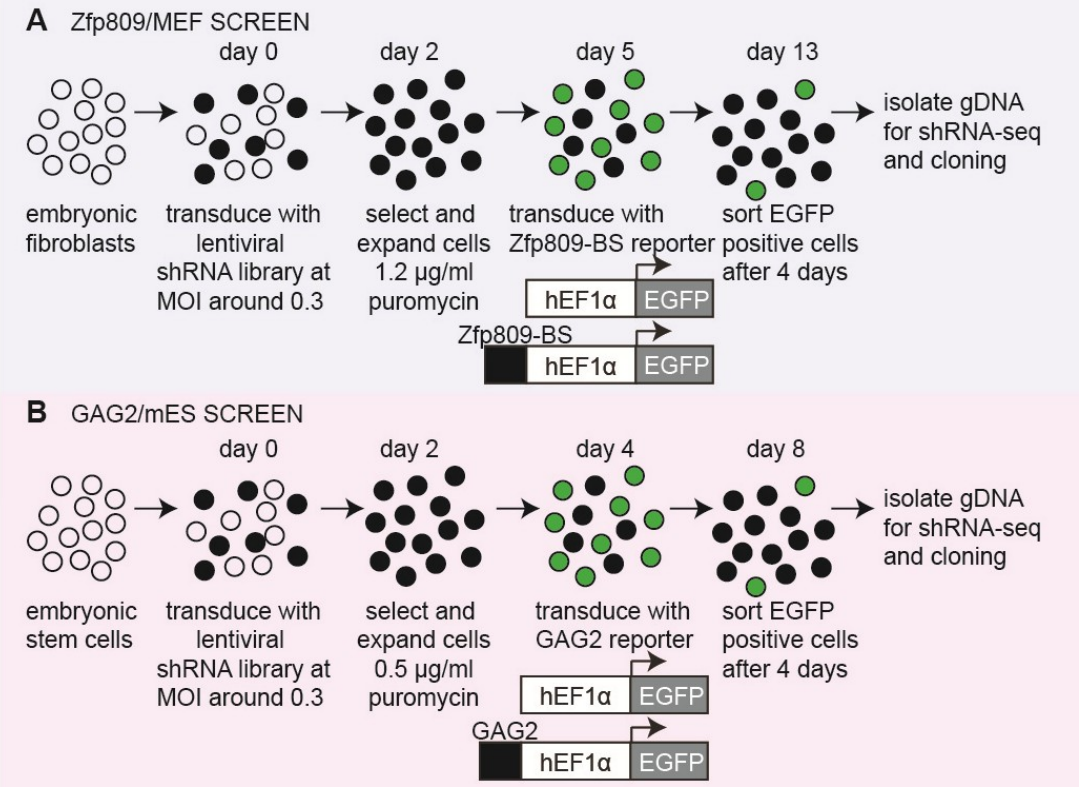


Figure 2.12 Optimized and final conditions of the two independent genome-wide shRNA screening approaches
A) The screening protocol was optimized for the identification of genes involved in heterochromatic silencing by Zfp809 and Trim28 in MEF cells. Another motivation for this screen was to test the feasibility of the shRNA screening combined with lentiviral reporters in a robust differentiated cell line. MEF cells were transduced with a genome-wide pool of mixed lentivirus-encoded shRNAs. A low multiplicity of infection (MOI) was used to ensure only one shRNA integrated in each cell. After selection of transduced cells with an optimized puromycin concentration for three days, the cell pool was split and either transduced with an EGFP reporter virus containing one or no Zfp809 binding site. The EGFP reporter containing the Zfp809 binding site was silenced unless particular shRNAs led to a bypass of silencing. Cells that bypassed silencing were EGFP-positive and were sorted using a FACS sorter. The shRNA sequences integrated in these cells were identified by deep-sequencing of PCR products from shRNA sequences. The EGFP reporter without a Zfp809 binding site was used as a negative control and to normalize for unequal representations in the shRNA pool. B) Optimized conditions for the GAG2/mES screen are shown for identifying genes involved in IAP silencing in mouse ES cells. The screen was carried out with the GAG2 sequence as described in A) except for the indicated changes in the protocol.

In summary, two pooled genome-wide shRNA screens were set up. The first screen aimed for the identification of factors that together with Setdb1 and Trim28 silence the GAG region of IAP retrotransposons. The second screen was done to prove the feasibility of pooled shRNA screening on EGFP reporter assays and to find additional factors required for *Zfp809*-dependent and *Trim28*-dependent silencing in MEF cells.

2.13. Genome-wide screening for Zfp809 and IAP-GAG-dependent silencing identifies many potential genes involved in heterochromatic gene silencing

Two genome-wide shRNA screening experiments have been carried out as described in Figure 2.12. EGFP-positive cells that bypass silencing were isolated using a FACS sorter and genomic DNA was isolated. In parallel cells transduced with an EGFP reporter without a silencing sequence were also harvested to normalize for unequal representations of shRNAs in the initial virus pool.

The shRNA sequences integrated in the genome of the cells were isolated via PCR and the sequences were identified using next-generation sequencing. Sequence reads were aligned to the respective shRNA sequences in the TRC library documentation (Moffat et al., 2006). ShRNAs that were not identified by sequencing or shRNAs identified with less than 20 reads in the EGFP-only control sample were removed from the analysis. The number of total reads between different samples was normalized and reads enriched more than three times were designated as hits (Figure 2.13A).

In total 1144 genes scored as hits in the Zfp809/MEF control screen. Please note that Trim28 was among the two strongest hits scoring with three independent shRNAs (Figure 2.13B). 2379 genes scored as hits in the GAG silencing screen with *DNA topoisomerase 2-beta* as the strongest hit.

The number of scoring genes is relatively high, probably reflecting the fact that FACS sorting displayed a bottleneck in the workflow. Since only 200.000 EGFP-positive cells could be obtained that bypassed silencing of the Zfp809 reporter and only 30.000 EGFP-positive cells could be obtained that bypass silencing of the GAG2 reporter, a high shRNA library representation during next generation sequencing could not be obtained. However, since *Trim28* was identified as the strongest hit in the control screen, both lists will likely contain a large number of *bona fide* genes involved in heterochromatic silencing. The full lists of hits from both genetic screens can be found in the appendix.

A

77723 shRNAs → remove shRNAs not detected in sequencing
→ remove shRNAs having less than 20 reads
→ correct for different number of total reads between samples
→ check for shRNAs that are 3-fold enriched over negative control
→ check for multiple shRNAs targeting the same gene in hit list

B

Zfp809/MEF SCREEN	2 genes scoring with 3 independent shRNAs: ***Trim28, Gak*** 42 genes scoring with 2 independent shRNAs 1100 genes scoring with 1 shRNA
GAG2/mES SCREEN	1 gene scoring with 4 independent shRNAs: ***Top2b*** 5 genes scoring with 3 independent shRNAs: ***Actrt2, Dach2, Rundc3b, Hoxb3, Zbtb20*** 147 genes scoring with 2 independent shRNAs 2226 genes scoring with 1 shRNA

Figure 2.13 Genome-wide screening for Zfp809 and IAP-GAG-dependent silencing identifies many potential genes involved in reporter silencing
A) Workflow of how the shRNA-seq reads of the two different samples of each screen were processed after alignment to the shRNA library to identify hits of the primary screen. Next generation library preparation and sequencing was carried out at Partners HealthCare Center for Personalized Genetic Medicine (PCPGM) Cambridge, MA, USA. Alignment of reads to the shRNA library has been performed by Prof. Dr. Gunnar Schotta. B) The table shows the most prominent hits of the two screens and the number of genes identified. In the Zfp809/MEF screen *Trim28* is the strongest hit.

2.14. A secondary genetic-screen identifies *Atrx* as a regulator of GAG-sequence silencing of IAP retrotransposon sequences

Genome-wide screening for *Zfp809*-dependent silencing in MEFs identified the corepressor *Trim28* as the strongest hit, which is a known and validated regulator for *Zfp809*-mediated repression (Wolf and Goff, 2007, 2009). This underlines that the method for genome-wide shRNA screening on reporter genes identifies functional regulators of silencing. However, since a lot of the regulation on IAP retrotransposons is still unknown, I focused on the validation of the GAG-EGFP shRNA screen (Figure 2.13B). In addition, our genome-wide screening approach identified a long list of potential genes involved in IAP silencing. Therefore, secondary screening of individual shRNA sequences was required.

A small candidate library of 75 lentiviral shRNA vectors was cloned directly from isolated DNA of the primary screen to validate individual shRNAs that bypass reporter gene silencing (Figure 2.14A). These lentiviral shRNA vectors were packaged into lentivirus and mES cells were transduced. After selection of knockdown cells with puromycin, the GAG2-EGFP reporter or an EGFP control lentivirus was transduced and GAG2-dependent silencing was determined using FACS as described in Figure 2.9.

Silencing potential of the individual shRNAs was compared to an shRNA targeting *Trim28* as a positive control and to an shRNA targeting a scrambled control sequence. The best performing shRNAs in the screen were sequenced using Sanger sequencing. An shRNA

targeting *Atrx* was found to have the strongest effect in the secondary screen (Figure 2.14A). The effect of the shRNA on *Atrx* mRNA level was determined by RT-qPCR showing that Atrx mRNA levels were indeed reduced upon shRNA transduction (Figure 2.14B).

In summary, a secondary screen of the primary GAG2-mES shRNA screen verified shRNAs that significantly reduce GAG2-EGFP reporter gene silencing. The highest scoring shRNA of the secondary screen targets the *Atrx* gene.

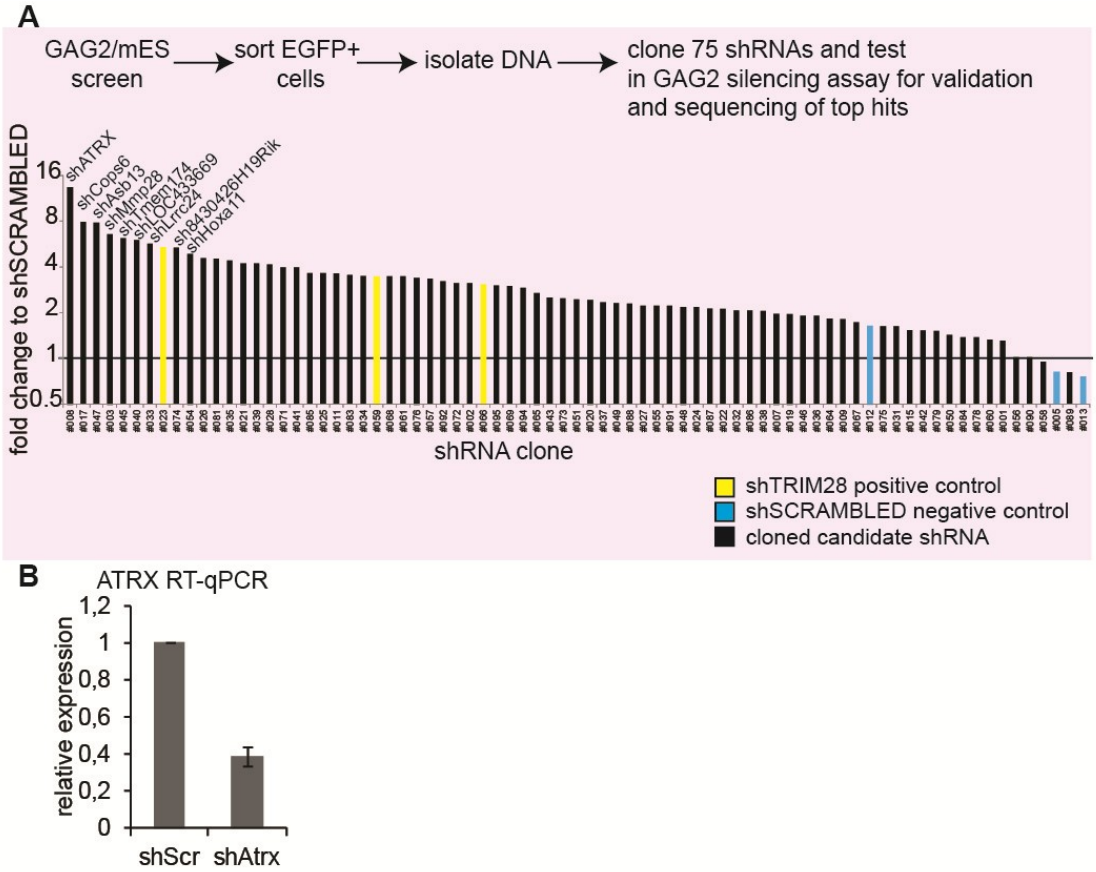


Figure 2.14 A secondary screen identifies Atrx as a regulator of IAP/GAG silencing
A) 75 shRNA vectors were directly cloned from DNA of EGFP-positive cells of the primary screen and analyzed in a secondary screen. Mouse ES cells were transduced with the 75 shRNA viruses in 96-Wells and selected with puromycin after two days. After additional two days the cells were transduced with either GAG2-EGFP reporter virus or an EGFP-only control virus. The different knockdown cells were analyzed for their silencing potential of the GAG2-EGFP reporter after additional two days. The relative percentage of EGFP-positive cells for every knockdown cell line was normalized to the relative percentage of the EGFP-only control. Control shRNAs where included in the experiment as an internal reference (shRNA against a scrambled sequence shown in blue, shRNA against Trim28 shown in yellow). The best performing shRNA-vectors were sequenced using standard Sanger sequencing. The targeted genes of these shRNAs are listed above the barplot. B) The identified shRNA against Atrx downregulates Atrx mRNA. Cells were transduced with shAtrx or an shRNA targeting a scrambled sequence. After selection of transduced cells with puromycin, RNA was isolated and reverse transcriptase-qPCR was performed. The barplot represents three individual experiments. Error bars indicate the standard deviation.

2.15. *Atrx* knockout cells show strongly retarded silencing kinetics of the GAG2.22-EGFP reporter

ShRNA-screening revealed that an shRNA targeting *Atrx* strongly impairs silencing of the GAG-EGFP reporter (Figure 2.14). Atrx is a known chromatin-associated SNF2-type ATPase that contains an N-terminal ADD-domain which binds to methylated H3K9 and which is implicated into the human X-lined alpha-thalassemia mental retardation syndrome (Eustermann et al., 2011; Gibbons et al., 2000; Gibbons et al., 1995; Iwase et al., 2011; Picketts et al., 1996) (Introduction 1.3.3). Therefore, a direct role of *Atrx* into *Setdb1*- and *Trim28*-dependent silencing of IAP retrotransposons was suggested. However, RNAi experiments often result in a substantial number of off-target effects and acquired hits require careful validation (Jackson et al., 2003). Thus, the connection between *Atrx* and retrotransposon-silencing was validated by analyzing mES cells deficient for *Atrx*.

CRISPR/Cas9 technology was used to disrupt the *Atrx* locus with RNA-guided-nucleases (RGN) (Figure 2.15A). Male mES cells were co-transfected with a plasmid containing a puromycin-resistance and the pX330 plasmid encoding a codon-optimized Cas9 enzyme and a small guide RNA (sgRNA or crRNA) targeting the N-Terminal region of the *Atrx* ORF (Cong et al., 2013; Mali et al., 2013) (Figure 2.15A). After selection of transfected cells using a short puromycin treatment, individual cell clones were analyzed for mutations at the *Atrx* locus. Four cell clones having frame-shift mutations were identified using Sanger sequencing of PCR products and verified by western blotting (Figure 2.15A and C).

Next, *Atrx*-deficient cell lines were analyzed for their ability to silence the GAG2.22-EGFP reporter cassette as described in Figure 2.9. All four *Atrx*-deficient cell lines are not completely defective in silencing, but show strongly retarded silencing kinetics (Figure 2.15B). However, this is consistent with the finding that knockdown of *Atrx* using a lentiviral shRNA leads to impaired silencing of the GAG2-EGFP reporter gene. To rule out CRISPR-dependent off-target effects, an independent mouse ES cell line was targeted with different sgRNAs against Atrx giving the same results (data not shown).

In summary, *Atrx*-deficient cells were generated via RNA-guided nucleases. Knockout of *Atrx* results in impaired silencing of the GAG2.22-EGFP reporter gene.

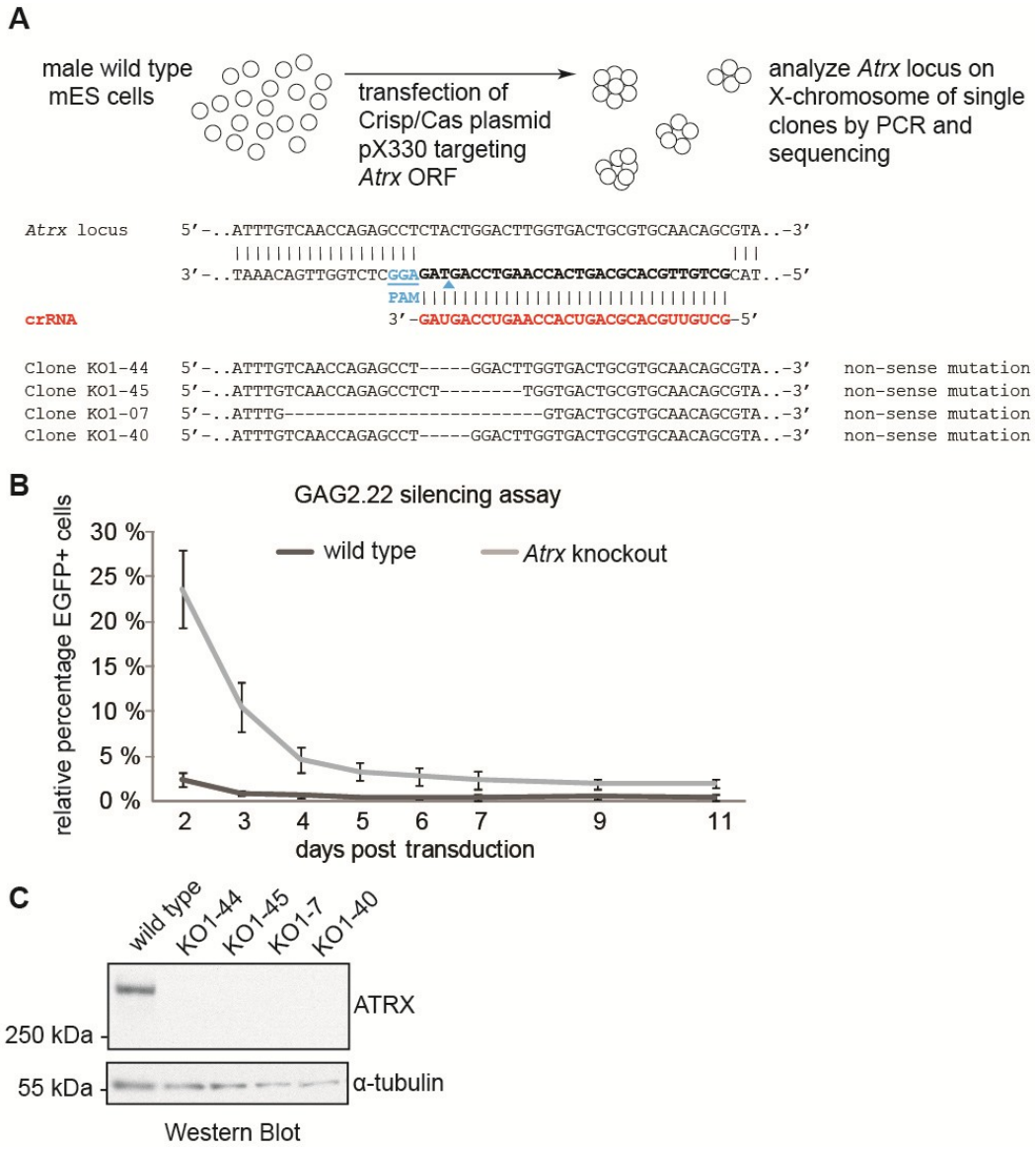


Figure 2.15 *Atrx* knockout cells show strongly retarded silencing kinetics of the GAG2.22-EGFP reporter
A) Generation of *Atrx* knockout cells via transient transfection of CRISPR/Cas9 plasmids. The used crRNA and the targeted position of the *Atrx* locus are shown. Single mouse embryonic stem cell colonies were expanded after CRISPR/Cas9 transfection and analyzed for mutations in the *Atrx* open reading frame. The *Atrx* locus also lies on the X-chromosome in mice, and male cells only contain one *Atrx* allele. The Sanger sequencing result of four independent clones revealed non-sense mutations in these cells. B) *Atrx* knockout cells were transduced with the GAG2.22-EGFP reporter as described in Figure 2.9 and the silencing of the reporter was monitored at the indicated time points using FACS. The plot represents three experiments of four independent *Atrx* knockout cell clones and two *Atrx* expressing control cell lines. Error bars represent the standard deviation. The plot in B) contains data points acquired by Katharina Schmidt (Schmidt, 2014). C) Western blot verifying that the generated *Atrx* knockout cells lost Atrx protein expression.

2.16. Atrx and H3K9me3 are enriched at endogenous IAP elements in mouse ES cells and are recruited to novel integrations of the GAG2.22 sequence

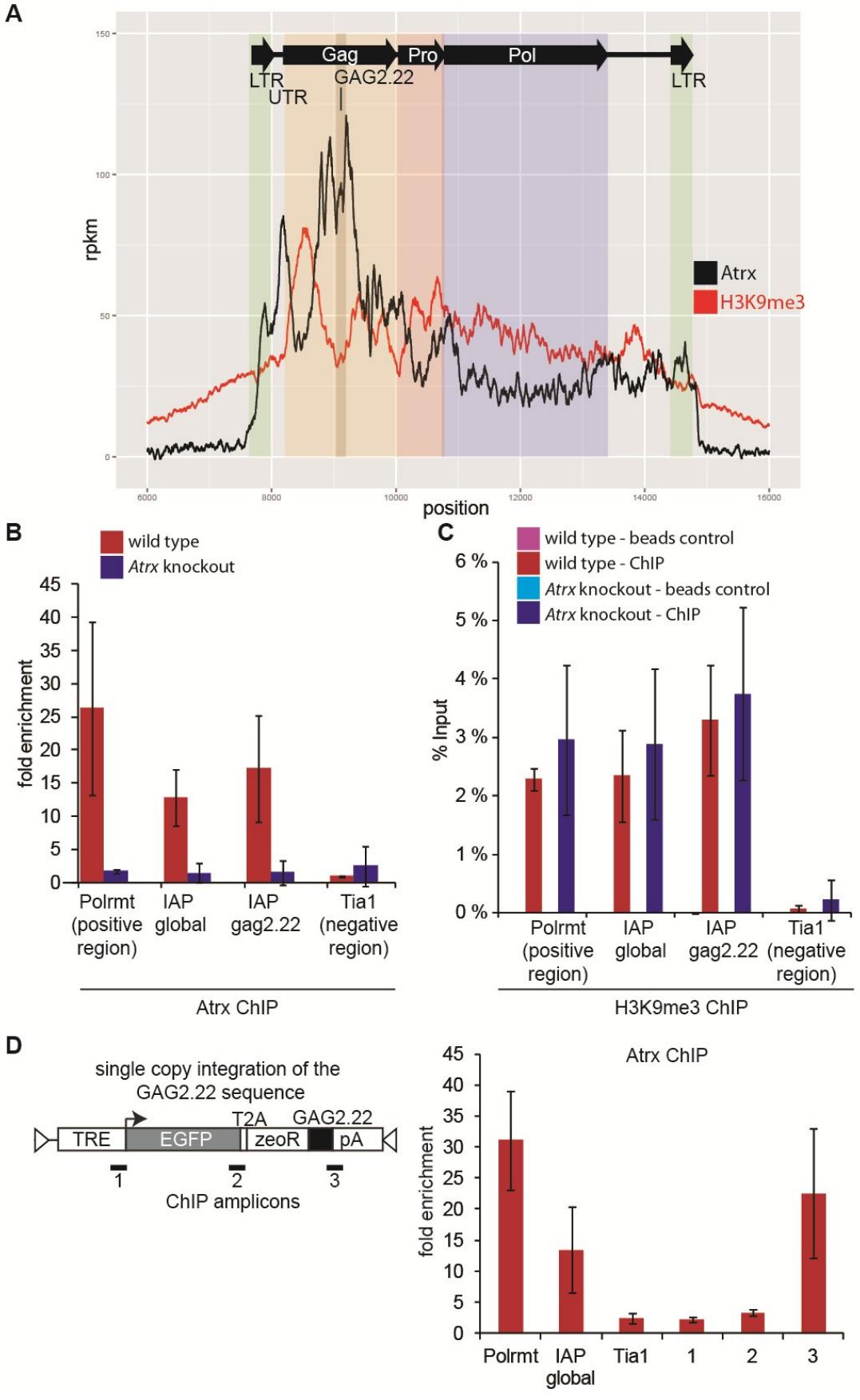
Depletion of *Atrx* via RNAi or RNA-mediated nucleases leads to defects in silencing of a lentiviral EGFP reporter using the GAG2.22 sequence element of IAP retrotransposons

(Figure 2.14 and Figure 2.15). Therefore, it was suggested that Atrx also plays a role at endogenous IAP elements in mES cells. Published datasets of Atrx ChiP-sequencing experiments in mES cells and ChIP-sequencing reads for H3K9me3 that were generated with mES cells in our lab were aligned to the mouse genome ((Law et al., 2010) and unpublished data). Full-length IAP element copies in the genome were identified and aligned at the GAG2.22 sequence. The relative number of ChIP-sequencing reads over these full length IAP elements shows a strong enrichment of Atrx at these sites (Figure 2.16A). Similarly, H3K9me3 is also found at IAP sequences in mES cells, as it has been reported before (Karimi et al., 2011; Mikkelsen et al., 2007). Interestingly, Atrx binding peaks directly next to the GAG2.22 sequence while H3K9me3 is more widely distributed over the IAP sequence and flanking genomic regions (Figure 2.16A). Similar data were obtained when Atrx enrichment on IAP elements was measured using ChIP-qPCR (Figure 2.16B). Enrichment of H3K9me3 on IAP elements and positive control regions is not changed in *Atrx* knockout mES cells, indicating that Atrx depletion has not an immediate impact on H3K9me3 in mES cells and seems to be downstream of H3K9 trimethylation (Figure 2.16C).

I analyzed Atrx binding in cells engineered to carry a single copy integration of an EGFP-GAG2.22 sequence via recombinase-mediated cassette exchange (RMCE) to see whether the GAG2.22 sequence can autonomously recruit Atrx (Figure 2.10 and 2.16D). Atrx is bound to the GAG2.22 sequence but does not spread over the reporter locus (Figure 2.16D).

In summary, Atrx protein and H3K9me3 are present at endogenous IAP retrotransposons in mES cells. *Atrx* deficiency does not lead to a detectable reduction of H3K9me3 in these cells. However, the GAG2.22 sequence autonomously recruits Atrx to nearby chromatin.

Figure 2.16 Atrx and H3K9me3 are enriched at endogenous IAP elements and are recruited to novel integrations of the GAG2.22 sequence
A) Full-length IAP copies were aligned at the GAG2.22 site and the relative number of ChiP-Seq reads for H3K9me3 and Atrx was plotted on the Y-axis. The GAG2.22 sequence is highlighted in dark ocher inside the GAG sequence. Atrx ChIP-seq data have been published by others (Law et al., 2010), H3K9me3 ChIP-seq data originate from unpublished datasets generated in our lab. Data analysis was performed by Prof. Gunnar Schotta. Atrx not only localizes to IAP elements *in vivo* but also Atrx enrichment peaks next to the GAG2.22 sequence. B) Atrx chromatin immunoprecipiation. Atrx is bound to positive control regions (Polrmt) and to IAP elements, but not to a negative control region (Tia1). The barplot indicates an average of two independent ChIP experiments. *Atrx* knockout cells were used as a negative control to control for antibody specificity. Error bars indicate the standard deviation. C) H3K9me3 chromatin immunoprecipitation in wild type and *Atrx* knockout cells. *Atrx* knockout cells maintain prominent levels of H3K9me3 at a positive control region (Polrmt) and at IAP elements. Tia1 serves as a negative control. The barplot represents an average of two experiments of wild type cells and three experiments of different *Atrx* knockout cell clones. Error bars indicate the standard deviation. D) Atrx ChIP in cells engineered to carry a single copy integration of an EGFP-GAG2.22 transgene via RMCE (see Figure 2.10). The newly integrated GAG2.22 sequence recruits Atrx. However, Atrx does not spread to neighboring chromatin. ChIP-Data have been normalized to background binding to the beads. The barplot shows an average of two experiments. Error bars indicate the standard deviation.


A
Gag
Pro
Pol
LTR
UTR
GAG2.22
LTR
rpkm
Atrx
H3K9me3
position
B
wild type
Atrx knockout
fold enrichment
Polrmt (positive region)
IAP global
IAP gag2.22
Tia1 (negative region)
Atrx ChIP
C
wild type - beads control
wild type - ChIP
Atrx knockout - beads control
Atrx knockout - ChIP
% Input
H3K9me3 ChIP
D
single copy integration of the GAG2.22 sequence
TRE
EGFP
T2A
GAG2.22
zeoR
pA
ChIP amplicons
Atrx ChIP
fold enrichment
Polrmt
IAP global
Tia1
1
2
3

2.17. *Atrx* knockout cells have normal DNA methylation on IAP retrotransposon sequences and a normal cell cycle distribution

Atrx knockout cells show an impaired silencing on GAG2.22-EGFP and Atrx binds to endogenous IAP elements *in vivo* (Figure 2.15B and Figure 2.16A and B). Yet, H3K9me3 in *Atrx* knockout cells appears to be unaltered (Figure 2.16C). Since endogenous IAP elements are heavily methylated by DNA methylation and Atrx has been reported to influence DNA methylation at repetitive elements (Gibbons et al., 2000; Walsh et al., 1998), DNA methylation analysis was performed in *Atrx* knockout cells to identify potential changes upon *Atrx* depletion. Genomic DNA of wild type and *Atrx* knockout cells was bisulfite-converted to identify changes of DNA methylation. Individual sequences were analyzed by Sanger sequencing (Figure 2.17A). No significant difference in the level of DNA methylation on endogenous IAP elements could be detected between wild type and *Atrx* knockout mES cells (Figure 2.17B).

Atrx depletion has been implicated into chromosome cohesion and congression defects and defects of the meiotic spindle (De La Fuente et al., 2004; Ritchie et al., 2008). In addition, *Atrx* has been involved in telomere integrity (Heaphy et al., 2011; Lovejoy et al., 2012; Wong et al., 2010). Therefore, I wondered whether aneuploidy or differences in the cell cycle distribution occur upon *Atrx* knockout. However, no change in DNA content or cell cycle distribution was detected in *Atrx* knockout cells (Figure 2.17C).

In summary, *Atrx* knockout cells show no perturbed DNA-methylation or cell cycle distribution.

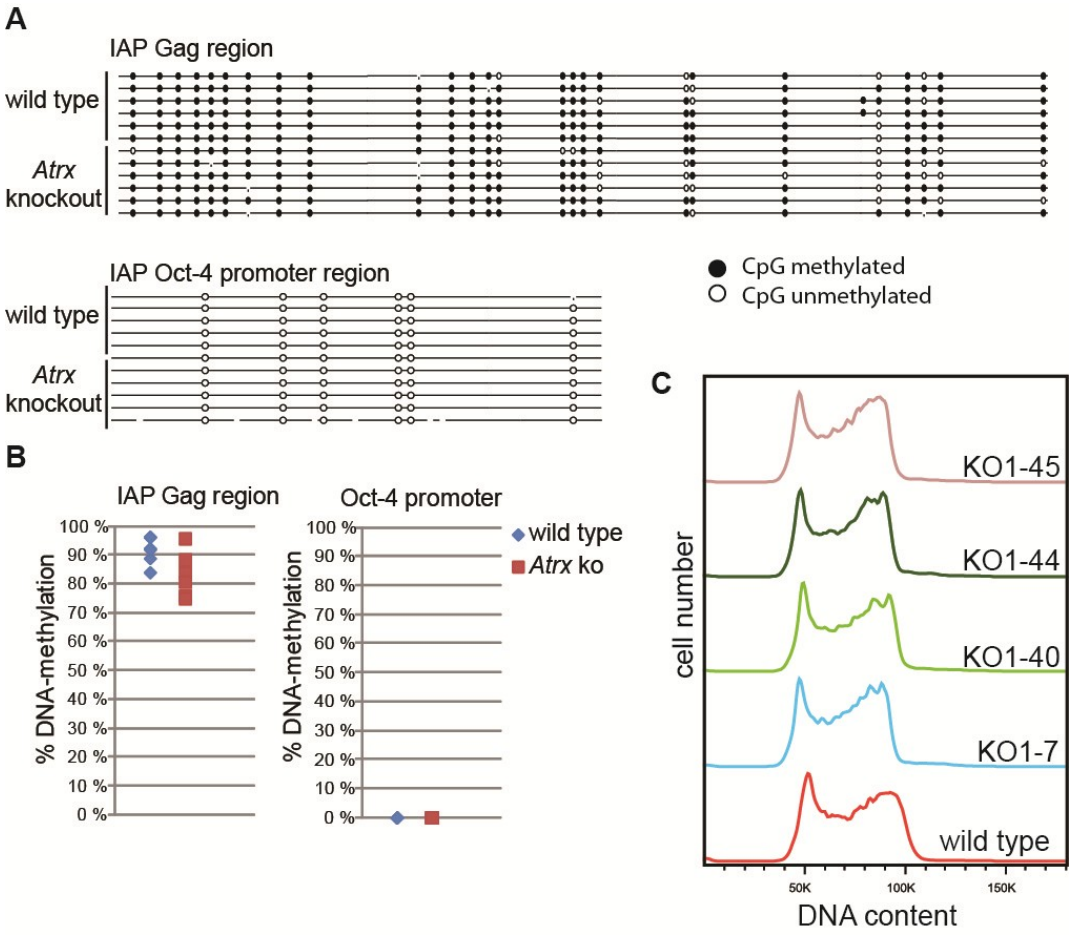


Figure 2.17 *Atrx* knockout cells have normal DNA methylation on IAP retrotransposon sequences and a normal cell-cyle distribution.
A) DNA methylation status of *Atrx* knockout cells at IAP Gag regions and at the Oct-4 promoter. DNA was isolated from cell lines, bisulfite-converted and PCR products of the respective regions were subcloned for Sanger sequencing. The cytosine-methylation state was analyzed via BiQ Analyzer (Bock et al., 2005). B) Quantification of the DNA methylation data. C) *Atrx* knockout mouse embryonic stem cell lines show a normal cell cycle distribution in PI-FACS.

2.18. *Daxx* knockout cells phenocopy the impaired silencing kinetics of *Atrx* knockout cells

Atrx is a known interacting protein of the histone H3.3 chaperone Daxx and several publications suggest a functional interaction of these proteins during deposition of histone H3.3 (Drane et al., 2010; Elsasser et al., 2012; Goldberg et al., 2010; Lewis et al., 2010; Wong et al., 2010). I generated *Daxx* knockout mES cells to see, whether Atrx and Daxx also functionally interact during the repression of IAP retrotransposons. Wild type mES cells were transfected with CRISPR/Cas9 plasmids targeting the *Daxx* ORF (Figure 2.18A), and cell lines that were deficient for Daxx protein were obtained (Figure 2.18B). *Daxx* knockout cells were analyzed for their ability to silence the GAG2.22-EGFP carrying the strong silencing sequence of IAP retrotransposons (Figure 2.9, Figure 2.18C). Similar to *Atrx* knockout mES cells, *Daxx* knockout cells show strongly retarded silencing kinetics. Interestingly, the

strength of remaining silencing activitiy of the GAG2.22-EGFP reporter is highly comparable between *Atrx* and *Daxx* knockout cells. Importantly, knockout of *Daxx* does not lead to a loss of *Atrx* expression (Figure 2.18D). Please note that also *Trim28* expression was not changed in *Atrx* and *Daxx* knockout cells (Figure 2.18D).

In summary, *Daxx* knockout cells phenocopy the impaired silencing kinetics of *Atrx* knockout cells. Because depletion of *Daxx* does not alter the expression of *Trim28* or *Atrx*, the Daxx protein seems to be involved in the process.

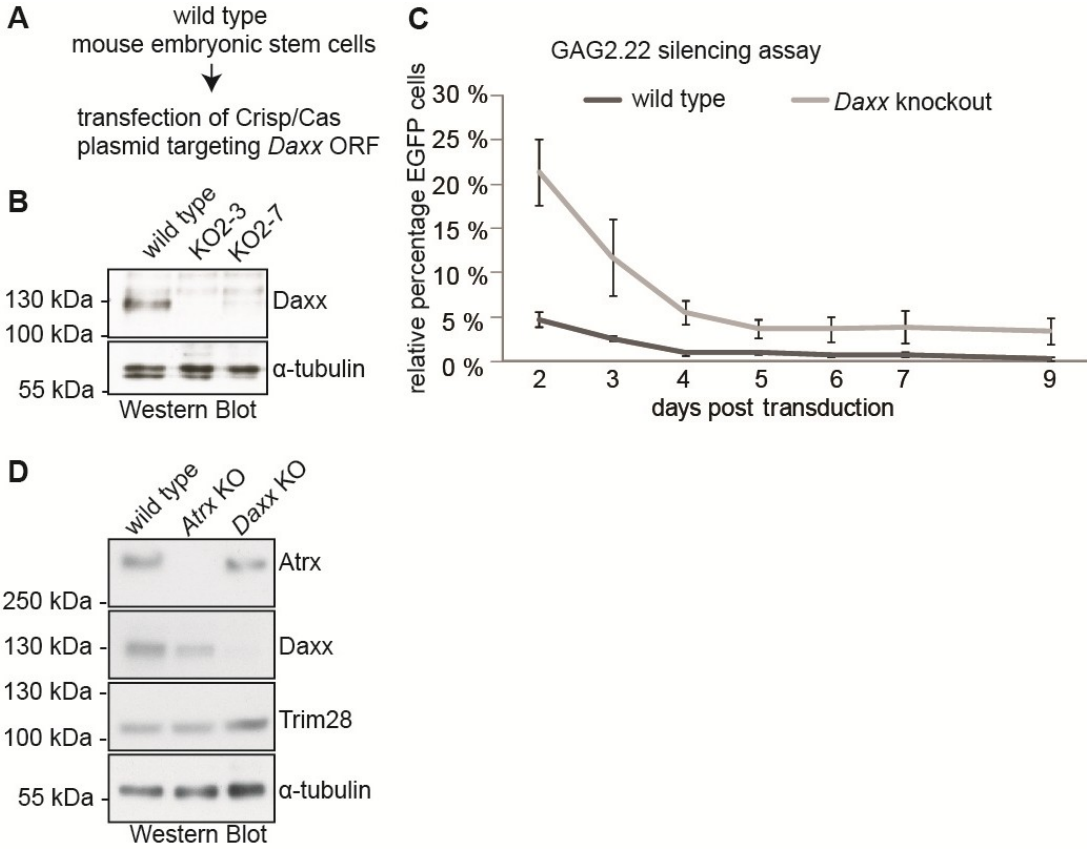


Figure 2.18 *Daxx* knockout cells phenocopy the impaired silencing kinetics of Atrx knockout cells
A) Mouse ES cells were transiently transfected with CRISPR/Cas9 plasmids targeting the *Daxx* ORF. In total four cell clones were isolated harboring frameshift mutations on both *Daxx* alleles. B) A representative western blot showing the loss of Daxx protein in two of the generated *Daxx* knockout cell lines. α-tubulin was used as a loading control. C) *Daxx* knockout cells were transduced with the GAG2.22-EGFP reporter as described in Figure 2.9. and the silencing of the reporter was monitored at the indicated time points using FACS. The plot represents three experiments of four independent *Daxx* knockout cell clones and the parental wild type cell line. Error bars represent the standard deviation. D) Western Blot for testing expression levels of Atrx and Trim28 in different genetic backgrounds. Trim28 is not strongly deregulated when Atrx or Daxx is depleted from the cells. *Daxx* knockout cells show normal Atrx levels. α-tubulin served as a loading control.

2.19. H3.3 interaction mutants of Daxx and wild type Daxx protein rescue the silencing defect of *Daxx* knockout cells

Daxx knockout cells and *Atrx* knockout cells show impaired silencing of the GAG2.22-EGFP reporter, which is a model system of IAP retrotransposon silencing (Figure 2.15, Figure 2.18). However, how one of the proteins on the molecular level participates in this process is unclear. Daxx has been shown to act as a histone chaperone for H3.3/H4 tetramer (Elsasser et al., 2012; Liu et al., 2012). In particular, arginine 257 of Daxx has been shown to be essential for binding of Daxx to the H3.3/H4 tetramer and glutamate 231 was identified to ensure that Daxx has selectivity for H3.3 rather than H3.1 (Elsasser et al., 2012). In addition, the C-terminus of Daxx has been shown to contain a Sumo-interaction motif that is essential for Daxx to act as a corepressor (Chang et al., 2005; Chang et al., 2011; Kuo et al., 2005; Lin et al., 2004; Lin et al., 2006). Therefore, I wondered, whether the histone chaperone activity of Daxx or its Sumo-interaction motif is important for its repressive function on the GAG2.22-EGFP reporter assay.

Thus, wild type *Daxx* cDNA and mutant *Daxx* cDNA were reintroduced into *Daxx* knockout cells to analyze, if the Daxx proteins harboring different mutations would rescue the reporter gene silencing defect (Figure 2.19A). After generation of stable cell lines, the GAG2.22-EGFP reporter was transduced and the silencing capabilities were measured after two days. *Daxx* knockout cells only transduced with an empty vector performed like *Daxx* knockout cells as shown in Figure 2.18 and showed a defect in GAG2.22-EGFP reporter gene silencing (Figure 2.19B). However, reintroduction of wild type *Daxx* cDNA and Daxx mutants carrying mutations E231A and R257A partially rescued the silencing phenotype in these cells, but reintroduction of Daxx lacking the Sumo-interaction motif (SIM) at its C-Terminus (amino acids 732 to 739) did not (Figure 2.19B). This indicates that H3.3 binding by Daxx or its binding specificity for H3.3 are not required for the repressive function of Daxx. In addition, the Sumo-interaction motif (SIM) seems to be required for the silencing activity of Daxx. Importantly, all Daxx proteins were overexpressed to a similar extent, ruling out the possibility that the SIM-mutant of Daxx was simply not expressed (Figure 2.19C).

In summary, Daxx mutants carrying mutations that affect H3.3 binding or H3.3 binding specificity can rescue IAP-GAG reporter silencing to the same extend as the wild type protein. However, Daxx carrying a deletion at its C-terminal Sumo-interaction motif did not rescue reporter gene silencing.

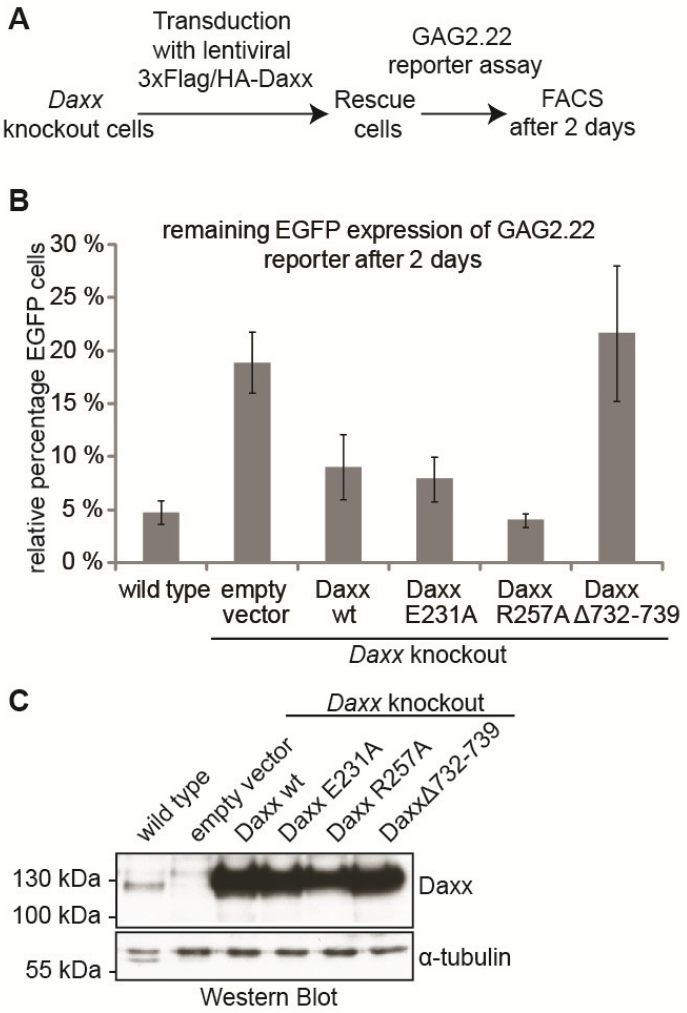


Figure 2.19 Wild type Daxx protein and H3.3 interaction mutants of Daxx rescue the silencing defect of *Daxx* knockout cells
A) *Daxx* knockout cells were stably transduced with Daxx, a Daxx mutant that has impaired selectivity between H3.3 and H3.1 (E231A), a Daxx mutant that has been shown to lose interaction with H3.3 (R257A) and a Daxx mutant lacking the C-Terminus carrying a phosphorylation and Sumo-interaction site (Elsasser et al., 2012; Lin et al., 2006). B) Rescue cells were transduced with the GAG2.22-EGFP reporter as described in Figure 2.9 and the silencing of the reporter was monitored two days after transduction using FACS. All Daxx constructs except the C-Terminal truncation mutant were able to rescue the silencing defect. The barplot represents an average of three experiments performed for the E231A and the R257A rescue cells and six experiments of the other cell lines. Error bars indicate the standard deviation. C) Daxx expression levels in the stable rescue cell lines indicating a strong overexpression of the transgene. α-tubulin was used as a loading control.

2.20. Association of Daxx with H3.3 is dependent on arginine 257 and the Daxx C-terminus

Daxx mutants carrying mutations in arginine 257 or glutamate 231 can rescue the silencing defect of *Daxx* knockout cells in reporter gene assays (Figure 2.19). This indicates that binding to H3.3 or binding specificity for H3.3 are both dispensable for the repressive function of Daxx. However, the finding that these two mutations affect histone H3.3 binding and binding specificity to H3.3 originates from analysis of human Daxx and crystallization

experiments of N-terminal Daxx truncations (Elsasser et al., 2012). Therefore, co-immunoprecipitations were carried out to verify, whether the generated mutations of the mouse Daxx protein really affect H3.3 binding.

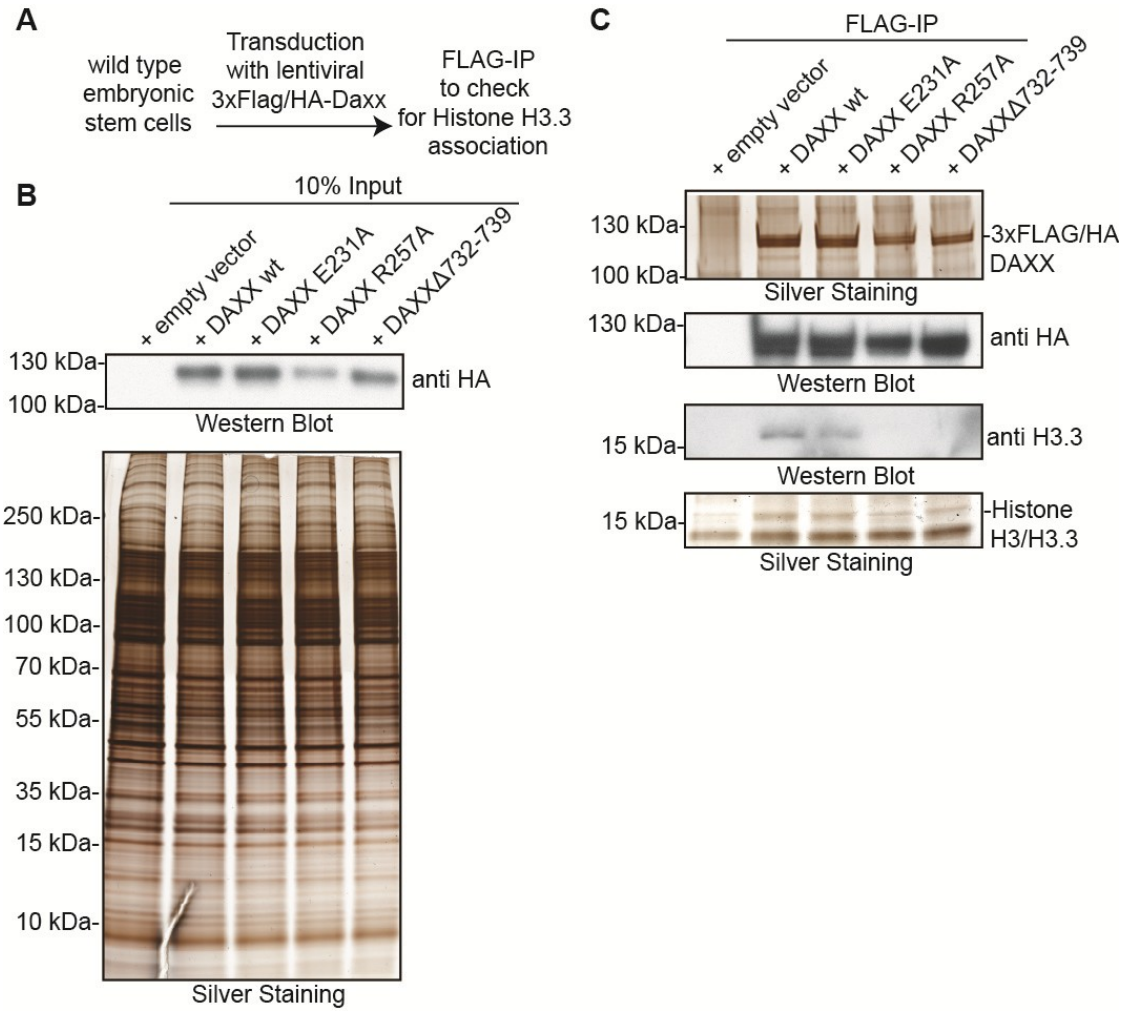


Figure 2.20 Association of Daxx with H3.3 is dependent on arginine 257 and the Daxx C-terminus
A) 3xFLAG/HA tagged wild type Daxx and Daxx mutants were stably transduced into wild type mES cells and nuclear extracts were used for FLAG-IP. B) Western blot showing the expression levels of the overexpressed proteins. The silver staining serves as a loading control. C) High amounts of Daxx protein can be immunoprecipitated by FLAG-IP. H3.3 only co-precipitates with wild type Daxx and the glutamate 231 to alanine mutant. H3.3 can also be seen as faint bands in the silver staining of the IP.

I generated wild type cell lines that overexpressed wild type 3xFLAG/HA-tagged Daxx or different Daxx mutants (Figure 2.20A). FLAG-CoIPs with these different Daxx overexpression cell lines were performed and histone H3.3 binding was analyzed via western blotting and silver staining. Even though Daxx mutants were slightly differently expressed in the individual cell lines (Figure 2.20B), similar amounts could be bound to FLAG-agarose beads, as estimated from the amount of bound Daxx protein via silver staining (Figure 2.20C). However, only wild type Daxx and Daxx E231A showed binding to histone H3.3 (Figure 2.20C). Please note that the binding of histone H3.3 is slightly reduced in the Daxx

E231A mutant. This is consistent with a loss of binding selectivity for histone H3.3 over histone H3.1 in this mutant (Elsasser et al., 2012). Interestingly, the C-terminal truncation of Daxx lacking the Sumo-interaction motif, is also defective in histone H3.3-binding (Figure 2.20C).

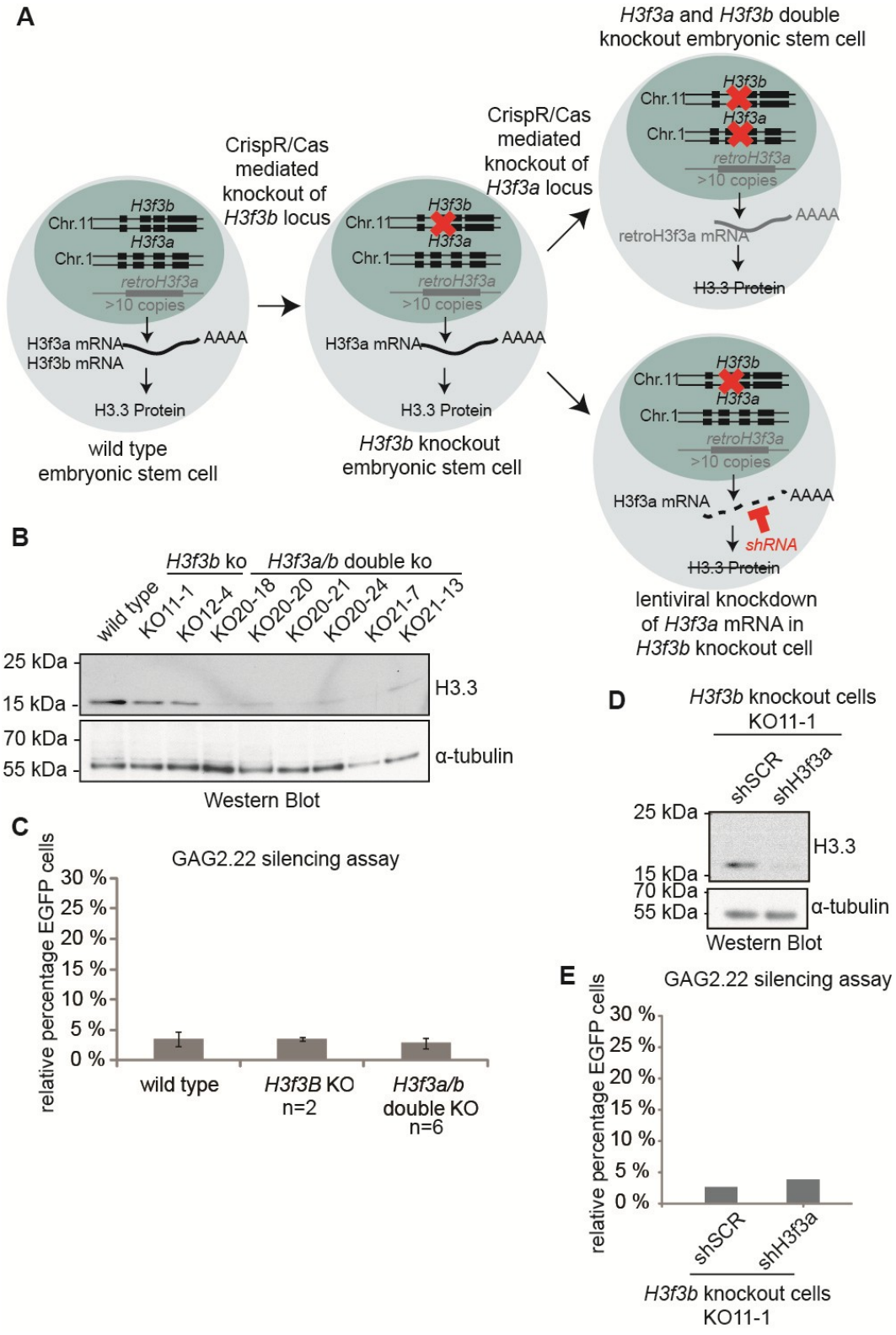
In summary, the histone H3.3-binding mutant of Daxx (R257A) and Daxx lacking the C-terminal Sumo-interaction motif fail to bind to histone H3.3 *in vitro.* The Daxx mutant that has been shown to affect H3.3-binding selectivity of Daxx (E231A) shows slightly reduced H3.3-binding.

2.21. H3.3 depleted cells show no change in GAG2.22 reporter gene silencing

Daxx mutants which do not bind to histone H3.3 or which are not selective for H3.3-binding rescued the reporter gene silencing defect of *Daxx* knockout cells (Figure 2.20, Figure 2.19). This suggests that histone H3.3-binding by Daxx is not required for Daxx to act as a co-repressor on IAP-GAG-EGFP reporter genes. However, the molecular function of Daxx has recently been described as being a histone H3.3 chaperone and impaired binding of H3.3 *in vitro* does not imply that H3.3 is not an important factor for the repressive activity of Daxx *in vivo.*

Histone H3.3 deficient mES cells were generated to investigate whether incorporation of histone H3.3 has indeed no influence on IAP-GAG reporter gene silencing. The histone H3.3 protein is encoded by two independent autosomal loci in mice, *H3f3a* and *H3f3b.* Furthermore the mouse genome contains a large number of promoter-less retrotransposed copies of the histone *H3f3a* gene, that in many cases carry point mutations. Depletion of H3.3 from mES cells was carried out in two consecutive steps.

Figure 2.21 H3.3 depleted cells show no change in GAG2.22 reporter gene silencing A) Generation of H3.3 depleted cells. In mice H3.3 is encoded by two genes *H3f3a* and *H3f3b*. In addition the mouse genome encodes additional retrotransposed copies of *H3f3a* that lack promoter sequences and frequently carry point mutations. *H3f3b* was deleted in the cells by transfection of a CRISPR/Cas9 plasmid targeting the 5' region of the *H3f3b* open reading frame to generate frameshift mutations on both alleles. Mutated cell clones were identified by showing a strong reduction of *H3f3b* mRNA-level using RT-qPCR most likely due to frameshift mutations resulting in non-sense-mediated mRNA decay (data not shown). *H3f3b* knockout cell clones were then either transfected with two CRISPR/Cas9 plasmids removing the first coding exon of the gene or transduced with an shRNA targeting *H3f3a*. B) H3.3 protein levels are massively reduced upon mutating *H3f3b* and *H3f3a* using CRISPR/Cas9 genome editing. Alpha-tubulin serves as a loading control. C) *H3f3a/b* double knockout and *H3f3b* single knockout cells show no defect in silencing the GAG2.22-EGFP reporter gene. The silencing assay was carried out as described in Figure 2.9. The number of different cell lines used for the analysis is indicated below the barplot. Error bars indicate the standard deviation. D) Western blot showing H3.3 depletion upon lentiviral knockdown of *H3f3a* in *H3f3b* knockout cells. α-tubulin serves as a loading control. E) The barplot shows the same experiment as described in C) except for using *H3f3b* knockout cells after *H3f3a* knockdown. Again, no strong derepression of the IAP-GAG2.22-EGFP reporter could be observed in H3.3 depleted cells.


A
H3f3a and H3f3b double knockout embryonic stem cell
CrispR/Cas mediated knockout of H3f3b locus
CrispR/Cas mediated knockout of H3f3a locus
H3f3b
Chr.11
H3f3a
Chr.1
retroH3f3a
>10 copies
H3f3a mRNA
H3f3b mRNA
AAAA
retroH3f3a mRNA
H3.3 Protein
wild type embryonic stem cell
H3f3b knockout embryonic stem cell
shRNA
lentiviral knockdown of H3f3a mRNA in H3f3b knockout cell
B
H3f3b ko
H3f3a/b double ko
wild type
KO11-1
KO12-4
KO20-18
KO20-20
KO20-21
KO20-24
KO21-7
KO21-13
25 kDa
15 kDa
70 kDa
55 kDa
H3.3
α-tubulin
Western Blot
C
GAG2.22 silencing assay
relative percentage EGFP cells
30 %
25 %
20 %
15 %
10 %
5 %
0 %
wild type
H3f3B KO n=2
H3f3a/b double KO n=6
D
H3f3b knockout cells KO11-1
shSCR
shH3f3a
E
GAG2.22 silencing assay
H3f3b knockout cells KO11-1

First, the *H3f3b* locus was mutated via RNA-guided nucleases (Figure 2.21A). Cell clones carrying mutations on both *H3f3b* alleles were identified by a strong reduction of *H3f3b* mRNA, most likely occurring through non-sense-mediated RNA decay (data not shown). Next, *H3f3b* knockout cell lines were used to either mutate the *H3f3a* gene locus or to knockdown *H3f3a* transcripts via lentiviral shRNAs (Figure 2.21A). Mutating the *H3f3a* ORF using RNA-guided nucleases was not possible, due to a high number of homologues retrotransposed copies of *H3f3a* in the mouse genome. Therefore, the first coding exon of *H3f3a* was completely removed by two RNA-guided nucleases targeting the flanking introns of this exon. Depletion of H3.3 protein in *H3f3b* knockout cells either by RNA-guided nucleases or by RNAi was monitored using western blot (Figure 2.21B and D). However, a very small signal of H3.3 protein was still detectable in these cells, probably due to weak expression of retrotransposed *H3f3a* copies, insufficient specificity of the H3.3 antibody or incomplete knockdown (Figure 2.21B and D). Importantly, even though H3.3 protein level was largely reduced in these cells, no change on reporter gene silencing of the GAG2.22-EGFP reporter could be observed (Figure 2.21C and E).

In summary, depletion of histone H3.3 protein from mES cells has no influence on silencing the GAG2.22-EGFP reporter cassettes that mimics IAP retrotransposon silencing. This finding is consistent with the observation that Daxx mutants deficient in H3.3-binding can still rescue the silencing defect of *Daxx* knockout cells (Figure 2.19).

2.22. *Atrx* is important for efficient *Trim28*- and *Setdb1*-dependent silencing at endogenous IAP elements

Trim28 and *Setdb1* depletion leads to a strong derepression of endogenous retroviral elements including IAP retrotransposons (Matsui et al., 2010; Rowe et al., 2010). Silencing of IAP retrotransposons originates to a great extent from the GAG2.22 sequence element (Figure 2.8, Figure 2.10). A generated lentiviral reporter system based on reporter gene silencing of this IAP-GAG sequence element phenocopies important properties of endogenous IAP silencing, including *Setdb1*- and *Trim28*-dependency and restriction of the silencing to embryonic stem cells. In addition, I found *Atrx* and *Daxx* to catalyze the silencing of the GAG2.22-EGFP reporter within a short time-frame (Figure 2.14, Figure 2.18). However, it is unknown, whether *Atrx* or *Daxx* deficiency also leads to derepression of endogenous IAP elements.

Therefore, endogenous IAP expression of *Atrx* and *Daxx* knockout cells was analyzed via RT-qPCR. IAP expression was not found to be deregulated in these cells, indicating that *Atrx* and *Daxx* are dispensable for the maintenance of silencing of these elements (Figure 2.22A). However, GAG2.22 reporter experiments *in vitro* suggested that the main function of Atrx

and Daxx is to catalyze the silencing process (Figure 2.15 and Figure 2.18). This argues that they are not absolutely required for the silencing process but are important for a high silencing efficiency. Thus, I sought to analyze transcript levels of IAP retrotransposons upon knockdown of *Trim28* to test whether an impaired silencing efficiency at IAP retrotransposons can be observed when Atrx is depleted.

Atrx knockout mES cells show a higher derepression of endogenous IAP retrotransposons than wild type cells when *Trim28* is depleted (Figure 2.22B). This suggests that *Atrx* might indeed catalyze heterochromatic gene silencing. Similarly, treating the cells with 5-azacytidine, a compound that is known to inhibit DNA methylation and to increase DNA damage at DNA-methylated regions (Christman, 2002), leads to a stronger upregulation of endogenous IAP transcripts in *Atrx* depleted cells (Figure2.22C).

This suggests that endogenous IAP elements are more sensitive to inhibition of DNA-methylation by DNA damage-inducing agents in *Atrx* knockout cells. Interestingly, also silencing of the GAG-EGFP reporter is further impaired in *Atrx* knockout cells, when *Trim28* or *Setdb1* are depleted (Figure 2.22D). This indicates that the rate-limiting concentration of Setdb1 and Trim28 required for efficient reporter gene repression is higher in *Atrx* knockout cells.

In summary, *Atrx* seems to increase the efficiency of *Trim28*- and *Setdb1*-dependent silencing at endogenous IAP elements and exogenous reporters. This implies that *Atrx* depletion leads to higher plasticity of heterochromatin.

2.23. *Atrx* depletion leads to a higher heterochromatin plasticity when heterochromatin is challenged by activating factors

If *Atrx* increases the efficiency of *Trim28*- and *Setdb1*-dependent heterochromatic silencing, *Trim28*- and *Setdb1*-dependent heterochromatin will have a higher plasticity in the absence of Atrx. Higher heterochromatin plasticity might result in an easier destabilization of heterochromatin, if strong activating factors or transcription challenge the heterochromatic state.

I induced strong transcription over the heterochromatinized reporter locus generated in Figure 2.10 to test whether Atrx is required when strong activating factors and transcription challenge the heterochromatic state (Figure 2.23A). For this, a reverse tet-transactivator was stably introduced into the cells carrying an EGFP-GAG2.22 locus under the control of an inactive tet-response element (Figure 2.10). Transcription at this reporter locus was induced

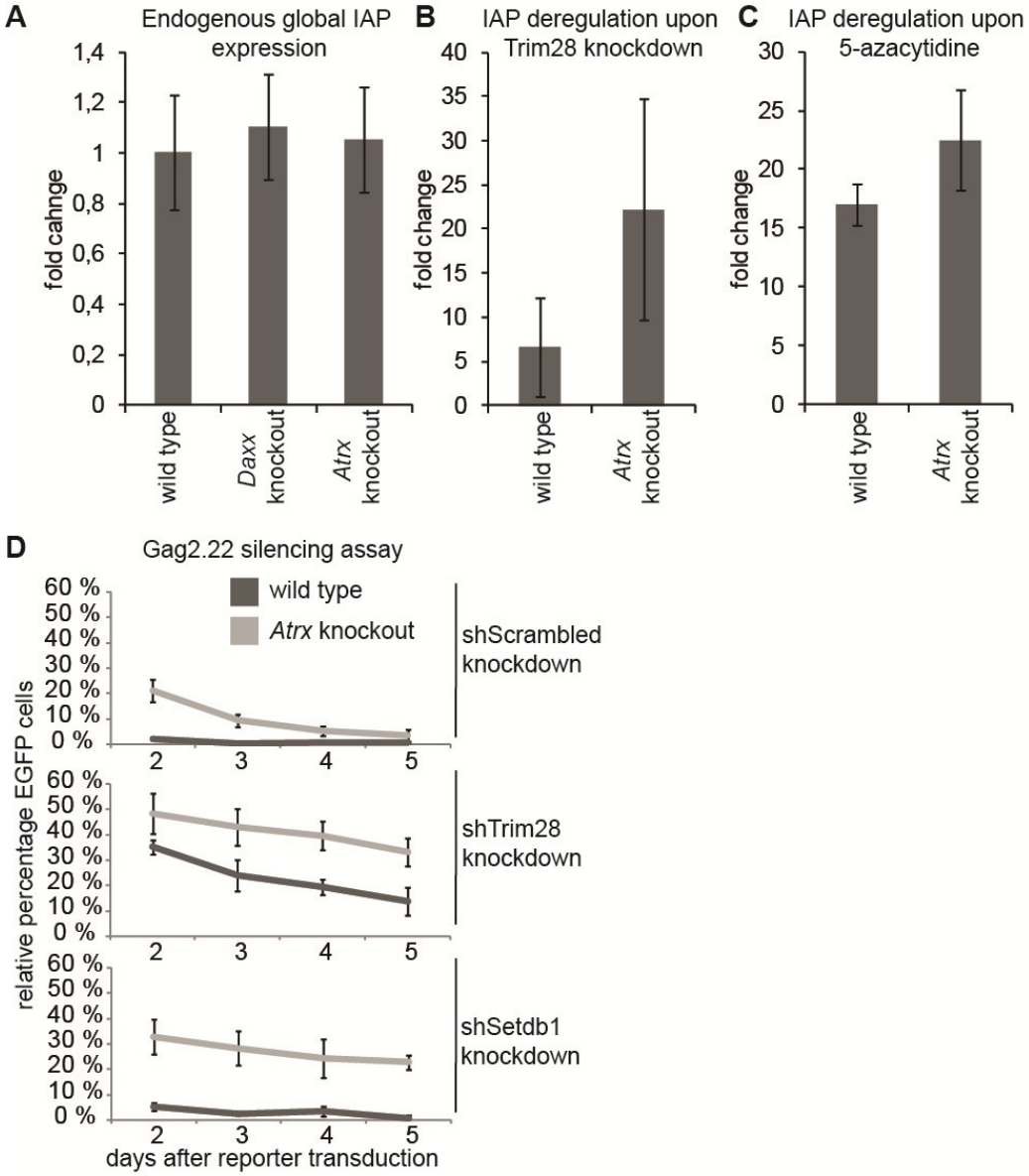


Figure 2.22 *Atrx* is important for efficient Trim28- and Setdb1-dependent silencing at endogenous IAP elements
A) Endogenous IAP expression levels are not changed in *Daxx* or *Atrx* knockout cells. RNA of wild type, *Daxx* and *Atrx* knockout mouse ES cells was reverse transcribed and IAP expression was analyzed using qRT-PCR. The barplot represents the average of three to eight biological replicates (wild type n=6, Atrx ko n=8, Daxx ko n=3). Error bars represent the standard deviation. B) *Trim28* depletion leads to a stronger upregulation of endogenous IAP transcripts in *Atrx* knockout cells. Wild type and *Atrx* knockout mES cells were transduced with shRNAs against *Trim28* and endogenous IAP transcripts were analyzed as described in A). The barplot represents three experiments using three different shRNAs against *Trim28.* Error bars indicate the standard error of the mean. C) *Atrx* knockout cells show higher upregulation of endogenous IAP elements upon treatment with 5-azacytidine. Cells were treated for 24 h with 7µM 5-azacytidine or DMSO as a control. Endogenous IAP expression was measured using RT-qPCR. Upregulation of endogenous IAP transcripts upon 5-azacytidine treatment was compared to DMSO treated cells. Error bars indicate the standard error of the mean. Wild type samples n=4, *Atrx* knockout samples n=8 Unpaired two-sided t-test p-value = 0,017 D) *Trim28* and *Setdb1* depletion leads to an increased silencing defect in *Atrx* knockout cells. Wild typed and *Atrx* knockout mES cells were transduced with shRNAs against *Trim28*, *Setdb1* and a scrambled control sequence. Four days after shRNA transduction the cells were transduced with the GAG2.22-EGFP reporter virus. Silencing potential of the GAG2.22 sequence was analyzed in FACS as described in Figure 2.9. The plot corresponds to three independent experiments. Error bars represent the standard deviation. The experiment in D) has been carried out by Katharina Schmidt (Schmidt, 2014).

by adding 1 µg/ml doxycyclin and activation of the reporter gene was monitored by EGFP expression. In addition, cells were treated with shRNAs targeting Atrx or a non-silencing scrambled control shRNA. Only few control knockdown cells (< 5 %) could reactivate the reporter gene upon doxycyclin addition, probably due to its heterochromatinization (Figure 2.23B and Figure 2.10B). However, a much higher number of cells could reactivate the reporter locus in *Atrx* knockdown cells (Figure 2.23B).

In summary, a heterochromatinized reporter locus can be easier reactivated by transactivating factors when *Atrx* is depleted from the cells. This suggests that, Atrx is important to ensure low heterochromatin plasticity.

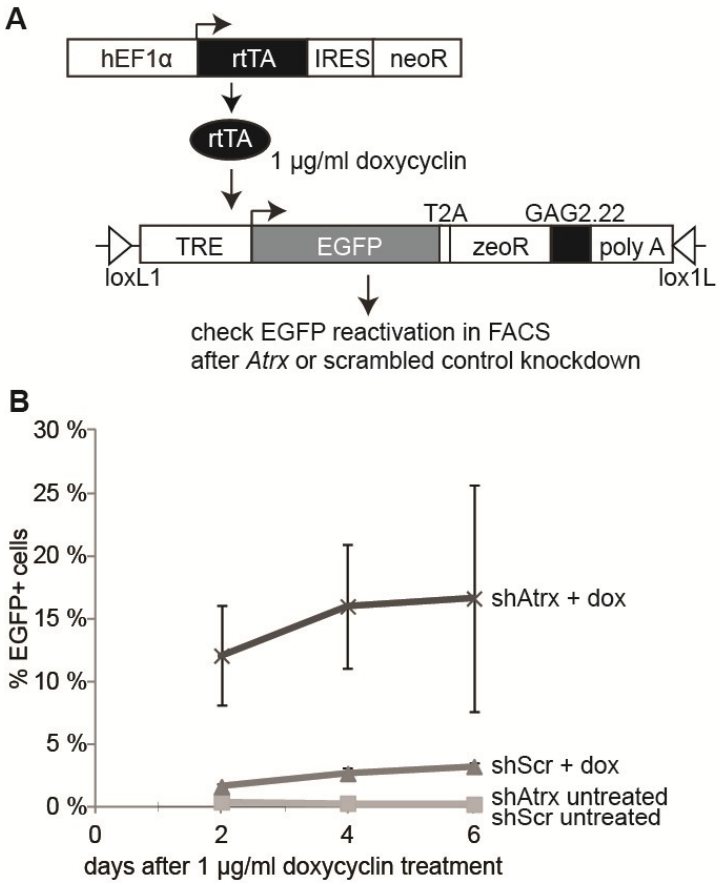


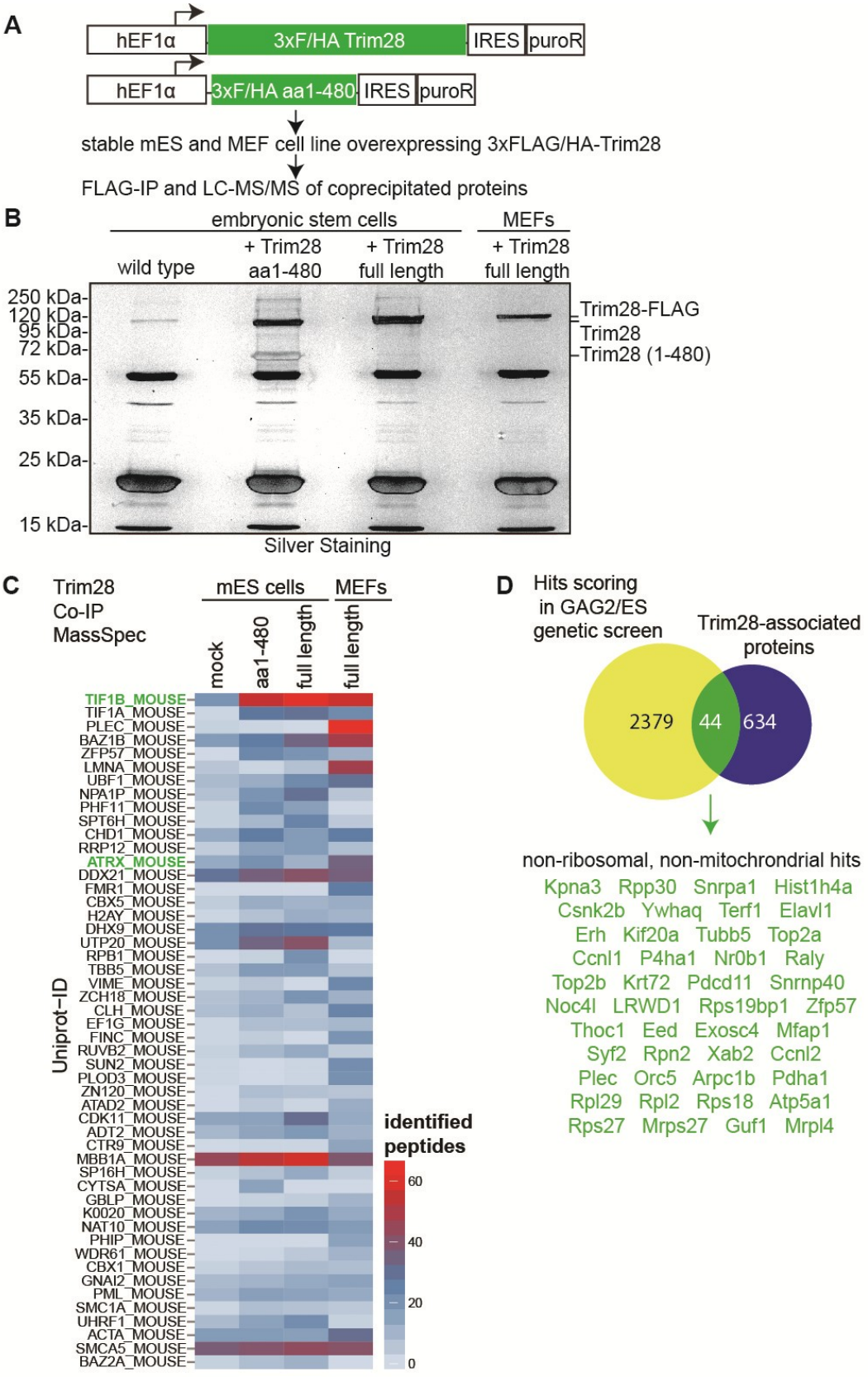
Figure 2.23 Atrx depletion leads to a higher heterochromatin plasticity when heterochromatin is challenged by activating factors
A) HA36 mouse ES cells carrying a stable integration of an inducible EGFP-GAG2.22 cassette were stably transduced with a doxycyclin-dependent reverse tet-transactivator that allows the induction of transcription over the EGFP-GAG2.22 cassette (also see Figure 2.10). B) Atrx depleted cells show a much stronger reactivation of the heterochromatinized EGFP-GAG2.22 cassette. Cells were either transduced with an shRNA targeting Atrx or an shRNA targeting a scrambled control sequence. Four days after shRNA transduction, the transcription over the EGFP-GAG2.22 transgene was induced using 1 µg/ml doxycycline and the EGFP expression of the cells was monitored using FACS. The plot represents an average of three experiments (for day six n=2). Error bars represent the standard deviation.

2.24. Trim28 is associated with Atrx and other known heterochromatin proteins

Trim28 and *Setdb1* are required for silencing of IAP retrotransposons and the GAG2.22-EGFP reporter gene ((Matsui et al., 2010; Rowe et al., 2010), Figure 2.9, Figure 2.11). In this thesis it was found, that depletion of *Atrx* and *Daxx* decelerates heterochromatic silencing of a GAG2.22-EGFP reporter gene that mirrors many aspects of endogenous IAP silencing (Figure 2.14, Figure 2.18). In addition, Atrx ensures low heterochromatin plasticity at endogenous IAP retrotransposon sequences and reporter loci (Figure 2.22, Figure 2.23). Yet, how Atrx or Daxx cooperate with Trim28 or Setdb1 on protein level to ensure rapid heterochromatic silencing is unknown and it is very likely that additional proteins are involved in this process. Since Trim28 is probably a factor that acts in the initiation of heterochromatic silencing, I wondered which proteins interact with Trim28. Hence, Trim28-interacting proteins might be involved in the silencing process of IAP retrotransposons. Moreover, Trim28 associates with KRAB zinc finger proteins, a large family of mammalian transcription factors, which are capable of directly recruiting Trim28 to genomic regions (Bellefroid et al., 1991; Friedman et al., 1996; Kim et al., 1996; Moosmann et al., 1996; Schultz et al., 2001). Consequently, a KRAB Zinc finger protein or other Trim28-associated proteins might be involved in recognition of the GAG2.22 sequence of IAP retrotransposons. Therefore, I screened for proteins associated with Trim28 using co-immunoprecipitation experiments followed by mass-spectrometry analysis.

First, mouse cell lines were generated which overexpress 3xFLAG/HA-tagged Trim28 or the N-terminal RBCC domain of Trim28 (Figure 2.24A). This was done, to achieve a higher amount of immunoprecipitated material for mass-spectrometry.

Figure 2.24 Trim28 is associated with Atrx and other known heterochromatin proteins A) Experimental design of the Trim28 pull-down experiment. Mouse ES cells and mouse embryonic fibroblasts where stably transduced with lentiviral vectors encoding either triple FLAG/HA-tagged Trim28 or a truncation of Trim28 encoding the N-terminal RBCC domain (aa1-480). Stable cell lines where harvested for FLAG-co-immunoprecipitation and Trim28-associated proteins were identified using liquid chromatography coupled tandem-mass spectrometry (LC-MS/MS). B) Silver staining of SDS-PAGE gel showing eluted proteins after FLAG-immunoprecipitation. The prominent bands at 55 and 20 kDa correspond to the FLAG-antibody chains. The wild type sample in the first lane contains no FLAG-tagged protein and serves as a negative control for proteins nonspecifically associated to the FLAG-beads. The RBCC domain (aa1-480) of Trim28 co-precipitates the endogenous full-length Trim28 protein (second lane). Trim28 protein is strongly stained at 110 kDa (indicated on the right of the gel). C) Heatmap showing the 50 proteins most prominently associated with Trim28. The number of peptides were identified using Scaffold3™. Proteins were sorted by the difference of the average number of identified peptides among all three IPs compared to the wild type control IP. The Uniprot-ID of the strongest hit "TIF1B_MOUSE" corresponds to Trim28 protein and that Atrx is identified as a Trim28-interacting protein. The complete list of Trim28-interacting proteins can be found in the appendix. The mass-spectrometry after in-gel digestion and the peptide search using Scaffold3™ were performed by Dr. Ignasi Forné. D) Hits from the genetic GAG2/mES IAP silencing screen were analyzed for Trim28-associated proteins. In total 44 proteins were found of which 35 are not mitochondrial or ribosomal proteins.


A
hEF1α
3xF/HA Trim28
IRES
puroR
hEF1α
3xF/HA aa1-480
IRES
puroR
stable mES and MEF cell line overexpressing 3xFLAG/HA-Trim28
FLAG-IP and LC-MS/MS of coprecipitated proteins
B
embryonic stem cells
MEFs
wild type
+ Trim28 aa1-480
+ Trim28 full length
+ Trim28 full length
250 kDa-
120 kDa-
95 kDa-
72 kDa-
55 kDa-
35 kDa-
25 kDa-
15 kDa-
Trim28-FLAG
Trim28
Trim28 (1-480)
Silver Staining
C
Trim28 Co-IP MassSpec
mES cells
MEFs
mock
aa1-480
full length
full length
Uniprot-ID
TIF1B_MOUSE
TIF1A_MOUSE
PLEC_MOUSE
BAZ1B_MOUSE
ZFP57_MOUSE
LMNA_MOUSE
UBF1_MOUSE
NPA1P_MOUSE
PHF11_MOUSE
SPT6H_MOUSE
CHD1_MOUSE
RRP12_MOUSE
ATRX_MOUSE
DDX21_MOUSE
FMR1_MOUSE
CBX5_MOUSE
H2AY_MOUSE
DHX9_MOUSE
UTP20_MOUSE
RPB1_MOUSE
TBB5_MOUSE
VIME_MOUSE
ZCH18_MOUSE
CLH_MOUSE
EF1G_MOUSE
FINC_MOUSE
RUVB2_MOUSE
SUN2_MOUSE
PLOD3_MOUSE
ZN120_MOUSE
ATAD2_MOUSE
CDK11_MOUSE
ADT2_MOUSE
CTR9_MOUSE
MBB1A_MOUSE
SP16H_MOUSE
CYTSA_MOUSE
GBLP_MOUSE
K0020_MOUSE
NAT10_MOUSE
PHIP_MOUSE
WDR61_MOUSE
CBX1_MOUSE
GNAI2_MOUSE
PML_MOUSE
SMC1A_MOUSE
UHRF1_MOUSE
ACTA_MOUSE
SMCA5_MOUSE
BAZ2A_MOUSE
identified peptides
60
40
20
0
D
Hits scoring in GAG2/ES genetic screen
Trim28-associated proteins
2379
44
634
non-ribosomal, non-mitochrondrial hits
Kpna3 Rpp30 Snrpa1 Hist1h4a
Csnk2b Ywhaq Terf1 Elavl1
Erh Kif20a Tubb5 Top2a
Ccnl1 P4ha1 Nr0b1 Raly
Top2b Krt72 Pdcd11 Snrnp40
Noc4l LRWD1 Rps19bp1 Zfp57
Thoc1 Eed Exosc4 Mfap1
Syf2 Rpn2 Xab2 Ccnl2
Plec Orc5 Arpc1b Pdha1
Rpl29 Rpl2 Rps18 Atp5a1
Rps27 Mrps27 Guf1 Mrpl4

Proteins associated with 3xFLAG/HA-Trim28 were eluted from FLAG-beads after IP with SDS loading buffer and separated on an SDS page for silver staining (Figure 2.24B). Please note that endogenous Trim28 full-length protein seems to co-precipitate with the FLAG-tagged N-terminal Trim28-RBCC domain (Figure 2.24B).This is consistent with previous reports that show Trim28 oligomerization (Peng et al., 2000a; Peng et al., 2000b; Peng et al., 2002). The full lane of separated proteins was cut into eight gel pieces. Proteins inside the gel-fragments were treated with DTT, cysteine residues were alkylated using iodoacetamide and proteins were in-gel digested using trypsin before mass-spectrometry. Peptides were identified using *Scaffold3*™ software.

A lot of known Trim28-binding proteins were consistently found to be associated in the immunoprecipitation experiments (Figure 2.24C), including Atrx, Hp1 proteins and KRAB domain containing proteins like Zfp57 (Quenneville et al., 2011; Ryan et al., 1999) (Figure 2.24). Most of them were associated in MEF and mouse ES cells (Figure 2.24C). The full list can be found in the Appendix.

Identification of Trim28-associated proteins could confirm known Trim28-interaction partners and could identify novel proteins that might be functionally involved in Trim28-dependent silencing. Which of these proteins is functionally involved in Trim28-dependent silencing is unclear and might be a basis for future experiments. Candidate proteins could be hits that were identified in the genetic screening experiments and which are physically associated with Trim28 protein (Figure 2.24D). However, since the number of (false positive) hits identified in the genetic screen is presumably quite high and mass-spectrometry experiments usually suffer from a high number of false negative hits, additional experiments and secondary screens might further increase the confidence of the identified hits.

2.25. DNA pull-down experiment for GAG2.22-binding proteins using SILAC

Please note that the DNA pull-down experiments and the SILAC mass-spectrometry screening in this chapter have been performed by Dr. Falk Butter, who collaborated with me in this project.

GAG2.22, a 160 bp fragment of the GAG region of IAP retrotransposons silences EGFP reporter genes in mES cells (Figure 2.8, Figure 2.9). Reporter assays using the GAG2.22 sequence element could phenocopy properties of endogenous IAP retrotransposon silencing including Trim28- and Setdb1-dependency (Figure 2.9, Figure 2.11). Moreover, GAG2.22 reporter assays have been useful to identify novel players involved in IAP retrotransposon silencing like Daxx and Atrx (Figure 2.14, Figure 2.18). However, the way the GAG2.22 sequence is only recognized and silenced in mES cells is still elusive. Trim28-associated

proteins have been analyzed to identify factors that might be functionally important for silencing (Figure 2.24). Yet, this proteomic screen does not give any specific information which is the recognition factor for the GAG2.22 sequence.

A DNA pull-down assay was performed using the GAG2.22 sequence and control sequences to identify DNA-binding proteins that can specifically bind to the GAG2.22 sequence. Mouse ES cells grown in medium containing arginine and lysine either labeled with light or heavy isotopes were used to generate nuclear extracts. “Heavy” nuclear extracts were used for a DNA pull-down experiment using the GAG2.22 sequence and “light” nuclear extracts were used for a DNA pull-down experiment using the control sequence (Butter et al., 2010) (Figure 2.25A). Eluted proteins were pooled, trypsinized and the mass-ratios of the identified peptides were used to identify proteins specifically binding to one of the sequences (Figure 2.25A). The experiment was also performed by swapping “heavy” and “light” nuclear extract to increase the statistical reliability.

First, feasibility of the approach was tested in a control experiment to verify the known binding of the protein Zfp809 to its DNA-binding site. A point mutated Zfp809-binding site was used as a negative control sequence. Specific binding of Zfp809 and its associated protein Trim28 to the Zfp809-binding site could be detected (Figure 2.25B). This indicated that the experimental approach was sensitive and specific enough to unravel Protein-DNA-interactions at heterochromatic target sequences.

Next, proteins specifically binding to the GAG2.22 sequence flanked by additional 60 bp were identified using a 220 bp fragment of the *Spectinomycin*-resistance gene as a control sequence (Figure 2.25C). Since the number of differently binding proteins was high, a second approach was performed, searching for proteins binding to GAG2.22 but not to a point mutant sequence (Figure 2.25D). The point mutant sequence was identified by Katharina Schmidt (Schmidt, 2014).

The first approach identified numerous proteins binding to GAG2.22 that are ubiquitously expressed or not restricted to mES cells (Figure 2.25C). Since GAG2.22 silencing is limited to mES cells, it was not very likely that these proteins are the cell type-specific recognition factors for the GAG2.22 sequence. However, the mouse specific KRAB zinc finger protein Zfp71-rs1 was also found to bind to the GAG2.22 sequence (Figure 2.25C). In a second screening approach were a GAG2.22 point mutant sequence was used as a control, the ubiquitous protein Ubtf was the only specifically enriched protein (Figure 2.25D). I generated knockdown cell lines for Ubtf and Zfp71-rs1, to verify whether these proteins have a functional role in silencing the GAG2.22 sequence element. Silencing of a GAG2.22-EGFP reporter gene was not impaired upon knockdown of either Ubtf or Zfp71-rs1 (Figure 2.25E).

In summary, DNA pull-downs can work as a powerful tool for identification of DNA-binding proteins. However, DNA-binding proteins that specifically recognize the GAG2.22 sequence element of IAP retrotransposons in mES cells could not be identified.

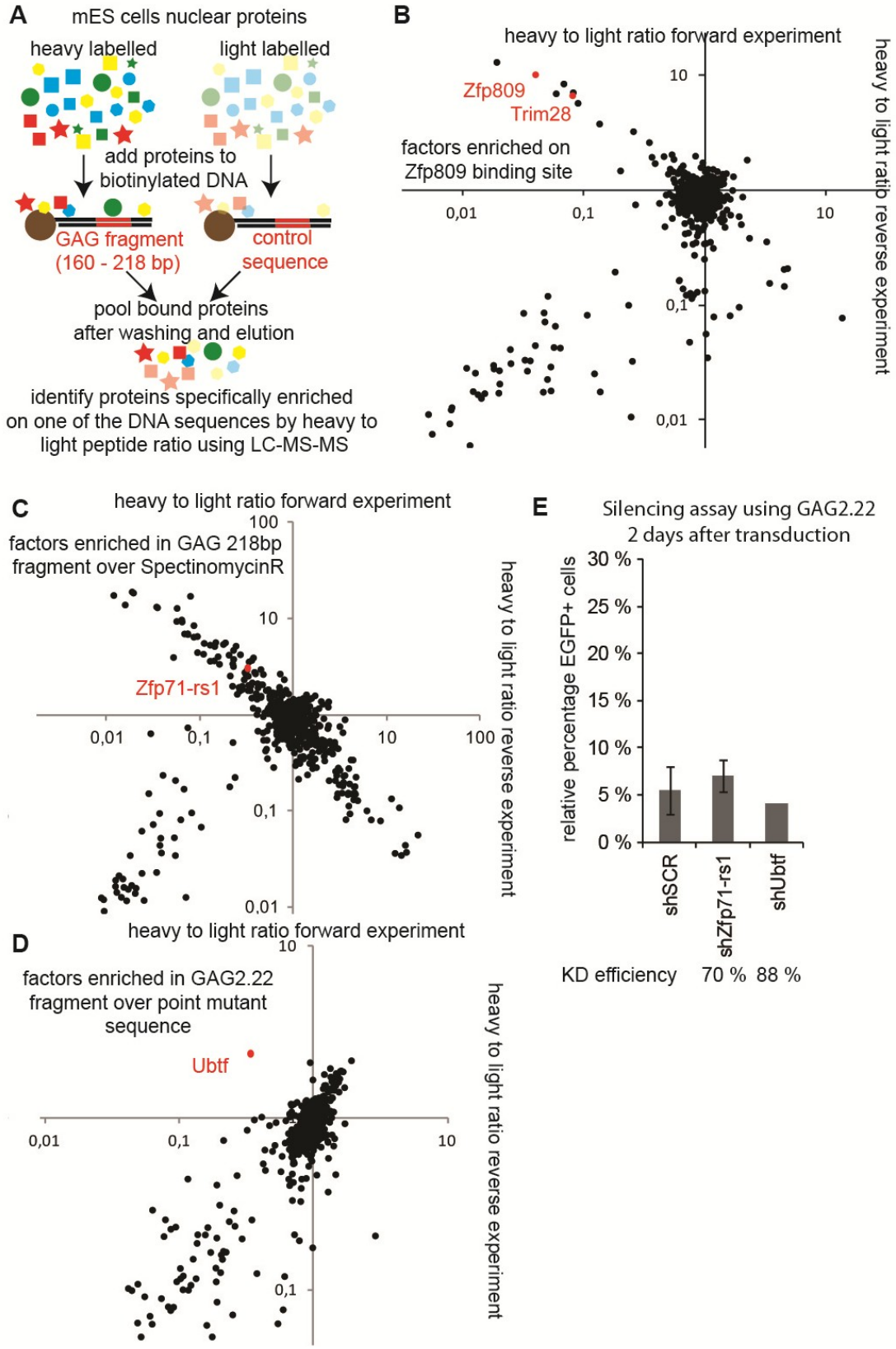


Figure 2.25 DNA pull-down experiment for GAG2.22-binding proteins using SILAC
A) Nuclear extracts isolated from mES cells labeled with either heavy or light isotopes are incubated with biotinylated DNA fragment immobilized on streptavidin beads. The bound proteins are pooled after elution and specific binding to one of the DNA sequences is detected by a change in the mass-ratio in mass-spectrometry. As a control the experiment is repeated with heavy and light nuclear extract swapped. B) A proof-of-principle experiment using a Zfp809-binding site and a control sequence where the binding site was mutated identifies Zfp809 and its binding partner Trim28. Enrichment is plotted as heavy to light ratio of identified peptides on the x-axis and the y-axis. In the reverse experiment the light and the heavy extract were swapped leading to reciprocal values of enrichment C) A slightly bigger GAG2.22 sequence (218 bp) was compared to a sequence fragment of the spectinomycin resistance gene of equal size in a first DNA pull-down experiment. Only factors in the top left corner are specifically enriched at the GAG2.22 sequence showing a heavy to light ratio >1 in the forward experiment and a heavy to light ratio <1 in the reverse experiment. D) The plot shows the result of the second experiment using the GAG2.22 sequence (160 bp) and a point mutant sequence as a control. E) Zfp71-rs1 was identified as a potential protein from experiment 1 for being a mouse-specific protein containing a KRAB domain that facilitates binding to Trim28. Ubtf was the only hit in experiment 2. Knockdown of either Zfp71-rs1 or Ubtf resulted in no significant changes of reporter gene silencing. Knockdown efficiencies on RNA level are indicated and were either measured via qRT-PCR or taken from the prevalidated TRC clone collection at Sigma Aldrich. The knockdown experiment for Zfp71-rs1 was performed two times and for Ubtf only once. Error bars indicate the standard deviation. SILAC pull-down experiments and data analysis was performed by Dr. Falk Butter.

3. Discussion

3.1. The pseudoautosomal exons of *midline1* gene are derepressed in male *Suv39h* and *Suv4-20h* knockout cells

In this thesis I tried to identify genes that are regulated by heterochromatin proteins to use them as reporter systems for monitoring heterochromatin status. I first focused on gene expression datasets generated from *Suv39h* and *Suv4-20h* knockout cells, because these enzymes act together in a consecutive pathway in pericentric heterochromatin formation (Schotta et al., 2004; Schotta et al., 2008).

The X-chromosomal *midline1* gene flanks the pseudoautosomal boundary in mouse (Palmer et al., 1997), and I found that its pseudoautosomal exons are consistently derepressed in *Suv39h* and *Suv4-20h* knockout cells (Figure 2.1). These pseudoautosomal *midline1* transcripts only arise in male *Suv39h* and *Suv4-20h* knockout cells and are oriented in the same direction as the full-length *midline1* transcript (Figure 2.2). This suggests that these transcripts might originate from the Y-chromosomal pseudoautosomal region.

Aggravatingly, I could confirm previous studies that indicate a high heterogeneity and polymorphic structure of the mouse pseudoautosomal region (Figure 2.2) (Kipling et al., 1996a; Kipling et al., 1996b; Palmer et al., 1997). This is due to the fact, that the pseudoautosomal region in mice is comparably short and crossing over events during male meiotic prophase I exclusively take place in this chromosomal region (Perry et al., 2001). Even though, the pseudoautosomal exons of *midline1* are present in multiple copies in the genomes of different cell lines, the strong derepression of these pseudoautosomal exons in *Suv39h* and *Suv4-20h* knockout cells does not correlate with the copy number of pseudoautosomal *midline1* exons (Figure 2.2). This indicates that Suv39h and Suv4-20h enzymes might indeed have a functional role in repressing transcripts from the pseudoautosomal region. In addition, ChIP sequencing studies revealed that the pseudoautosomal region is enriched for H3K9me3 and H4K20me3 in mouse ES cells (Figure 2.1) (Mikkelsen et al., 2007). Thus, the observed derepression of the pseudoautosomal *midline1* transcripts might be a consequence of an impaired heterochromatin structure of the Y-chromosome. The pseudoautosomal *midline1* transcripts might originate from spurious promoter sequences of the Y-chromosome that are normally silenced by Suv39h and Suv4-20h enzymes. The mouse Y-chromosome is also highly repetitive and enriched for H3K9-methylation, suggesting that H3K9me3 could be involved in repressing Y-chromosomal transcription (Peters et al., 2002).

Moreover, the pseudoautosomal region is evolutionary selected to facilitate recombination during male meiosis and its repetitive nature might require a dense heterochromatin structure to avoid recombination events in somatic cells. Interestingly, already in the early 20th century it was found that centromeric or pericentromeric DNA regions repress recombination events, and it was proposed that heterochromatin plays a role in repression of recombination (Beadle, 1932; Mahtani and Willard, 1998; Mather, 1939). Quantitative analysis of recombination events in *Drosophila* also indicate that impairment of the heterochromatin status directly increases the rate of recombination events at heterochromatic regions (Westphal and Reuter, 2002).

Apart from *midline1* no other genes were strongly and consistently deregulated in *Suv39h* and *Suv4-20h* knockout cells. This indicates that Suv39h and Suv4-20h might only play a minor role in gene regulation even though they have important roles in pericentric and telomeric chromatin (introduction 1.1.2). Consistently, *Suv39h* knockout cells show only slightly reduced levels of H3K9me3 outside of heterochromatin (Bulut-Karslioglu et al., 2012). However, Suv4-20h enzymes are involved in non-coding RNA-mediated repression of ribosomal DNA repeats and retrotransposon sequences (personal communication Prof. Ingrid Grummt – article in press), underlining that a direct role of Suv4-20h enzymes in transcriptional regulation cannot be ruled out.

Due to constant recombination events in male meiosis the pseudoautosomal region is extremely diverse between different mammalian species, which suggests that sequence-dependent features are not important for the general function of the pseudoautosomal region (Helena Mangs and Morris, 2007). In contrast to the short pseudoautosomal region in mice, human sex-chromosomes, for example, contain two pseudoautosomal regions, which encode almost 20 genes (Helena Mangs and Morris, 2007). How chromatin regulation is involved in the function of the pseudoautosomal region and consequently mammalian speciation will be an interesting topic in future research.

3.2. Setdb1, Trim28, Prdm6 and Hp1 proteins are transcriptional repressors

To identify heterochromatin proteins that can directly influence gene expression, I cloned a candidate-library of heterochromatin-associated proteins into a Gal4/UAS reporter system. I found that Hp1 proteins (Cbx1, Cbx3 and Cbx5), Setdb1, Trim28 and Prdm6 can strongly repress transcription from the constitutively active mouse PGK promoter (Figure 2.3).

Setdb1, Trim28 and Hp1 proteins are known to collaborate in transcriptional silencing and to recruit each other when tethered at reporter gene loci (1.3.1. and 1.3.2.) (Ayyanathan et al., 2003; Hathaway et al., 2012; Schultz et al., 2002; Sripathy et al., 2006).

Prdm6 has not been implicated into Setdb1- or Trim28-dependent processes, but has been described to be a tissue-specific repressor that is required for mouse development, for the development and function of the vascular system and for neurogenesis (Davis et al., 2006; Gewies et al., 2013; Kinameri et al., 2008; Wu et al., 2008). Prdm6 also has a SET-domain, which has been described to have no detectable methyltransferase activity (Davis et al., 2006). However, others have found that immunopurified Prdm6 shows H4K20-methylation activity *in vitro* (Wu et al., 2008). Interestingly, Prdm6 has also been found to associate with Hp1 proteins (Davis et al., 2006), which could explain its ability to repress reporter gene transcription (Figure 2.3). In addition, Hp1 has been found to bind to Suv4-20h H4K20-methyltransferases (Schotta et al., 2004), which would suggest that the observed H4K20me-activity of immunoprecipitated Prdm6 by Wu et al. is indirectly mediated by co-precipitating Suv4-20h enzymes (Wu et al., 2008).

However, I could not identify a strong silencing effect for other Hp1-associated proteins like Suv4-20h or Suv39h enzymes in luciferase assays (Figure 2.3). This was surprising, because Su(var)3-9, the *Drosophila* homologue of Suv39h, is a potent repressor in luciferase assays (Boeke et al., 2010). However, the repressive activity of Su(var)3-9 seems to be largely dependent on the co-recruitment of histone deacetylases rather than the SET-domain activity of Su(var)3-9 (Boeke et al., 2010). Possibly, not all heterochromatin proteins are capable of silencing the PGK promoter. Promoter-specific effects might therefore mask the transcriptional silencing potential of individual H3K9me3-associated proteins.

3.3. Establishment of an ecotropic, lentiviral EGFP-based silencing assay utilizing a binding site for Zfp809.

In this study, I established an EGFP-based silencing reporter assay which is stably integrated into the genome using ecotropic lentiviruses (Figure 2.4, 2.5 and 2.6). Similar systems have been generated earlier to monitor EGFP silencing by cis-acting DNA elements (Haas et al., 2003; Rowe et al., 2010). Stably integrated EGFP transgenes have also widely been used in chromatin research to study reporter silencing with single cell resolution (Akhtar et al., 2013; Ashe et al., 2008; Daxinger et al., 2013; Hathaway et al., 2012).

I picked EGFP as a reporter gene due to the brightness of fluorescence of the EGFP protein and its technical compatibility with available FACS devices, even though it is a poor choice for a silencing reporter due to its long half-live of 26 h (Corish and Tyler-Smith, 1999). However, in my assays the lentiviral EGFP transgene is silenced before full expression of the EGFP transgene is reached, which prevents an accumulation of large amounts of stable EGFP protein inside the cells (Figure 2.6). An improvement of the EGFP reporter system could be the use of fast-degrading EGFP-derivatives even though the lowered half-life of

these variants is accompanied with a fundamental decrease of detectable fluorescence (Corish and Tyler-Smith, 1999). Alternatively, *td-tomato* reporter genes have successfully been used by others to monitor fast fluctuations of reporter gene expression (Macfarlan et al., 2012).

In experiments were I triggered transcriptional silencing by using three Zfp809-binding sites, I found that silencing of the lentiviral reporter occurred within 12 to 24 h (Figure 2.6). Yet, it is unclear, whether the lentiviral EGFP reporter containing the silencing sequence is expressed at all after integration in the genome, because the small amount of EGFP-fluorescence arising shortly after transduction could as well originate from non-integrated viral RNA. Gene expression from non-integrated lentiviral RNA is well analyzed in integrase-defective lentiviral vector systems (Nightingale et al., 2006). In summary, this suggests that heterochromatic silencing of the reporter could occur even in the first twelve hours after transduction or immediately after proviral integration. Consequently, epigenetic retrotransposon-defense mechanisms might require the rapid silencing of newly integrated retrotransposon copies before they can give rise to further retroviral transcripts and retrotransposition events.

3.4 The GAG region of IAP-Ez retrotransposons contains a 160 bp sequence (GAG2.22) that silences strong constitutive promoters in mouse ES cells, but does not recruit a DNA-binding factor in DNA pull-downs *in vitro*.

In previous studies it was shown that the UTR region of mouse IAP retrotransposons silences neighboring promoters in lentiviral reporter assays (Rowe et al., 2010). I tried to verify this finding by checking whether my lentiviral vector system can detect silencing activities of these IAP retrotransposon sequences. Indeed, I could not only detect a silencing activity of the UTR sequence, but also found a novel, stronger silencing activity inside the GAG sequence of the IAP element (Figure 2.8). This sequence element could be narrowed down to a size of 160 bp, which are sufficient for silencing of strong constitutively active promoters like the PGK and the EF1α promoter (Figure 2.9).

Interestingly, removing larger parts of the 160 bp GAG2.22 silencing sequence at its 5’ or 3’ end or removing internal sequences resulted in a severe reduction of silencing activity (experiments performed by Katharina Schmidt - data not shown) (Schmidt, 2014). Presuming that the DNA sequence of the GAG2.22 sequence is recognized by DNA-binding proteins to facilitate transcriptional repression, the DNA-binding site is either unusually long or it is bound by multiple DNA-binding factors that cooperate in transcriptional repression. Due to the intrinsic symmetry of the DNA double strand, many sequence-specific DNA-binding factors bind to palindromic sequences. Searching for palindromic sequences inside the GAG2.22 sequence identifies only few, very short palindromes (http://www.alagu-molbio.net/)

(Figure 3.1A). However, assuming that a dimer of DNA-binding proteins might also recognize binding sites on the sense and antisense strand that are more divergent, I aligned the GAG2.22 sequence with its reverse complement sequence using the YASS alignment software and the LALIGN algorithm (Huang and Miller, 1991; Noe and Kucherov, 2005). Interestingly, the GAG2.22 sequence contains two independent partially palindromic sequences (Figure 3.1B). These hypothetical transcription factor binding sites are separated by a spacer of 53 bp and consequently cover a region from nucleotide 9 to 142 of GAG2.22 (Figure 3.1B and C). It might therefore be possible that two independent transcription factors recognize and target the GAG2.22 sequence for silencing (Figure 3.1C). If this binding model was true, why should evolution select for two recognition motifs in such a short sequence? The use of two repressor binding sites probably generates a more robust system, because mutations of the GAG2.22 sequence would not easily result in a failure of repression. Consistently, we have found that silencing is strongly decreased but not lost when the GAG2.22 sequence is shortened at the 5' and the 3' end (experiments performed by Katharina Schmidt - data not shown) (Schmidt, 2014). In addition, the use of two repressor binding sites might also guarantee a higher specificity of binding for the silencing machinery. This might avoid "off-target" effects at other regions of the genome that could be disastrous for the organism.

We tried to identify DNA-binding factors that are recruited to GAG2.22 using DNA pull-down experiments followed by SILAC mass spectrometry in collaboration with Dr. Falk Butter (Figure 2.25). Even though binding of Zfp809 and its associated protein Trim28 to its known DNA-binding site could easily be detected, we were not able to identify a specific binding of a repressive GAG2.22-binding protein (Figure 2.25). A protein that was found to be enriched at the GAG2.22 sequence compared to a point mutant sequence was Ubtf (Figure 2.25), a very abundant protein involved in rRNA transcription. Depletion of Ubtf did not lead to a relief of GAG-silencing, suggesting that it might bind to the GAG2.22 sequence but is not involved in transcriptional repression (Figure 2.25). The failure of identifying specific hits by DNA pull-down experiments could be either explained in the way that the GAG2.22-recognizing protein cannot be identified with the used approach or by the fact that the GAG2.22 sequence is recognized by a much more complex mechanism. Indeed, if two independent GAG2.22-binding proteins are important for GAG2.22 recognition, these proteins will interact and result in a complex rearrangement of the DNA like looping or compaction. The formation of such a higher order repressor complex might be highly unfavorable *in vitro* and might not occur on immobilized linear DNA. The existence of such a repressor complex at the GAG2.22 sequence could also explain that the GAG2.22 sequence is slightly depleted of histones and histone modifications in ChIP-sequencing data (Figure 2.16A and data not shown). Future experiments using GAG2.22-sequences mutated in the proposed repressor-binding

sequences will test the proposed binding-model. Concatemers of the isolated individual binding-sites in DNA pull-downs might also generate a higher affinity for the isolation of the individual transcription factors.

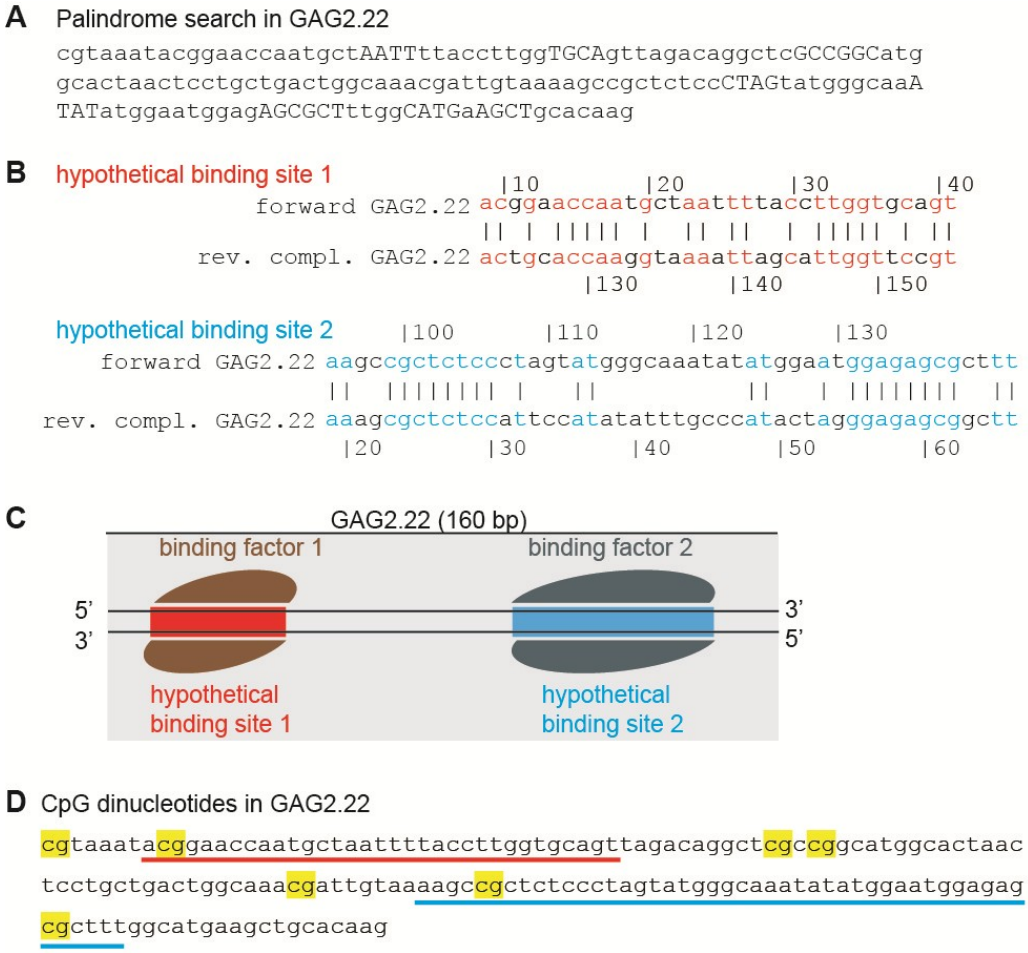


Figure 3.1 Analysis of the GAG2.22 sequence and a hypothetical binding model of GAG2.22 recognition
A) A search for palindromes (highlighted in capital letters) inside the GAG2.22 sequence does not reveal long palindromic sequences. B) Alignment plots of the forward and reverse complement strand of GAG2.22 (in 5' to 3' direction) reveal two symmetric and partially palindromic sequences that might be potential repressor binding sites. C) A hypothetical binding model: Two transcription factor dimers might be involved in recognition of the GAG2.22 sequence to ensure that silencing specifically occurs on IAP elements. D) CpG dinucleotides are highlighted in yellow inside the GAG2.22 sequence. Hypothetical repressor binding sites are indicated in red and blue. All sequences in this figure are shown in 5' to 3' direction.

On the other hand, transcriptional repression of transposons in plants has been described to occur via complex RNA-dependent mechanisms (Law and Jacobsen, 2010). A possible explanation why we could not identify a binding protein for the GAG2.22 sequence might therefore be that the GAG2.22 sequence is only triggering epigenetic silencing when transcribed into RNA. Yet, the GAG2.22 sequence is upstream of the promoter sequence in the cloned reporter vectors and might not be transcribed at all. Conversely, most active promoters produce low abundant antisense transcripts that may still be sufficient for the

recruitment of silencing factors (Core et al., 2008; Djebali et al., 2012; Seila et al., 2008). However, the IAP-UTR and IAP-GAG sequence are both capable of silencing adjacent promoters even when they are cloned in antisense direction, which strongly argues against an RNA-directed mechanism (data not shown) (Rowe et al., 2010).

An interesting observation is the high density of CpG-dinucleotides inside the GAG2.22 sequence which are normally underrepresented in mammalian genomes due to mutagenic effects of CpG methylation (Figure 3.1D) (Shen et al., 1992). These CpG dinucleotides partially overlap with the hypothetical binding sites described above (Figure 3.1D). Methylation of cytosines inside CpG dinucleotides is one of the major repression mechanisms of endogenous retroviruses in later stages of development (reviewed in chapter 1.2.2.) and IAP elements are heavily DNA methylated throughout development (Figure 2.17). Therefore, it is possible that the accumulation of CpG dinucleotides inside the GAG2.22 sequence also has a role during the initiation of silencing and recognition of the GAG2.22 sequence. Recently it was shown that transcription factor binding and CpG-density are important regulators for the establishment of a repressive chromatin state by DNA-methylation enzymes in mouse ES cells (Lienert et al., 2011b). Therefore, recognition of the GAG2.22 sequence might also involve the detection of unmethylated CpG dinucleotides.

3.5. The GAG2.22 sequence silences reporter genes in a Setdb1- and Trim28-dependent manner and autonomously recruits H3K9me3.

Knockdown of Trim28 leads to an impairment of GAG2.22-mediated silencing (Figure 2.11). Trim28 has been extensively characterized as an epigenetic master regulator that is recruited via KRAB domain zinc finger proteins (Iyengar and Farnham, 2011). Strikingly, the KRAB domain zinc finger protein family shows a coevolution with endogenous retroviruses and contains many mouse-specific proteins (Thomas and Schneider, 2011). Consequently, it is very likely that mouse-specific KRAB domain zinc finger proteins are responsible for GAG2.22-dependent transcriptional repression in mouse ES cells. Many KRAB domain proteins are specifically expressed in mouse ES-cells (Corsinotti et al., 2013), and testing these proteins for their ability to repress GAG2.22 might be a promising approach to identify the mouse-IAP retrotransposon silencing factor in the future. Therefore, I grouped these KRAB zinc fingers for their average expression in mouse ES cells relative to their expression in MEFs based on a previous study (Corsinotti et al., 2013) (Figure 3.2).

In addition to Trim28, I also found that Setdb1 is important for GAG2.22-mediated repression (Figure 2.9). Setdb1 and Trim28 have both been shown to be master regulators of endogenous IAP elements and other endogenous retroviruses in mouse ES cells (Karimi et

al., 2011; Matsui et al., 2010; Rowe et al., 2010). Setdb1 has been extensively characterized to be recruited by Trim28 in a sumoylation-dependent manner and is most likely the enzyme that induces H3K9me3 at Trim28 target sites (Ivanov et al., 2007) (reviewed in 1.3.2). Previously, it has been shown that Setdb1 requires mCAF1 (Atf7ip) to act as an H3K9me3 methyltransferase (Schultz et al., 2002; Wang et al., 2003). Whether also Atf7ip is involved in Trim28/Setdb1-dependent silencing of endogenous retroviruses, could be analyzed by using the reporter systems established in this thesis.

Gene symbol	# of zinc finger	expression fold change mES vs. MEFs (2log)	hypothetical DNA coverage in nt	absolute expression mES cells	cDNA tested for GAG2 silencing in MEFs
Zfp459	3	12,3	12	515	
Zkscan16	8	7,3	32	15	
Gm14124	19	6,8	76	567	
Gm9631	17	5,6	68	913	
Zfp936	11	5,2	44	493	yes
Zfp78	11	4,9	44	3	
Zfp606	13	4,9	52	31	
2610305D13Rik	3	4,1	12	2913	
Zfp473	16	3,7	64	348	yes
Gm12588	9	3,5	36	435	
Zfp811	13	3,5	52	59	
Zfp456	10	3,3	40	4	
Zfp819	9	3,3	36	122	yes
Gm14443	16	2,7	64	493	
Gm13154	10	2,4	40	93	
C030039L03Rik	17	2,3	68	190	
Zfp961	11	2,0	44	2334	
4930522L14Rik	20	2,0	80	490	
Gm9894	4	1,9	16	0	
Zfp7	16	1,8	64	676	yes
Zfp114	10	1,4	40	50	
Zfp59	16	1,4	64	539	
Zfp345	14	1,4	56	137	
Zfp418	14	1,4	56	193	
Zfp953	8	1,3	32	192	

Figure 3.2 25 KRAB zinc finger proteins that are potential candidates for being the mouse-IAP retrotransposon silencing factor
KRAB zinc finger proteins were grouped for their expression in mES cells relative to their expression in MEFs based on published RNA-seq analyses (Corsinotti et al., 2013). The indicated expression units are relative RNA-seq read counts (Corsinotti et al., 2013). For some of the candidate proteins, I cloned the respective cDNAs and generated stable MEF-cell lines to see whether these cells show ectopic GAG2-repression. However, I could not induce GAG-dependent silencing in MEFs by overexpressing the indicated cDNAs (data not shown).

The UTR-region of IAP retrotransposons also recruits Trim28-dependent transcriptional silencing and DNA *de novo* methylation (Rowe et al., 2013a; Rowe et al., 2010). Consistently, I found that GAG2.22-dependent reporter silencing also involves Trim28 and Setdb1 (Figure 2.9), indicating that IAP element sequences might indeed have multiple nucleation sites for recruiting Setdb1 and Trim28. In my hands, reporter silencing mediated by the IAP-GAG sequence was massively stronger than IAP-UTR-mediated silencing (Figure 2.9). This indicates that the GAG-mediated silencing might be the predominant way of how IAP-Ez retrotransposons are targeted for epigenetic silencing in early mouse

development. Moreover, the IAP-Ez-GAG region encodes the structural capsid protein, which is important for the formation of retroviral particles and which shows a much higher evolutionary conservation than the UTR region of IAP elements (Kuff and Lueders, 1988; Rowe et al., 2010). One might speculate that the GAG2.22-targeting mechanism, which covers a large portion of a functional and coding IAP gene, robustly prevents the bypass of epigenetic silencing by inserting small point mutations.

Interestingly, *Suv39h* and *Suv4-20h* enzymes are not essential for GAG2.22-mediated silencing (Figure 2.9). Consistently, *Suv4-20h* knockout ES cells don't show a massive upregulation of endogenous IAP elements and *Suv39h* knockout cells largely maintain H3K9me3 methylation on IAP elements (Bulut-Karslioglu et al., 2012; Matsui et al., 2010). However, quantitative assays show that *Suv4-20h* knockout ES cells have a slight upregulation of endogenous IAP elements, probably because Suv4-20h contributes to silencing of IAP elements by non-coding-RNA (personal communication Prof. Ingrid Grummt – article in press). In addition, others have shown that many H3K9me3-binding proteins are dispensable for retrotransposon silencing (Maksakova et al., 2011). This is in accordance with the luciferase screening approach I have shown in this study (Figure 2.3).

The most prominent pathway of retrotransposon-repression in later development is DNA methylation, although it is dispensable for silencing IAP elements at the ES-cell stage (extensively reviewed in chapter 1.2.2). In accordance with these previous studies, GAG2.22-dependent silencing is also occurring in *Dnmt1* knockout and *Dnmt3a/b* double knockout cells (Figure 2.9). Although, DNA-methylation is dispensable for retroviral silencing in mouse ES cells and early embryonic stages, Trim28-dependent silencing has recently been implicated to play a vital role in the deposition and maintenance of DNA-methylation at endogenous retroviral sequences, imprinted loci and other genomic regions in early mouse development (Messerschmidt et al., 2012; Quenneville et al., 2012; Quenneville et al., 2011; Rowe et al., 2013a). The recruitment of Trim28 to IAP elements might therefore not only be important to silence IAP elements during early development but also to ensure maintenance and establishment of DNA-methylation, because at later stages of development DNA-methylation is crucial for IAP silencing when Trim28 and Setdb1 become dispensable for this process (see chapter 1.2.2) (Rowe et al., 2013a; Walsh et al., 1998).

I used a recombinase-mediated cassette exchange (RMCE) system to integrate the GAG2.22 sequence into chromatin (Baubec et al., 2013; Lienert et al., 2011b), and I could find that it is sufficient to recruit the heterochromatic H3K9me3 mark (Figure 2.10). The GAG2.22 sequence not only induces silencing of nearby promoters but also autonomously recruits histone modifications to nearby chromatin. This finding is an example of how DNA-sequences can directly determine the chromatin state and it questions the hypothesis that

the chromatin state is mostly inherited epigenetically and independently of the underlying DNA sequence. How many of these functional “chromatin-defining” sequences exist *in vivo* is unknown but could be identified by an experiment I propose in Figure 3.3. In short, the genome is fragmented, cloned into transfer vectors and the genomic fragments are reinserted into the cellular genome. If a fragment *de novo* recruits a histone modification, it will be isolated by histone modification ChIP. ChIP-PCR using universal primers amplifying newly inserted DNA-fragments will enrich for “epigenome regulating sequences” that can be deconvoluted by next generation sequencing. Similar experiments that try to unravel the functionality of different DNA-sequences in a genome-wide manner have already been performed by others (Akhtar et al., 2013; Arnold et al., 2013).

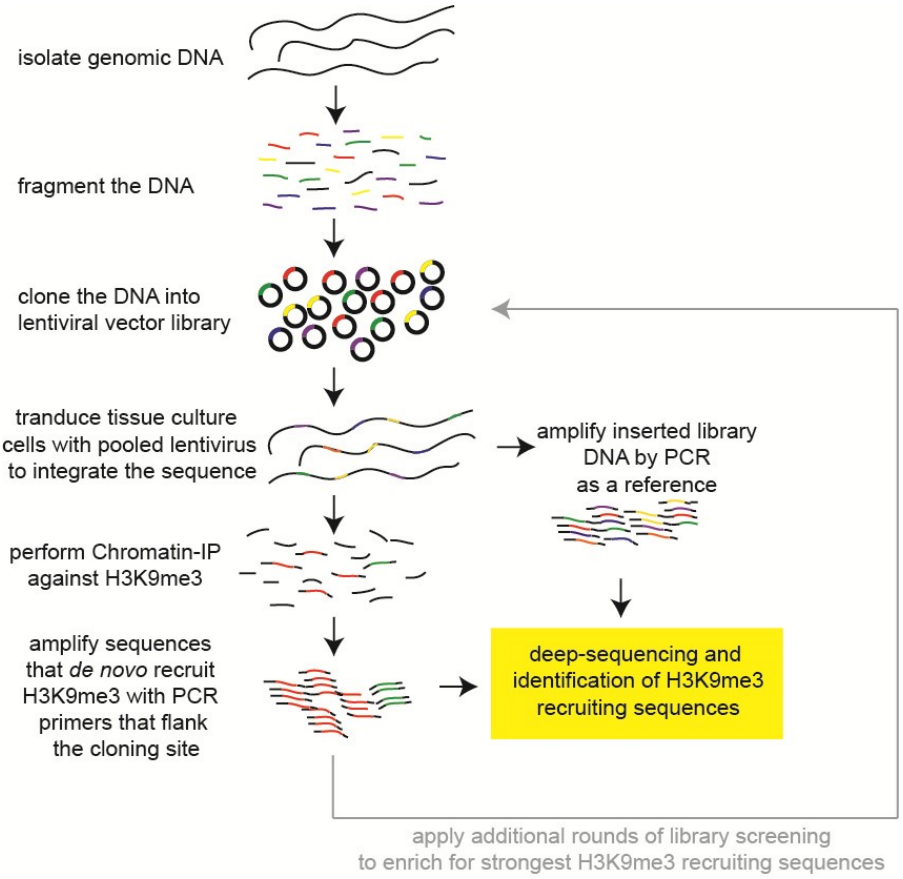


Figure 3.3 Sequencing Project for Identification of DNA-Elements that Recruit Modifications (*SPIDER-Mod*)
This proposed experiment could identify sequences that recruit specific chromatin modifications. Here H3K9me3 is used as an example. First, genomic DNA is sheared, end-repaired and DNA fragments are cloned into a lentiviral vector library with high coverage. Next, the lentiviral library is prepared and stably integrated into a cell line. After chromatin immunoprecipitation (ChIP) of H3K9me3-modified regions virally integrated DNA elements that are modified by H3K9me3 are enriched by PCR and identified using deep sequencing. Differences in library representation are controlled by amplifying and sequencing all integrated DNA sequences. For the analysis of active chromatin marks piggy-pac transposase systems can be used, that do not have an intrinsic preference for integrating into active chromatin domains. Instead of cloning fragments that represent the entire genome, a subpool of DNA-fragments that is endogenously modified by H3K9me3 could also be used for library generation.

3.6. Genome-wide shRNA screens for Zfp809- and GAG2.22-dependent silencing in MEFs and mouse ES cells

In this study, I performed two genome-wide shRNA-screening experiments to identify novel players that are involved in heterochromatic silencing (Figure 2.12 and 2.13). In the screen for Zfp809-dependent silencing in MEF cells Trim28 was identified as the strongest hit, which is the master regulator of Zfp809-dependent silencing (Figure 2.13). This indicates that the established screening-assay is capable of detecting functional hits.

However, the number of potential off-target hits seems to be high in both screens, indicated by the fact that a substantial number of scoring shRNAs target genes that are not expressed (see list in appendix and Figure 2.14). This is probably due to fact that the cell-sorting step in the screening outline displayed a bottleneck in the screening experiment. Because only very limited numbers of EGFP-positive cells could be sorted, a high coverage of the shRNA library could not be maintained throughout the experiment (chapter 2.13). Nonetheless, the top hits should still be informative and it will be very interesting to validate them in the context of heterochromatin silencing.

Apart from Trim28, the second strongest hit that scored with three independent shRNAs in the Zfp809/MEF screen was *Gak* (Figure 2.13). Gak is a kinase of 150 kDa that has been implicated into clathrin-coated vesicle uncoating and which is associated with Cyclin G and Cdk5 (Greener et al., 2000; Kanaoka et al., 1997). Recent studies have also revealed a nuclear localization of Gak and have implicated Gak in the formation of the mitotic spindle and the attachment of tubulin-fibers to the chromosome (Sato et al., 2009; Shimizu et al., 2009; Tanenbaum et al., 2010). Therefore, Gak could be a novel interesting protein that is involved in heterochromatic silencing, maybe by influencing higher order chromosome structure.

The top hit in the screen for GAG2-dependent silencing in mES-cells was *Top2b*. This gene encodes DNA topoisomerase II beta, an enzyme that is required for the topological organization of DNA by transiently breaking and rejoining DNA inside the nucleus. Interestingly, topoisomerase II beta localizes to pericentric heterochromatin in mouse cells and was found to be associated with Trim28 in my mass-spectrometry-based screen (Figure 2.24) (Cowell et al., 2011). Topoisomerase II beta has also been implicated into gene transcription in general and has been shown be important for the regulation of imprinted genes (Baranello et al., 2012; Huang et al., 2012; King et al., 2013; Vos et al., 2011). In addition to *Top2b*, I could also identify *Top2a* as an interactor of Trim28 in mouse ES cells and as hit in the GAG2/mES cell shRNA screen (Figure 2.24). This further supports the hypothesis that topoisomerase II is an important player in heterochromatic gene silencing. A

potential model that could explain an involvement of topoisomerase II into heterochromatin formation could be that heterochromatin formation requires the relaxation of DNA supercoils generated by chromatin condensation. Alternatively, topoisomerase II might collaborate with chromatin remodeling enzymes to remove DNA supercoils generated by the sliding and deposition of nucleosomes (Varga-Weisz et al., 1997). Interestingly, I identified the SNF-2-type ATPase chromatin remodeler Atrx as a modulator of GAG2-dependent reporter silencing (Figure 2.14). Analyzing *Top2b* function in heterochromatin formation in the context of Atrx might therefore be an interesting project to study in the future.

3.7. Atrx and Daxx catalyze heterochromatic gene silencing

Secondary screening of shRNAs enriched in the primary shRNA screen revealed that the SNF2-type chromatin remodeler Atrx is important for silencing of the IAP-GAG reporter (Figure 2.14). To verify this finding and to identify whether the Atrx-associated protein Daxx is also involved in this process, I generated *Atrx* and *Daxx* knockout cells using RNA-guided nucleases (Figure 2.15 and 2.18). Interestingly, GAG-silencing was not lost in *Atrx* and *Daxx* knockout cells, but was massively delayed (Figure 2.15 and 2.18), indicating that Atrx and Daxx are not mandatory for reporter silencing but rather accelerate heterochromatic silencing. But how are Daxx and Atrx targeted to chromatin?

Daxx has already been found to be important for transcriptional repression mediated by sumoylated transcription factors where it is targeted by its Sumo-interaction motif (reviewed in chapter 1.3.4.) (Chang et al., 2005; Croxton et al., 2006; Hollenbach et al., 1999; Kim et al., 2003; Kuo et al., 2005; Lehembre et al., 2001; Li et al., 2000b; Lin et al., 2004; Lin et al., 2006; Lin et al., 2003; Obradovic et al., 2004; Park et al., 2007; Shih et al., 2007; Zhao et al., 2004). Interestingly, also the nuclear localization of Daxx to PML nuclear bodies and to pericentric heterochromatin seems to be dependent on its Sumo-interaction motif (Ishov et al., 1999; Lin et al., 2006; Shih et al., 2007). Consequently, I could show that in rescue experiments wild-type Daxx can restore the repressive activity of Daxx but not a Daxx mutant lacking its Sumo-interaction motif (Figure 2.19). This suggests that Daxx might also be recruited to the newly inserted GAG2.22 reporter in a Sumo-dependent manner. Interestingly, also Setdb1 is recruited via Sumo-dependent interactions and is associated with sumoylated Trim28, which has been shown to have autosumoylation activity (Ivanov et al., 2007). It can therefore be hypothesized that Daxx is also recruited to sumoylated Trim28. Alternatively, Daxx might also be targeted to sumoylated transcription factors, which might be sumoylated during heterochromatin formation. The transcription factor YY1, which binds the promoter region of IAP elements and other cellular promoters (including the human EF1α promoter), has recently been shown to greatly accelerate heterochromatin formation on

M-MLV retroviruses but not to be absolutely essential for silencing in general (Gerstein et al., 2012; Schlesinger et al., 2013). Interestingly, YY1 also binds to IAP elements *in vivo* and can also be sumoylated (Deng et al., 2007; Schlesinger et al., 2013), underlining the possibility that also sumoylated YY1 might recruit Daxx to IAP sequences. Moreover, also Atrx has been found to be a potential Sumo-target in proteome-wide screens (Westman et al., 2010), suggesting that the Sumo-interaction motif of Daxx might also be important for its interaction with Atrx. Which sumoylated protein is actually important to recruit Daxx is unknown and will be hard to identify among the high number of sumoylated proteins involved in heterochromatin formation. Yet, there is still the possibility that Atrx is the most important targeting factor for Daxx.

Atrx itself has been shown to bind unmodified H3K4 in combination with di- or trimethylated H3K9 with its ADD-domain (Eustermann et al., 2011; Iwase et al., 2011). In addition, Atrx has an Hp1-interaction motif that together with the ADD-domain recruits Atrx to heterochromatic regions (Eustermann et al., 2011; Iwase et al., 2011). Therefore, it is possible that Atrx and Daxx are recruited to chromatin independently. However, Atrx and Daxx associate on protein level (Xue et al., 2003), which rather supports the hypothesis that both proteins act as a stable Atrx/Daxx complex that uses many different protein interaction modules to be tethered to certain chromatin regions. Testing *Daxx/Atrx* double knockout cells in my reporter assay will further reveal whether Daxx and Atrx act in the same silencing pathway and consequently, whether they can act independently from each other.

Atrx is enriched at the GAG2.22 sequence of endogenous IAP elements *in vivo* and is recruited to new insertions of the GAG2.22 sequence (Figure 2.16). Although H3K9me3 is present all over the analyzed regions, Atrx does not spread over the complete IAP retrotransposon sequence *in vivo* or on a reporter gene *in vitro,* but rather stays localized at the GAG2.22 sequence (Figure 2.16 and 2.10). This argues against the hypothesis that Atrx is recruited exclusively via its ADD-domain and its Hp1 interaction motif because one would rather expect a colocalization with H3K9me3. A possible explanation for this observation could be that the Atrx/Daxx complex has the highest affinity to proteins at the nucleation site of silencing (presumably Trim28). In fact, Atrx was found as a Trim28-interacting protein using mass-spectrometry-based screening experiments (Figure 2.24). Rescue experiments using Atrx mutants in Atrx deficient cells will reveal which part of the protein is responsible for its targeting.

In addition, one could analyze whether the ATPase activity of the SNF2-type domain of Atrx is required for its repressive activity. Some ATRX-syndrome patients have mutations in the SNF2-type ATPase domain of Atrx (Gibbons et al., 2008), suggesting that this domain might be important *in vivo*. However, *in vitro* assays did not reveal any specific physiological

function of the ATPase domain of Atrx, yet (Xue et al., 2003). Due to its binding to certain repetitive DNA elements in ChIP experiments, it has been postulated that the ATPase of Atrx might actively dissolve G-quadruplex structures that form in G-rich DNA regions (Law et al., 2010). However, this may be the case, but does not explain, why Atrx is required for quick heterochromatin formation.

Consistent to the fact that *Atrx* and *Daxx* knockout cells are not defective in reporter silencing, *Atrx* and *Daxx* knockout cells don't show an upregulation of endogenous IAP elements and depletion of Atrx does not result in a reduction of H3K9me3 or DNA-methylation at endogenous IAP elements (Figure 2.16, 2.17 and 2.23). A good, explanation for this finding is that Atrx and Daxx are probably not required for maintenance of H3K9me3 chromatin or DNA-methylation, but are rather involved in the establishment of heterochromatin. In line with this hypothesis, an impaired heterochromatin establishment during developmental transitions might also explain the reduced DNA-methylation at certain chromosomal regions in Atrx patients (Gibbons et al., 2000), while direct effects on already established DNA methylated regions cannot be observed (Figure 2.17). However, I cannot rule out that Atrx has a role during maintenance of heterochromatin, because loss of Atrx might also be compensated by an increased activity of Trim28, Setdb1 and the residual heterochromatin machinery. Consistently, knockdown experiments showed that Setdb1 and Trim28 are rate-limiting for *de novo* silencing of the GAG-reporter at a much higher concentration when Atrx is missing (Figure 2.22). In addition, *Atrx* knockout cells show much stronger derepression of endogenous IAP elements when *Trim28* is depleted (Figure 2.22). Therefore, Atrx might facilitate a low plasticity of heterochromatin and render it much more resistant against influences that impair heterochromatin status. Consequently, I could show that reactivation of an EGFP reporter gene that is silenced by H3K9me3 is much easier upon Atrx knockdown (Figure 2.23). How loss of Atrx affects endogenous genes that are controlled by H3K9me3 will be an interesting aspect to study in the future. In fact, some imprinted genes have been shown to be deregulated in neuronal tissues in *Atrx* knockout mice, suggesting a potential link to the ATRX syndrome phenotype (Kernohan et al., 2010). In addition, low heterochromatin plasticity at telomeric regions could contribute to the higher frequency of telomeric recombination events observed in the alternative lengthening of telomeres pathway observed in many *Atrx*-deficient tumors (reviewed in 1.3.3).

Daxx was also found to incorporate histone H3.3 at telomeres and other regions (Drane et al., 2010; Elsasser et al., 2012; Goldberg et al., 2010; Lewis et al., 2010; Liu et al., 2012). Even though H3.3-binding seems to be one of the main functions of Daxx, Daxx mutants that are deficient in H3.3-binding can still silence the IAP reporter gene (Figure 2.19 and 2.20). This suggests that H3.3-binding and H3.3 itself don't have a functional role in Daxx-mediated

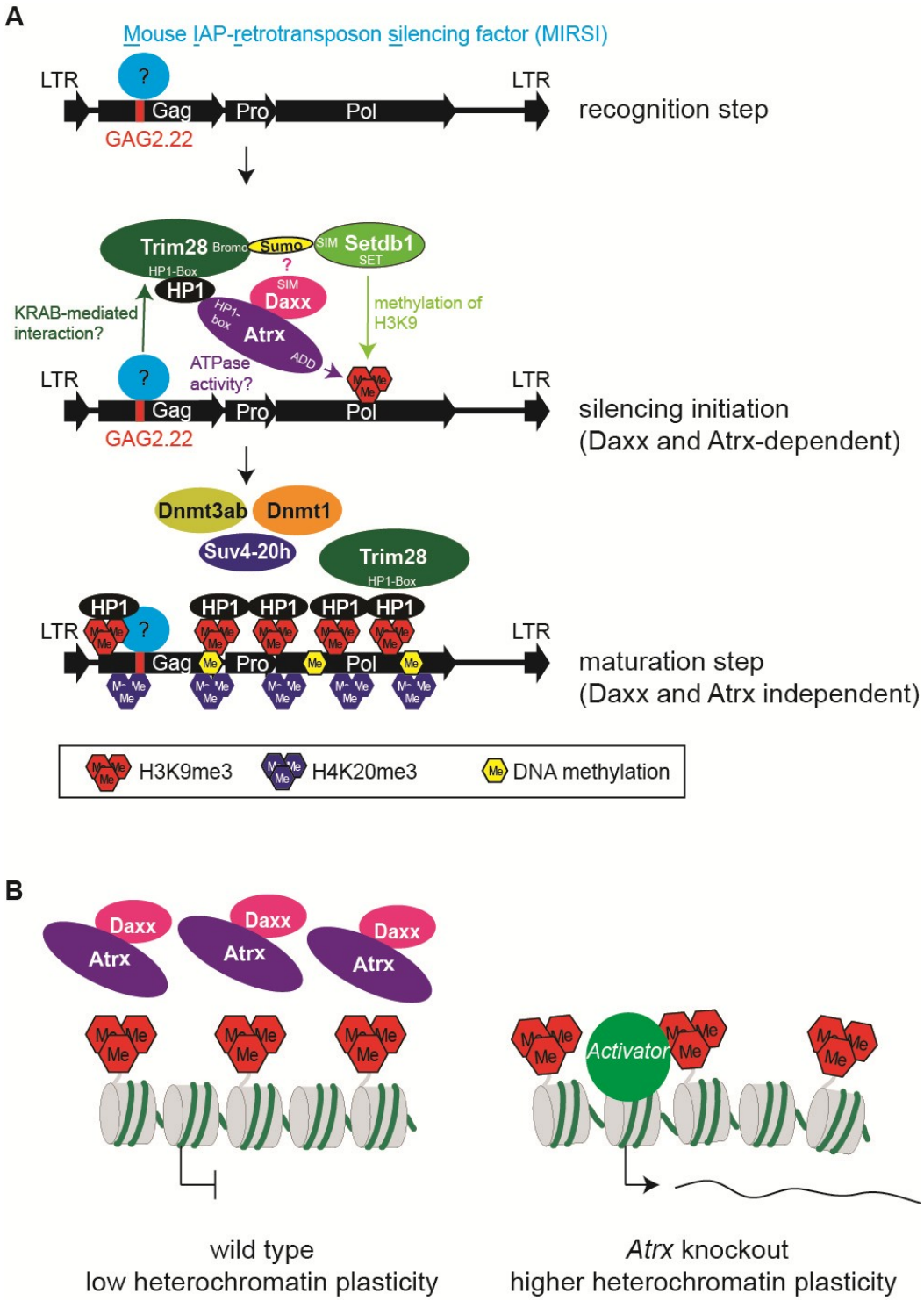


Figure 3.4 A silencing model for IAP retrotransposons
A) Silencing of newly integrated IAP retrotransposons might start with the recognition of the GAG2.22 sequence (and possibly other IAP regions) by a mouse IAP retrotransposon silencing inducing protein. This sequence-specific factor might then recruit Trim28, potentially via a KRAB domain. Sumoylation of Trim28 could be a binding-platform for Setdb1 and maybe the Daxx/Atrx complex. However, Atrx could also be recruited via binding of its ADD domain to H3K9me3, which has been catalyzed by Setdb1. Alternatively, Atrx could also bind Hp1 or Trim28. After establishment of the H3K9me3-domain, additional factors like Dnmts and Suv4-20h are recruited to set additional chromatin marks. These marks are most likely dispensable for heterochromatic silencing in mouse ES cells. Additional heterochromatin players (like HDACs, Atf7IP and the NURD complex) might also be involved in this pathway but are not shown here. B) A simplified illustration that shows how loss of Atrx leads to increased heterochromatin plasticity that allows activating factors to intrude into heterochromatin.

repression of IAP sequences. Consequently, H3.3-deficient cells can also silence IAP reporter genes similarly to wild type cells (Figure 2.21). Consistently, H3.3 incorporation is not changed at Atrx target loci when Atrx is depleted, indicating that the main function of Atrx outside telomeres is not to incorporate H3.3 (Law et al., 2010). It will be a very interesting topic in future research to investigate if H3.3 independent functions of Atrx and Daxx are also important in telomere biology.

3.8. Improvements in studying genetic interactions of heterochromatic silencing

Due to recent advancements in genome-editing by RNA-dependent nucleases, genome-wide screening-studies using genome-wide depletion screenings with gRNA/CRISPR libraries have become possible (Koike-Yusa et al., 2013; Shalem et al., 2014; Wang et al., 2014b). Repeating my genome-wide screening assays with these assays could reduce the amount of potential off-target hits substantially. In addition, one could circumvent the bottleneck of FACS-sorting in the established screening protocols by using an antibiotic resistance marker to monitor reporter gene expression (Figure 3.5). Pilot experiments using the *Zeocin*™ resistance gene, already showed the feasibility of antibiotic-based selection in such experiments (Figure 3.5). The use of *Zeocin*™ for these assays is specifically important, because it has a small therapeutic index. This means that cells expressing the zeoR resistance gene below a certain level can be counterselected by increasing *Zeocin*™ concentration (Nakatake et al., 2013). In this way, the antibiotic selection allows the adjustment of a certain expression threshold. In addition, the heterochromatin reporter can be challenged by a precise concentration of doxycyclin that only results in reporter expression if heterochromatin is impaired. This approach could be a simple and powerful screening platform for the identification of factors involved in heterochromatin maintenance.

Additionally, one could further exploit this approach to establish genetic screens for synthetic lethality. As an example, one could compare *Atrx*-deficient cells with wild type cells under very low doxycyclin induction to identify factors that are only required once Atrx is depleted. In this way, redundant pathways can be uncovered and a broad landscape of genetic interactions at mouse heterochromatin can be unraveled. Ultimately, the identified genetic interactions can then be validated in biochemical approaches in the context of other heterochromatic domains like pericentric or telomeric heterochromatin. Another way of validating these functional screens could be the use of mass-spectrometry-based interaction studies as performed in Figure 2.24. The combination of these approaches can give rise to a more specific understanding about which proteins are associated to heterochromatin and which proteins have functional roles at this domain.

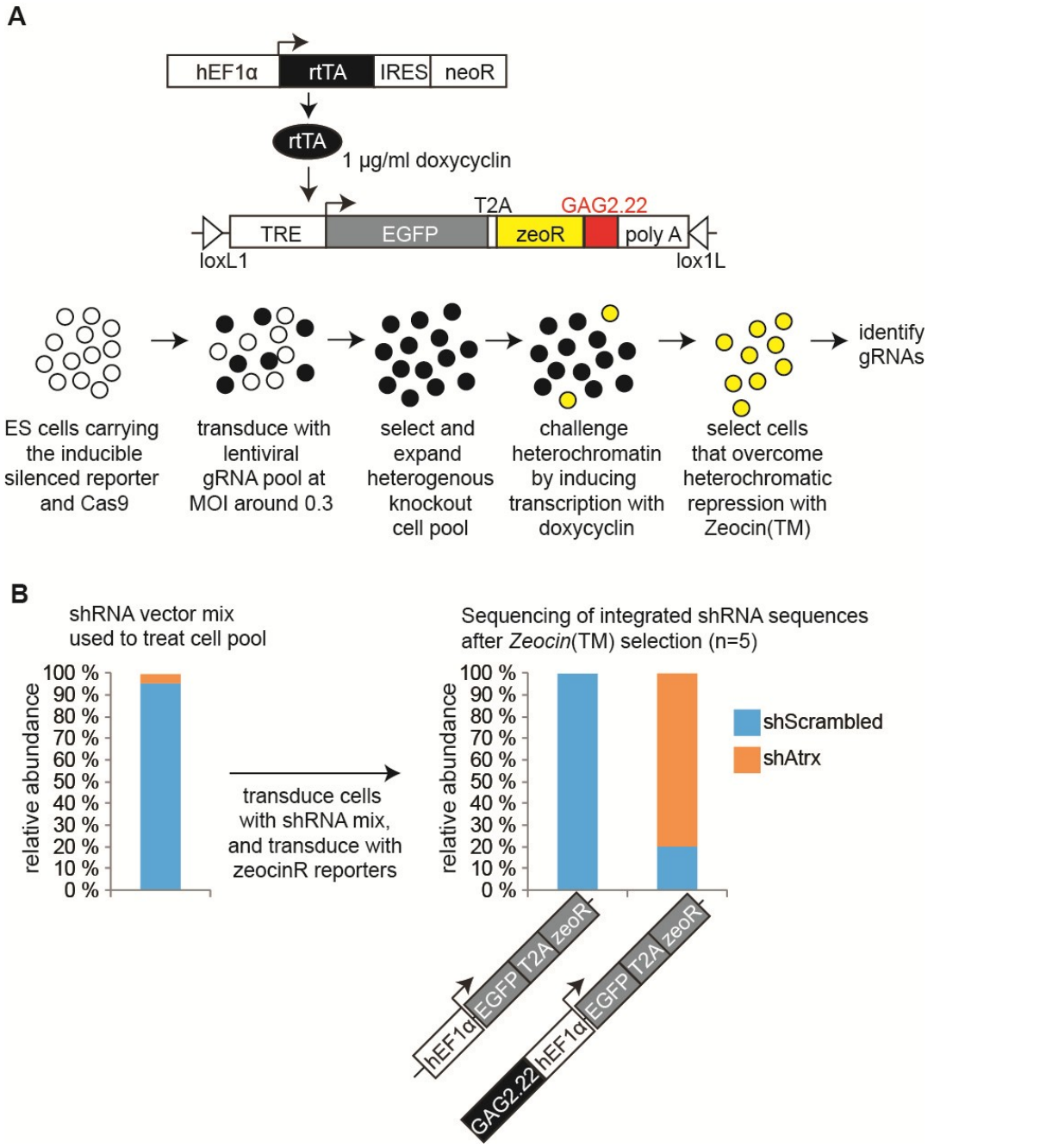


Figure 3.5 Improved screening systems to identify genetic interactions at heterochromatin A) A screening outline based on CRISPR/Cas9-mediated gene disruption. The screening cell line carries a heterochromatinized, silenced reporter gene that can be challenged with doxycyclin. Expression of the transgene only works when heterochromatin is impaired by the depletion of heterochromatin factors (see Figure 2.23). Reporter cells are depleted for individual genes by transduction with a lentiviral guide-RNA library. Addition of doxycyclin leads to expression of the reporter gene where heterochromatin is impaired by a loss of a heterochromatin factor. These cells can be selected by *Zeocin*™ selection and gRNA sequences can be identified by PCR and deep-sequencing. B) Proof-of-concept experiment showing that *Zeocin*™ selection can enrich for functional hits in a positive selection screen. A mixture of shRNA virus targeting Atrx or a scrambled control sequence was generated and mouse ES cells were transduced as described in Figure 2.12. Next, cells were transduced with a lentiviral GAG2.22/ZeoR-reporter and cells that bypass heterochromatic silencing were enriched by *Zeocin*™ selection. The barplot indicates the identified shRNA sequences of individual cell clones isolated from the experiment. The shRNA against Atrx was strongly enriched, even though it was strongly underrepresented in the initial virus mix. ShRNA sequences were identified by Sanger sequencing.

4. Material and methods

4.1. Material

4.1.1. Machines, buffers and reagents

Used kits, reagents, buffers and machines are indicated in the respective protocols in the methods section.

4.1.2. qPCR oligonucleotides

Primer design for quantitiative PCR was carried out using the *Primer3Plus* web interface (Untergasser et al., 2007) or designed manually.

Target	internal #	direction	sequence (5' to 3')	useage
Midline1 exon1	GS707	fw	tcaagtggccgagttgcttgg	RT-qPCR
	GS699	rw	gctatgtgcgcagggaagcag	
Midline1 exon2	GS708	fw	gctagcctgcgcccgactt	RT-qPCR
	GS709	rw	gcagacaccaaagccggca	
Midline1 exon3	GS1544	fw	agcttccttggactttgctg	RT-qPCR
	GS1545	rw	ttgaagcagaggctatgtgc	
Midline1 exon5	GS702	fw	gtcaatgcatcccgtcaaga	RT-qPCR & allele number
	GS703	rw	cttgccttctttaatctttgttc	
Midline1 exon6	GS704	fw	gtgatcaggctccgcaagttagc	RT-qPCR & allele number
	GS705	rw	gggcgtggtcattttccttcag	
Midline1 3' exons	GS494	fw	cctcccaggtcctaattcccg	semi-quantitative PCR
	GS495	rw	tctaatcgcgggagggttgg	
M18bp1	GS358	fw	ctccaaaaggccagcatcacg	semi-quantitative PCR
	GS359	rw	ttgccggaggtaggctgttcc	
Gapdh	GS276.1	fw	tcaagaaggtggtgaagcag	RT-qPCR
	GS276.2	rw	gttgaagtcgcaggagacaa	
Actin	GS278.1	fw	ggtcatcactattggcaacg	RT-qPCR
	GS278.2	rw	tccatacccaagaaggaagg	
Atrx	GS2493	fw	gagcttgacgtgaaacgaagag	RT-qPCR
	GS2494	rw	ttgttgctgttgctgctgag	
Daxx	GS2522	fw	gaacagttgcaggaagatcagg	RT-qPCR
	GS2523	rw	aaagtctgaaggcgatgtgg	
Trim28	GS1909	fw	cggaaatgtgagcgtgttctc	RT-qPCR
	GS1910	rw	cggtagccagctgatgcaa	
IAP global	GS2512	fw	cgggtcgcggtaataaaggt	RT-qPCR/ChIP-qPCR
	GS2513	rw	actctcgttccccagctgaa	
IAP GAG2.22	GS2600	fw	cttcatgccaaagcgctctc	ChIP-qPCR
	GS2601	rw	atggcactaactcctgctgac	
EGFP	GS1992	fw	cgacggcaactacaagac	ChIP-qPCR
	GS1993	rw	tagttgtactccagcttgtgc	

Target	internal #	direction	sequence (5' to 3')	useage
Tia1	GS477	fw	Gctcgccgccatcttggat	ChIP-qPCR
	GS478	rw	Ggctatggctgcggaagagc	
Polrmt	GS1650	fw	tcagcaaactccaatagcgcac	ChIP-qPCR
	GS1651	rw	ttgccgcacaacatggactt	
pLKO1 shRNA seq	GS2273	fw	cgagactagcctcgagc	library cloning
	GS2274	rw	ctgcgagggtactagtgag	
Rb1 intronic region	GS562	fw	gctgagcattctgctgcatcg	allele copy number
	GS563	rw	aagccgggagaaacagcctct	
Tspan32_2	GS890	fw	gccatgagagaggtgaggag	allele copy number
	GS891	rw	ggtaggtcccaagatgtga	
Osbpl5_1	GS924	fw	aaaaagccctgtccacatca	allele copy number
	GS925	rw	aaccaccaagagcattgtcc	
H3.3B	GS2879	fw	tttcaaagtgcagccatcgg	RT-qPCR
	GS2880	rw	tgtctttgggcatgatggtg	
Zfp71-rs1	GS2479	fw	tgttcctgggtcttgctttg	RT-qPCR
	GS2480	rw	ttttgccacactctgtgcac	
Zfp809	GS2119	fw	gcagaagcttcaagcaggag	RT-qPCR
	GS2120	rw	ggatgttgtccctttggaga	
GAG2.22/polyA	GS3012	fw	atggaatggagagcgctttg	ChIP-qPCR
	GS3013	rw	acagatggctggcaactagaag	
EGFP/T2A	GS3014	fw	tcctgctggagttcgtgac	ChIP-qPCR
	GS3015	rw	tcttccctctcctcccttgtac	
TRE	GS3016	fw	tgggaggcctatataagcagag	ChIP-qPCR
	GS3017	rw	aggtcaaaacagcgtggatg	

4.1.3. CRISPR guide-RNA sequences used with pX330 vector

Target	Internal Code	guide sequence (5'to 3')
Atrx ORF	#2_GS2621/GS2622	gctgttgcacgcagtcaccaagtccagtag
Atrx ORF	#1_GS2619/GS2620	cttctgtaagaaatgcatcctgcgcaacct
Daxx ORF	#4_GS2659/GS2660	agtacaatgatgctgtcatcgg
Daxx ORF	#1_GS2653/GS2654	tcaagtacaatgatgctgtcat
H3f3b ORF	#3_GS2827/GS2828	gcgcgcttttccgagccgcct
H3f3b ORF	#4_GS2829/GS2830	gcggaaaagcgcgccctctac
H3f3a upstream intron	#1_GS2939/GS2940	gataattagtttggaagggcg
H3f3a downstream int.	#2_GS2945/GS2946	gtatgtccgtgtaatttaac

4.1.4. Northern blot probes

probe	generation	size
X-chromosomal *midline1*	PCR on female mouse cDNA	790 bp
	Primer: 5'-tggactttgctgatgaccccg-3'	
	Primer: 5'-ttcaacatgttgacaagtttgg-3'	
pseudoautosomal *midline1*	PCR on male *Suv4-20h* dko cDNA	1292 bp
	Primer: 5'- gtcaatgcatcccgtcaaga-3'	
	Primer: 5'- gtccctcaaggtcgctgctccg-3'	

M18bp1	EcoRV-fragments of M18bp1 plasmid (#433) and gel purification of fragments	806 and 940 bp

4.1.5. DNA sequences used for DNA pull-down experiments

Zfp809-binding site

5'-ttaatttgggggctcgtccgggatcgggagacc-3'

Zfp809-binding site mutant

5'-ttaatttgggggggagctccgggatcgggagacc-3'

GAG2.22 + 40 bp

5'cgtaaatacggaaccaatgctaattttaccttggtgcagttagacaggctcgccggcatggcactaactcctgctgactggcaaacgattgtaaaagccgctct ccctagtatgggcaaatatatggaatggagagcgctttggcatgaagctgcacaagcgcaggcccgagcaaacgcagctgctttgactccagagcagagag attggacttttga-3'

Spectinomycin

5'gttggcgtcatcgagcgccatctcgaaccgacgttgctggccgtacatttgtacggctccgcagtggatggcggcctgaagccacacagtgatattgatttgctg gttacggtgaccgtaaggcttgatgaaacaacgcggcgagctttgatcaacgacctttttggaaacttcggcttcccctggagagagcgagattctccgcgctgta-3'

GAG2.22

5'cgtaaatacggaaccaatgctaattttaccttggtgcagttagacaggctcgccggcatggcactaactcctgctgactggcaaacgattgtaaaagccgctct ccctagtatgggcaaatatatggaatggagagcgctttggcatgaagctgcacaag-3'

GAG2.22_Mutant38

5'cgtaaatacggaaccaatgctaattttaccttggtgcagttag**g**caggctcgcc**c**gcatggcactaactcctgctg**g**ctggcaaacgattgtaaaagccgctct ccctagtatgggcaaatatatggaatggagagcgctttggcatgaagctgcacaag-3'

4.1.6. shRNA sequences used with pLKO1 vector backbone (TRC library)

Target	internal Code	sequence (5'to 3')
Trim28	A	GCAGTGAGGAACCAACGTAAA
Trim28	B	CCACCAGTCTTCAAGGTCTTT
Trim28	C	CCGCATGTTCAAACAGTTCAA
Trim28	D	AGACATCGTGGAGAATTATTT
Zfp809	A	GCATTACAGAAGAACTCATAA
Atrx	GAG8	CCCACGGATGAGAATGTAAAT
Cops6	GAG47	TAATGGGCCATTGTAAGCTGC
Mmp28	GAG3	GACCACTATTGGCATTTGGAT
Tmem174	GAG45	CGCTGTGTTCTTCAGAAGTTA
UNK	GAG40	CATCGCTTGTTCGTCATGCAT
Lrrc24	GAG33	GACACTGTTCCTGCAAGACAA
8430426H19Rik	GAG74	GAATGTAATCAGTGTGGTAAA
Hoxa11	GAG54	CTGCTCCTCTAACATGTATTT
Ubtf	A	GAGCCAGAAGGAGCACTATAA
Zfp71-rs1	C	GCCTTCCGTTATCCATCATTA
Zfp71-rs1	D	ACCAGATGCTTGATTTACGAA
Setdb1	B	GCCTTGATCTTCCATGTCATT
Scrambled control	-	CAACAAGATGAAGAGCACCAA

4.1.7 Plasmids used and cloned in this study

The following list contains the most important plasmids used and cloned for this study. Sequences, cloning strategies and vector maps are accessible in the Schotta lab plasmid database.

name	#	origin	marker	use
1L-PGK-HygTK-L1	1281	Dirk Schübeler Lab	Ampicillin	RMCE
1L-poly-L1	1280	Dirk Schübeler Lab	Ampicillin	RMCE
L1-Poly-1L/TetO-EGFP-T2A-Zeo-GAG2.22-pA	1294	this study	Ampicillin	RMCE
pBS-UAS-CMV-EGFP-Zeo	510	this study	Ampicillin	transient transfection
pBS-UAS-PGK-EGFP-Zeo	592	this study	Ampicillin	transient transfection
pcDNA3.1 hygro +/Gal4-VP16	605	Christian Haass Lab	Ampicillin	luciferase assay
pGal4-N1	589	Schotta lab	Kanamycin	luciferase assay
pGal4-N1/GW	506	Schotta lab	Kanamycin	luciferase assay
pGal4-N1_GW/Adnp	631	this study	Kanamycin	luciferase assay
pGal4-N1_GW/Atad2	632	this study	Kanamycin	luciferase assay
pGal4-N1_GW/Cbx1	552	this study	Kanamycin	luciferase assay
pGal4-N1_GW/CBX3	638	this study	Kanamycin	luciferase assay
pGal4-N1_GW/CBX5	635	this study	Kanamycin	luciferase assay
pGal4-N1_GW/Chd2	630	this study	Kanamycin	luciferase assay
pGal4-N1_GW/Dnmt3a	624	this study	Kanamycin	luciferase assay
pGal4-N1_GW/Dnttip2	634	this study	Kanamycin	luciferase assay
pGal4-N1_GW/M18bp1	522	this study	Kanamycin	luciferase assay
pGal4-N1_GW/MPP8	633	this study	Kanamycin	luciferase assay
pGal4-N1_GW/pogZ	629	this study	Kanamycin	luciferase assay
pGal4-N1_GW/Prdm6	623	this study	Kanamycin	luciferase assay
pGal4-N1_GW/Rbm39	622	this study	Kanamycin	luciferase assay
pGal4-N1_GW/Setdb1	636	this study	Kanamycin	luciferase assay
pGal4-N1_GW/Setdb2	626	this study	Kanamycin	luciferase assay
pGal4-N1_GW/Suv39h1	521	this study	Kanamycin	luciferase assay
pGal4-N1_GW/Suv39h2	520	this study	Kanamycin	luciferase assay
pGal4-N1_GW/Suv420h1	551	this study	Kanamycin	luciferase assay
pGal4-N1_GW/Suv4-20h2	519	this study	Kanamycin	luciferase assay
pGal4-N1_GW/Tardbp	625	this study	Kanamycin	luciferase assay
pGal4-N1_GW/TRIM28	637	this study	Kanamycin	luciferase assay
pGal4-N1_GW/Utf1	628	this study	Kanamycin	luciferase assay
pGal4-N1_GW/Zfp106	620	this study	Kanamycin	luciferase assay
pGal4-N1_GW/Zfp622	621	this study	Kanamycin	luciferase assay
pGal4-N1_GW/Zfp828	627	this study	Kanamycin	luciferase assay
pGL3/5xUAS/PGK/luciferase	646	this study	Ampicillin	luciferase assay
pIC-Cre	1279	Dirk Schübeler Lab	Ampicillin	RMCE
pLCIN (pLenti6-E2-Crimson-IRES-neo)	974	this study	Ampicillin	lentiviral
pLEIP-GW	966	this study	Ampicillin	lentiviral

name	#	origin	marker	use
pLenti6 puro	831	this study	Ampicillin	lentiviral
pLenti6-EF1a-3FLAG-IRES-PURO (pLFIP)	963	this study	Ampicillin	lentiviral
pLenti6-EF1a-Cre-ERT2-T2A-hyg	842	this study	Ampicillin	lentiviral
pLenti6-EF1a-CRE-IRES-PURO	1087	Katharina Schmidt	Ampicillin	lentiviral
pLenti6-EF1a-EGFP-IRES-PURO (pLEIP)	962	this study	Ampicillin	lentiviral
pLenti6-EF1a-rtTA-IRES-neo	1308	this study	Ampicillin	lentiviral
pLenti6-EFEGT-neo	940	this study	Ampicillin	lentiviral
pLenti6-EFEGT-neo/1xB2	948	this study	Ampicillin	lentiviral
pLenti6-EFEGT-neo/1xRBS	946	this study	Ampicillin	lentiviral
pLenti6-EFEGT-neo/3xB2	949	this study	Ampicillin	lentiviral
pLenti6-EFEGT-neo/3xRBS	947	this study	Ampicillin	lentiviral
pLenti6-EFEGT-neo/GAG1	1011	this study	Ampicillin	lentiviral
pLenti6-EFEGT-neo/GAG2	1012	this study	Ampicillin	lentiviral
pLenti6-EFEGT-neo/GAG2.1	1046	this study	Ampicillin	lentiviral
pLenti6-EFEGT-neo/GAG2.1+2 (A)	1050	this study	Ampicillin	lentiviral
pLenti6-EFEGT-neo/GAG2.1+2+3 (B)	1051	this study	Ampicillin	lentiviral
pLenti6-EFEGT-neo/GAG2.2	1047	this study	Ampicillin	lentiviral
pLenti6-EFEGT-neo/GAG2.2.1	1037	this study	Ampicillin	lentiviral
pLenti6-EFEGT-neo/GAG2.2.2	1038	this study	Ampicillin	lentiviral
pLenti6-EFEGT-neo/GAG2.2.3	1039	this study	Ampicillin	lentiviral
pLenti6-EFEGT-neo/GAG2.2.4	1040	this study	Ampicillin	lentiviral
pLenti6-EFEGT-neo/GAG2.2.5	1041	this study	Ampicillin	lentiviral
pLenti6-EFEGT-neo/GAG2.2.6	1042	this study	Ampicillin	lentiviral
pLenti6-EFEGT-neo/GAG2.2.7	1043	this study	Ampicillin	lentiviral
pLenti6-EFEGT-neo/GAG2.2.8	1054	this study	Ampicillin	lentiviral
pLenti6-EFEGT-neo/GAG2.2+3+4 (C)	1052	this study	Ampicillin	lentiviral
pLenti6-EFEGT-neo/GAG2.3	1048	this study	Ampicillin	lentiviral
pLenti6-EFEGT-neo/GAG2.3+4 (D)	1053	this study	Ampicillin	lentiviral
pLenti6-EFEGT-neo/GAG3	1013	this study	Ampicillin	lentiviral
pLenti6-EFEGT-neo/GAG4	1014	this study	Ampicillin	lentiviral
pLenti6-EFEGT-neo/GAG5	1015	this study	Ampicillin	lentiviral
pLenti6-EFEGT-neo/IAP-5'UTR	951	this study	Ampicillin	lentiviral
pLenti6-EFEGT-neo/IAP-GAG anti	952	this study	Ampicillin	lentiviral
pLenti6-EFEGT-neo/IAP-LTR	950	this study	Ampicillin	lentiviral
pLenti6-EFEGT-neo/POL	1016	this study	Ampicillin	lentiviral
pLenti6-EFEGT-neo/UTR1	1005	this study	Ampicillin	lentiviral
pLenti6-EFEGT-neo/UTR2	1006	this study	Ampicillin	lentiviral
pLenti6-EFEGT-neo/UTR3	1007	this study	Ampicillin	lentiviral
pLenti6-EFEGT-neo/UTR4	1008	this study	Ampicillin	lentiviral
pLenti6-EFEGT-neo/UTR5	1009	this study	Ampicillin	lentiviral

name	#	origin	marker	use
pLenti6-EFEGT-neo/UTR6	1010	this study	Ampicillin	lentiviral
pLenti6-mPGK-GW-EGFP-T2A-pur	936	this study	Ampicillin	lentiviral
pLenti6-mPGK-GW-FLAG-T2A-pur	938	this study	Ampicillin	lentiviral
pLenti6-PGEGT-neo	941	this study	Ampicillin	lentiviral
pLenti6-PGEGT-neo/1xB2	956	this study	Ampicillin	lentiviral
pLenti6-PGEGT-neo/1xRBS	954	this study	Ampicillin	lentiviral
pLenti6-PGEGT-neo/3xB2	957	this study	Ampicillin	lentiviral
pLenti6-PGEGT-neo/3xRBS	955	this study	Ampicillin	lentiviral
pLenti6-PGEGT-neo/IAP-5'UTR anti	959	this study	Ampicillin	lentiviral
pLenti6-PGEGT-neo/IAP-GAG	960	this study	Ampicillin	lentiviral
pLenti6-PGEGT-neo/IAP-LTR	958	this study	Ampicillin	lentiviral
pLenti6-puro	831	this study	Ampicillin	lentiviral
pLenti6-TR	172	Life Technologies	Ampicillin	lentiviral
pLFIP/DAXX E231A	1307	this study	Ampicillin	lentiviral
pLFIP/DAXX R257A	1306	this study	Ampicillin	lentiviral
pLFIP/Trim28 (aa1-480) no Stop	1095	this study	Ampicillin	lentiviral
pLFIP/Trim28 noStop	1094	this study	Ampicillin	lentiviral
pLFIP-DAXX-deltaSIM(aa1-731)	1273	this study	Ampicillin	lentiviral
pLFIP-DAXX-FL	1272	this study	Ampicillin	lentiviral
pLFIP-GW	967	this study	Ampicillin	lentiviral
pLIP-H3.3-HA	1274	this study	Ampicillin	lentiviral
pLP-eco env	811	this study	Ampicillin	virus packaging
pPIH	651	this study	Ampicillin	MLV retroviral
pPIH/Slc7a1(mCAT1)	1055	this study	Ampicillin	MLV retroviral
psPAX2	183	Didier Trono lab	Ampicillin	virus packaging

4.1.8. Antibodies

ChIP

epitope	host		company	product
H3K9me3	rabbit	polyclonal	Active-Motif	anti-H3K9me3 (339161.39162 13509002)
Atrx	rabbit	polyclonal	Santa Cruz	anti-ATRX (sc-15408 H1412)

Western blotting or immunoprecipitation

epitope	host		company	product	dilution
Daxx	rabbit	polyclonal	Santa Cruz	anti-DAXX (sc7152 E1412)	1:3000
Atrx	rabbit	polyclonal	Santa Cruz	anti-ATRX (sc-15408 H1412)	1:15000
H3.3	rabbit	polyclonal	Millipore	anti-H3.3 (09-838)	1:1000
HA	rat	monoclonal	Roche	anti-HA (clone 3F10)	1:1000
α-tubulin	mouse	monoclonal	Sigma-Aldrich	anti-α-Tubulin (clone B-5-1-2)	1:3000
Trim28	rabbit	polyclonal	Bethyl	anti-Trim28 (A300-275A-1)	1:3000
FLAG	mouse	monoclonal	Sigma-Aldrich	anti-FLAG-M2-affinity-agarose	30-50 µl per IP

4.1.9. Cell lines used and generated in this study

cell line	type	origin	generation
293T cells	human embryonic kidney cells	commercial	
HeLa cells	human cervix carcinoma cells	commercial	
wt26	feeder-independent mouse ES cells	Schotta lab	
HA36	feeder-independent mouse ES cells for RMCE	Dirk Schübeler lab	
B5.3	feeder-independent mouse ES cells	Schotta lab	
AinV15	feeder-independent mouse ES cells	commercial	
E19.8	immortalized mouse embryonic fibroblasts	Schotta lab	
W9	immortalized mouse embryonic fibroblasts	Schotta lab	
KO-1-07	Atrx knockout cells	this study	wt26 cells (CRISPR targeted with sgRNA mix against Atrx)
KO-1-40	Atrx knockout cells	this study	wt26 cells (CRISPR targeted with sgRNA mix against Atrx)
KO-1-44	Atrx knockout cells	this study	wt26 cells (CRISPR targeted with sgRNA mix against Atrx)
KO-1-45	Atrx knockout cells	this study	wt26 cells (CRISPR targeted with sgRNA mix against Atrx)
KO2-3	Daxx knockout cells	this study	wt26 cells (CRISPR targeted with sgRNA DAXX#4)
KO2-7	Daxx knockout cells	this study	wt26 cells (CRISPR targeted with sgRNA DAXX#4)
KO2-18	Daxx knockout cells	this study	wt26 cells (CRISPR targeted with sgRNA DAXX#4)
KO2-24	Daxx knockout cells	this study	wt26 cells (CRISPR targeted with sgRNA DAXX#4)
KO9-1	Daxx knockout cells	this study	HA36 cells (CRISPR targeted with sgRNA DAXX#1)
KO9-3	Daxx knockout cells	this study	HA36 cells (CRISPR targeted with sgRNA DAXX#1)
KO10-3	Atrx knockout cells	this study	HA36 cells (CRISPR targeted with sgRNA ATRX#1)
KO11-1	H3.3B knockout cells	this study	wt26 cells (CRISPR targeted with sgRNAH3.3#3)
KO12-4	H3.3B knockout cells	this study	wt26 cells (CRISPR targeted with sgRNAH3.3#4)
KO20-21	H3.3 double knockout cells	this study	KO11-1 cells (CRISPR-targted with sgRNAs H3.3A#1 and #2)
KO20-18	H3.3 double knockout cells	this study	KO11-1 cells (CRISPR-targted with sgRNAs H3.3A#1 and #2)
KO20-24	H3.3 double knockout cells	this study	KO11-1 cells (CRISPR-targted with

			sgRNAs H3.3A#1 and #2)
KO20-20	H3.3 double knockout cells	this study	KO11-1 cells (CRISPR-targted with sgRNAs H3.3A#1 and #2)
KO21-7	H3.3 double knockout cells	this study	KO12-4 cells (CRISPR-targted with sgRNAs H3.3A#1 and #2)
KO21-13	H3.3 double knockout cells	this study	KO12-4 cells (CRISPR-targted with sgRNAs H3.3A#1 and #2)
HA36_TetO-9	HA36 cells with TetO-EGFP-T2A-ZeoR-GAG2.22	this study	RMCE in HA36 cells
HA36_TetO-8	HA36 cells with TetO-EGFP-T2A-ZeoR-GAG2.22	this study	RMCE in HA36 cells
T86	inducible HA36_TETO-8 cells expressing rtTA	this study	lentiviral tranduction of HA36_TetO8 with rtTA
T87	inducible HA36_TETO-9 cells expressing rtTA	this study	lentiviral tranduction of HA36_TetO9 with rtTA
T37	HeLa cells expressing Slc7a1 (mCAT1)	this study	transfection and transduction of retroviral Slc7a1 in HeLa
T52	cells expressing 3xFLAG/HA-Trim28(RBCC)	this study	wt26 cells stably transduced with Trim28-RBCC domain
T53	cells expressing 3xFLAG/HA-Trim28 full length	this study	wt26 cells stably transduced with Trim28
T55	cells expressing 3xFLAG/HA-Trim28 full length	this study	w9 cells stably transduced with Trim28

**Daxx rescue cell lines are not listed here and were generated freshly multiple times by lentiviral transduction.*

4.1.10. Bacterial strains

For cloning purposes DH5α (Life Technologies) and Stellar (Clontech) *E.coli* strains were used. Propagation of vectors containing gateway-cassettes was performed in *ccdB*-resistant db3.1 bacteria (Life Technologies).

4.2. Molecular biology methods

4.2.1. Molecular cloning

Plasmid cloning and purification was performed using standard procedures (Sambrook et al., 2001). For traditional cloning restriction enzymes (NEB), *rapid*™ alkaline phosphatase (Roche) and T4 DNA ligase (Roche) was used. Cloning PCRs were performed using *Phusion*™ polymerase (NEB) or *Q5*™ polymerase (NEB) according to the manufacturers' protocols. DNA was purified using commercial column purification kits (Promega, Macherey-Nagel, Qiagen). For subcloning of cDNAs into different vectors, *Gateway*™ cloning (Life Technologies) was performed according to the manufacturer's protocol. Many different PCR fragments or multiple fragments were cloned in one step by *Infusion*™ cloning (Clontech)

according to the manufacturer's protocol or a self-made *Gibson assembly*™ cloning mix (NEB) (Gibson et al., 2010; Gibson et al., 2009; Hamilton et al., 2007). Gibson assembly mix was generated and aliquoted after combining 219 µl water, 100 µl 5x Gibson assembly buffer (450 mM Tris/Cl pH 7.5, 25 % (w/v) PEG 8000, 50 mM $MgCl_2$, 50 mM DTT, 1 mM of each dNTP, 5 mM NAD), 0.2 µl T5 exonuclease (10 U/µl) (NEB), 6.25 µl Phusion polymerase (2 U/µl) (NEB) and 50 µl Taq Ligase (40 U/µl NEB). Aliquots of 15 µl were stored at -80°C. 5 µl of purified DNA fragment mix was added to defrosted aliquots of Gibson assembly mix and cloning was performed as described elsewhere (Gibson et al., 2010; Gibson et al., 2009).

4.2.2. RNA purification and cDNA synthesis

For RNA purification, cells were lysed in RLT buffer (Qiagen) supplemented with 2-mercaptoethanol as instructed by the manufacturer and RNA was purified using *RNEasy*™ purification (Qiagen). Traces of DNA were removed by including a DNaseI (Qiagen) on-column digestion step as suggested by the manufacturer's protocol.

Between 0.5 µg and 5 µg total RNA was reverse transcribed into complementary DNA (cDNA) using *Superscript III*™ (Life Technologies) and random hexameric primers (NEB) according to the manufacturer's protocol. Reactions that were not treated with reverse transcriptase enzyme were used as a negative control to check for DNA background.

For strand-specific reverse transcription, a single strand-specific primer was used for reverse transcription.

4.2.3. Semi-quantitative reverse transcriptase PCR

Semi-quantitative PCR was performed by using *Jumpstart*™ polymerase (Sigma Aldrich) or Taq Polymerase (5-Prime) using primers against the respective cDNA targets according to the manufacturers' protocols. PCR-Products were visualizes using agarose gel electrophoresis and ethidium bromide (Roth) staining.

4.2.4. Quantitative PCR (qPCR)

The relative amount of cDNA after reverse transcription, the copy number of different alleles inside genomic DNA or enrichment of DNA after chromatin immunoprecipitation was measured by qPCR. PCR was carried out with the *Fast SYBR® Green Master Mix*™ (Applied Biosystems) in a *LightCycler480*™ (Roche) according to the *Fast SYBR Green Master Mix*™-protocol. Primers were evaluated for generating a single PCR product and for linear amplification in a wide range of DNA template dilutions (up to 1:10000). Every PCR-reaction was performed in a total volume of 10 µl in triplicates in a 384-well plate (Sarstedt). Two independent control genes (Gapdh and Actin) were used as reference genes for qRT-PCR experiments and geometric mean of reference C_t values was used as normalization as

described by others (Vandesompele et al., 2002). Three independent control genes were used for evaluating the copy-number of the *midline1* exons. For qRT-PCR of repetitive regions like IAP elements, negative control samples that were not treated with reverse transcriptase were used to control for genomic DNA background.

C_t-values were generated by the *LightCycler480*-Software (Roche) using the *2nd derivative max* function and fold changes were calculated using the $2^{-\Delta\Delta C}$ method.

4.2.5. Northern Blotting

Northern Blotting was performed using RNAse-free material, DEPC-treated water and RNAse-free reagents. 5-20 µg of total RNA was denatured in three volumes of RNA formaldehyde sample buffer (Ambion) at 65°C for 20 min and separated by a denaturing agarose gel (1 % agarose, 6,66 % formaldehyde) in 1x MOPS buffer (40 mM 3-(N-morpholino) propanesulfonic acid, 10 mM sodium acetate, 1 mM EDTA, adjusted to pH 7.0 using NaOH). RNA was blotted overnight on a positively-charged Nylon membrane (Roth) by capillary blot in 10x SSC (1.5M NaCl, 150 mM sodium citrate, pH 7.0). RNA was crosslinked to the membrane using a UV-autocrosslinker (Stratagene). The membrane was rehydrated and blocked for at least 30 min at 68°C using *QuickHyb*™ hybridization solution (Stratagene). 25 ng of radio-labeled DNA probes (see 4.2.6) were mixed with 25 µg salmon-sperm DNA, denatured at 95°C for 5 min and added to *QuickHyb*™ hybridization solution at 68°C for overnight hybridization with the membrane. Membranes were washed several times with 1xSSC 0.1 % SDS at 68°C under vigorous shaking until only low amounts of radioactivity could be detected at the edges of the membrane. X-ray films (Fuji) were exposed to the membrane in the presence of an intensifier screen for a few hours to several days at -80°C. Films were developed in a photo developer (Agfa).

4.2.6. Radiolabeling of DNA

DNA corresponding to the mRNA of interest were amplified by PCR from cDNA or isolated by restriction digest from plasmids containing the respective cDNA.

50 ng of dsDNA was radiolabeled using 5 µCi alpha-^{32}P -dCTP using the *PrimeIt (II)-Random Prime Labeling Kit*™ (Stratagene) according to the manufacturer's protocol and purified using *QuickSpin*™ Gel Filtration Columns (Roche). Radiolabeled DNA was stored at -20°C for up to three weeks.

4.2.7. DNA methylation analysis of genomic DNA

Genomic DNA was prepared using the *DNEasy Blood and Tissue Kit*™ (Qiagen) and subjected to bisulfite conversion using the *EpiTect Bisulfite Kit*™ (Qiagen) according to the manufacturer's protocol. *Jumpstart*™ Taq polymerase (Sigma Aldrich) was used to amplify the IAP-gag region and the Oct4 promoter sequence. PCR primers for bisulfite-converted

DNA were either designed using the MethPrimer tool (Li and Dahiya, 2002) or taken from previous studies (Rowe et al., 2013a).

PCR products were subcloned into *pBluescript-SK2+* vector (Stratagene) using *InfusionHD*™ cloning (Clontech) and analyzed by Sanger sequencing (Eurofins MWG Operon). Methylation analysis of sequencing data was performed using *BiQ Analyzer* (Bock et al., 2005).

4.3. Biochemical methods

4.3.1. Preparing whole cell protein extracts

Cells were washed once with 1xPBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 2 mM KH_2PO_4) and trypsinized using 1xPBS containing 1 x Trypsin/EDTA (PAA or Sigma Aldrich). Cells were resuspended in growth medium and counted using a *CasyCounter*™ (Roche). Equal amounts of cells were washed with 1xPBS and were resuspended in SDS-Buffer (50 mM Tris/Cl pH 7.5, 2 % SDS, 1x complete™ Roche Protease Inhibitor). Cell lysates were incubated at 98°C for 10 to 15 min on a heatblock and centrifuged for 15 min at 17000 x g at room temperature. Supernatants were used for western blotting. When required, protein concentration was measured using the *Pierce™ BCA Protein Assay Kit* (Thermo Scientific).

4.3.2. Preparing nuclear cell extracts by hypotonic lysis

Cells were washed once with 1xPBS and trypsinized using 1xPBS containing 1 x Trypsin/EDTA (PAA or Sigma Aldrich). Cells were resuspended in 1xPBS and counted using a *CasyCounter*™ (Roche). Cells were resuspended in at least five pellet volumes of swelling buffer A (10 mM HEPES/KOH pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM DTT and 0.5x complete protease inhibitors (Roche)). Cells were incubated for 10 min on ice, vortexed shorty and pelleted for 10 sec at 17.000 x g. Supernatant was discarded and nuclei were resuspended in around 30 µl buffer C per 1 Mio cells (20 mM HEPES/KOH pH 7.9, 25 % (v/v) glycerol, 300 mM NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 1x complete protease inhibitors (Roche)). Nuclear extract was incubated for 20 min on ice and centrifuged at 14000 x g for 15 min at 4°C. Supernatant was used for western blot or immunoprecipitation. If required, protein concentration was measured using the Bradford *Bio-Rad protein assay* (Bio-rad).

4.3.3. Immunoprecipitation of Trim28 for mass spectrometry

Cells were washed once with 1xPBS and trypsinized using 1xPBS containing 1 x Trypsin/EDTA (PAA or Sigma Aldrich). Cells were resuspended in 1xPBS and counted using a *CasyCounter*™ (Roche). 75 Mio mouse ES cells and 12 Mio MEF cells were resuspended in 1 ml DMEM and separated on a Ficoll gradient (20 % Ficoll, 80 mM Tris/Cl

pH 7.4, 8 mM $MgCl_2$, 8 mM $CaCl_2$, 1.6 % NP40, 1.28 % TritonX-100, 0.1 % DMSO). Nuclei were isolated by centrifugation through Ficoll using the following centrifugation protocol in a Hereaus #75006475 rotor: 400 rpm - 30 sec, 500 rpm – 30 sec, 600 rpm – 30 sec, 700 rpm – 30 sec, 800 rpm – 6 min.

Nuclei were resuspended in IP Buffer (50 mM Tris/Cl pH 7.5, 0.1 % NP40, 15 % Glycerol, 300 mM NaCl, 18 µM $ZnSO_4$, 0.5 mM DTT, 1 x complete protease inhibitor) and disrupted with two short pulses (20 % power, 3 sec) of a ultrasound sonifier (Branson). 1.5 µl (375 U) of *Benzonase*™ (Merck) was added and lysates were incubated at 37°C for 15 min to digest nucleic acids. Lysate was cleared by centrifugation at 17000 x g for 30 min at 4°C and supernatant was used for immunoprecipitation. A small aliquot was saved to use as an input control. Immunoprecipitation was carried out with 50 µl Flag-M2 affinity agarose (Sigma Aldrich) for 2 h at 6°C. Beads and immunoprecipitated material were transferred into fresh reaction tubes, washed three times with IP buffer and eluted in 2x Roti®-Load 1 sample buffer (Roth) for 5 min at 95°C.

4.3.4. Other immunoprecipitation protocols used in this study

Daxx, Trim28 and Atrx-immunoprecipitation experiments were performed as indicated in 4.3.3, except that *Benzonase*™-treated nuclear extracts were prepared by hypotonic lysis and buffer C (see 4.3.2) was used as an IP-buffer.

Co-immunoprecipitation of H3.3 by FLAG-immunoprecipitated DAXX mutants was performed as described elsewhere (Elsasser et al., 2012), except that RNase A was used instead of RNase I and beads were washed three times instead of two times after immunoprecipitation.

4.3.5. Western blotting

Roti®-Load 1 (Roth) Laemmli sample buffer was added to protein extracts. Samples were incubated at 95°C for 3 min, centrifuged for 2 min at 17000 x g at room temperature and loaded onto precast SDS gels (SERVA) or self-made SDS gels. After separation of proteins by SDS-PAGE, proteins were blotted onto PVDF membranes in a wet-blotting chamber (Bio-Rad) at 400 mA for 1.5 h at 6°C using ice-packs.

Membranes were blocked with 5 % skim milk (Santa Cruz or Heirler Cenovis) in 1xTBS (20 mM Tris/Cl, pH 7.6, 150 mM NaCl) for at least 30 min under mild agitation. Primary antibody incubation was performed between 1 h to 16 h in 1 % skim milk in 1xTBST (0.1 % Tween-20 in 1xTBS). Membranes were rinsed two times in 1xTBST and washed three times for 5 min in 1xTBST. Incubation with ECL-coupled secondary antibodies was performed for 1.5 h in 1 % skim milk in 1xTBST. Again, membranes were rinsed two times in 1xTBST and washed three times for 5 min in 1xTBST. Chemoluminescence signals were

detected using *Immobilon™ Western* kit (Millipore) and X-ray films (Fuji) according to the manufacturer's protocol.

4.3.6. Chromatin immunoprecipitation (ChIP)

Cells were washed once with 1xPBS and trypsinized using 1xPBS containing Trypsin/EDTA (PAA or Sigma Aldrich). Cells were resuspended in growth medium and counted. Twenty to thirty million cells were washed with 1xPBS and crosslinked in 10 ml of 1 % under mild agitation. For Atrx-ChIPs EGS (ethylene glycol bis[succinimidylsuccinate]) (Pierce - Thermo Scientific) was added to a final concentration of 2 mM to crosslink proteins for 30 min at room temperature before crosslinking with 1 % methanol-free formaldehyde in 1xPBS for 10 min at room temperature. Crosslinking was quenched by adding 2.5 M glycine to a final concentration of 125 mM. Crosslinked cells were washed two times with 1xPBS and cell pellets were either snap-frozen and stored at -80°C or directly processed.

Cells were resuspended in 1 ml lysis buffer (10 mM Tris/Cl pH 8.0, 1 mM EDTA, 0.5 mM EGTA, 0.25 % SDS, 0.1 % Sodium deoxycholate, 200 mM NaCl, 1x complete protease inhibitor) and sheared at 4°C to a chromatin size between 150 bp and 800 bp using focused-ultrasonification (Covaris). 10 % Triton-X100 was added to the lysates to a final concentration of 1 % and lysates were cleared by centrifugation at 14000 x g for 30 min at 4°C.

Nine volumes of IP buffer (20 mM Tris/Cl pH 8.0, 1 mM EDTA, 0.5 mM EGTA, 1 % Triton, 167 mM NaCl, 1x complete protease inhibitor) were added to one volume of lysate. 1 ml of diluted lysate was used for immunoprecipitation with 3-5 µg of antibody prebound to magnetic *Protein G Dynabeads™* (Life Technologies) over night. After over night incubation beads were transferred into fresh reaction tubes and washed three times with IP buffer and once with 1 x STE buffer (50 mM NaCl, 10 mM Tris/Cl pH 8.0, 1 mM EDTA). Immunoprecipitated DNA was eluted from beads in elution buffer (50 mM Tris/Cl pH 8.0, 50 mM $NaHCO_3$, 1x SDS, 10 mM EDTA) at 65°C for 30 min under vigorous shaking. DNA-Protein crosslinks were reversed by incubating the immunoprecipitated material or input chromatin for 4 to 16 h at 65°C in elution buffer supplemented with 200 mM NaCl. Immunoprecipitated material was then diluted by the addition of one volume 1 x TE buffer (10 mM Tris/Cl pH8.0, 1 mM EDTA) and incubated for 30 min with 10 µg/µl RNaseA (Sigma Aldrich) at 37°C and 2–4 h with 20 µg/ml Proteinase K at 56°C (Bioline). DNA was purified using *QIAquick™ PCR purification Kit* (Qiagen) according to the manufacturer's protocol and enrichment of specific DNA regions was analyzed by qPCR.

4.3.7. Silver staining of protein gels

Protein gels were fixed with 40 % ethanol/10 % acetic acid for 2 h or overnight and washed three times with 30 % ethanol for 20 min. Then, protein gels were treated with 0.02 % $Na_2S_2O_3$ for 1 min and washed three times with water for 20 seconds. Protein gels were treated with 0.2 % $AgNO_3$ for one hour in the dark and were washed again three times with water for 20 seconds. Ag^+ ions bound to proteins were then reduced to elemental silver for 5-10 min with a reducing developing buffer (3 % Na_2CO_3, 0.05 % formaldehyde, 0.0004 % $Na_2S_2O_3$). Reduction was quenched by washing the protein gels once with water and rinsing the gel for 5 min in 0.5 % glycine. Gels were washed once more with water for 30 min, photographed and stored at 4°C until the gel was cut for mass-spectrometry analysis.

4.3.8. Sample preparation for mass-spectrometry

To prepare samples for mass spectrometry, immunoprecipitated proteins were separated by gel electrophoresis and stained by silver staining. Gel lanes were cut into eight pieces corresponding to different ranges of molecular weights. The gel pieces were shredded into smaller fragments and transferred into 200 µl PCR tubes for further processing.

Gel pieces were incubated two times with 100 µl water, two times with 100 µl 20 mM NH_4CO_3 and three times with 100 µl acetonitril (10 min for every incubation step). Samples were then rehydrated and reduced with 50 µl 10 mM DTT/20 mM NH_4CO_3 for 1 h and alkylated with 50 µl 55 mM iodoacetamide/20 mM NH_4CO_3 for 30 min. Gel pieces were then washed 10 min with 100 µl 20 mM NH_4CO_3 and dehydrated again by incubating them three times for 10 min with 100 µl acetonitril. 10-15 µl of a 25 ng/µl solution of sequencing-grade Trypsin (Promega) in 20 mM NH_4CO_3 was added and gel pieces were incubated for 30 min at 4°C. Non-absorbed trypsin was removed and 50 µl 20 mM NH_4CO_3 was added before incubating the gel pieces over night at 37°C on a PCR-tube shaker. After overnight incubation 50 µl 50 % acetonitril, 0.25 % trifluoroacetic acid was added for 10 min and the supernatant containing trypsinized peptides was transferred into a fresh PCR tube. This extraction step was repeated once under the same conditions and twice using pure acetonitril. Supernatants containing extracted were pooled and peptides were dried in a speedvac centrifuge. Peptides were resolubilized in 10 µl of 0.1 % trifluoroacetic acid and stored at -20°C until final processing and loading on the liquid chromatography column coupled to an orbitrap mass spectrometer (Thermo Scientific).

The final processing, handling of the mass spectrometer and peptide analysis using the *Scaffold3*™ software was performed by Dr. Ignasi Forné.

4.3.9. DNA pull-down experiments and SILAC

DNA pull-down experiments have been carried out by Dr. Falk Butter as described earlier (Butter et al., 2010).

4.4. Cell biology methods

4.4.1. Cell culture of mES cells and MEF cells

Cell lines were cultured at 37°C, 5 % CO_2 under standard procedures as described earlier (Dambacher et al., 2012).

4.4.2. Transfection of mouse ES cells, MEF cells and HeLa cells

Mouse embryonic stem cell lines were transfected with *Jetprime*™ (Polyplus) in a 2:1 ratio of transfection reagent vs. plasmid according to the manufacturer's protocol. MEF cell lines were transfected using *Jetprime*™ (Polyplus), Lipofectamine2000™ (Life Technologies) or *MirusLT-1*™ (Mirus) according to the manufacturer's protocol. HeLa cell lines were transfected using *FugeneHD*™(Roche) using the manufacturer's protocol.

4.4.3. Transfection of HEK293T cells and lentivirus production

HEK293T cells were transfected using the calcium phosphate transfection method for producing lentivirus. HEK293T cells were never grown longer than fifteen passages and kept in MEF medium containing 500 µg/ml G418 (PAA). Cells were always maintained subconfluent until being used for virus production. Five million HEK293T cells were seeded onto 10 cm dishes in growth medium without G418. On the next day the medium was replaced with fresh growth medium without G418 one hour before transfection. For transfection, a mix of 24 µg plasmid DNA consisting of 8 µg lentiviral transfer vector, 8 µg psPAX2 and 8 µg pLP-eco-env was mixed with 120 µl of 2.5 M $CaCl_2$ and adjusted to 1200 µl with sterile water. Occasionally, the amount of lentiviral transfer vector was increased proportionally to its size when very large lentiviral transfer vectors (>10 kb) were used to ensure an equimolar ratio between packaging plasmids. Then, 1200 µl 2xHBS solution (HEPES 50 mM, NaCl 280 mM, Na_2HPO_4 1.5 mM, adjusted to pH 7.05 with NaOH) was added slowly and dropwise to the mix while vortexing it gently on a vortexer (Scientific Industries). The transfection mix was added immediately to the cells. Four to eight hours after transfection the medium which contained precipitates of calcium phosphate and DNA was removed, cells were washed with 1xPBS and fresh growth medium without G418 was added. Virus supernatant was harvested 48 h after transfection. Viral supernatants were harvested a second time 72 h after transduction when a maximum amount of lentiviral particles was required. Viral supernatants were centrifuged at 3000 x g for 5 min at room temperature to pellet cells and cellular debris before the supernatant was decanted, aliquoted, snap frozen in liquid nitrogen and stored at -80°C until further use.

For genome-wide shRNA screening, viral supernatants of GAG2-EGFP reporter virus and EGFP control virus were concentrated by polyethylenglycol precipitation. One volume of precipitation-solution (42.5 % (w/v) polyethylenglycol 8000, 1.33 M NaCl) was mixed with

three volumes of viral supernatant and incubated 30 min to overnight at 4°C. Virus was pelleted by centrifugation at 1,500 x g for 45 minutes at 4°C. The pellet was gently resuspended in 1/100th of the original volume of the supernatant in DMEM (Sigma Aldrich) and virus was stored at -80°C.

4.4.4. Lentiviral transduction

Mouse ES cells were seeded on gelatinized 6-well or 12-well dishes in ES medium containing 8 µg/ml *polybrene*™ (Hexadimethrinbromid H9268 – Sigma Aldrich) and viral supernatant was added to the cells. 6-well or 12-well plates were spun in a prewarmed centrifuge at 1000 x g for 1 h at 34°C to enhance viral transduction. After centrifugation the medium was carefully replaced by ES medium. Cells were either assayed by FACS or split two days after transduction. When required, selection antibiotics were added 48 h after transduction to generate stable cell lines. The concentrations used were experimentally determined for every individual cell line.

Concentrations used: 180 – 360 µg/ml G418 (PAA)
0.4 – 2 µg/ml puromycin (Sigma Aldrich)
150 µg/ml hygromycin B (PAA)

MEF and HeLa cells were transduced accordingly except for minor differences. MEF and HeLa cells were seeded one day before transduction on standard 6-well or 12-well dishes and different concentrations of selection antibiotics were used.

Concentrations used: 300 µg/ml hygromycin B (PAA)
1.2 – 2.5 µg/ml puromycin (Sigma Aldrich).

Virus titers of EGFP reporter lentivirus were compared by titrating viral supernatants. For this, different amounts of viral supernatant were used to transduce 100.000 T37 HeLa cells and the percentage of EGFP-positive was determined two days later with a *Canto*™ FACS counter (Becton Dickinson). The percentage of EGFP-positive cells was tried to be kept below 15 % to ensure a linear relationship between the amount of used virus and virus transduction. The number of infectious particles per volume was calculated with the following formula: **n x r / v** = transduction particles / ml

n = cell number at time of transduction, **r** = percentage of green cells 48 h post transduction, **v** = volume of viral supernatant used for transduction

4.4.5. Lentiviral knockdown experiments

Lentiviral shRNA sequences were taken from the TRC library (Moffat et al., 2006) or designed according to the TRC library guidelines (http://www.broadinstitute.org/rnai/public/).

Cloning of shRNA sequences into the lentiviral pLKO1 vector was performed as described earlier (Dambacher et al., 2012).

Cell lines were selected with puromycin 48 h after transduction. Knockdown cell lines were used for reporter assays starting at day four to six days after shRNA transduction. Knockdown was evaluated by qRT-PCR or western blotting at day four to six after shRNA transduction.

4.4.6. Genome-wide lentiviral shRNA screen and secondary screen

The two genome-wide shRNA screening has been performed as described in section 2.12. W9 MEFs and wt26 mouse ES cells were used for the two different screens, because they performed best in terms of transduction capability and signal to noise ratio. The amount of required shRNA lentivirus was predetermined by transducing both cell lines with a known concentration of pLKO1-Ubi-EGFP-puro transduction particles (Sigma Aldrich), followed by puromycin selection.

For the MEF/Zfp809 screen W9 cells were seeded on ten 15 cm dishes one day before transduction (1.8 Mio W9 cells /15 cm dish). Cells were transduced in 10 ml of growth medium containing 8 µg/ml *polybrene*™ with pooled Lentiplex™ shRNA viruses (Sigma Aldrich) with a multiplicity of infection < 0.3. Medium was changed the next day. 48 h after transduction transduced cells were selected with 1.2 µg/ml puromycin and split 1:2. Puromycin selection was continued until the end of the experiment. Four days after transduction cells were pooled and split on thirty 6-well plates (200.000 cells / well). On the next day cells were transduced with 1xZfp809-bs-EGFP reporter-lentivirus and EGFP control virus at an MOI of 1 and split onto thirty 15 cm dishes. Cells were maintained on thirty 15 cm dishes and after additional eight days EGFP-positive cells were sorted using a MoFlo cell sorter (Becton Dickinson) and genomic DNA was isolated using *DNEasy Blood and Tissue Kit*™ (Qiagen). Sequencing of stably integrated shRNA sequences was performed by the shRNA sequencing core facility at *Partners Healthcare* center (MA, USA).

The mES/GAG screen was performed with wt26 feeder-independent mouse ES cells growing on gelatinized tissue culture plates. Apart from using different EGFP reporter constructs the screening experiment was identical to the MEF/Zfp809 screen, except that cells were transduced directly after seeding and different time points were used for transduction, selection and sorting (see Figure 2.12). In addition, 0.4 µg/ml puromycin was used instead of 1.2 µg/ml puromycin to select stably transduced cells.

After preparation of genomic DNA, a small amount of DNA was used to PCR-amplify shRNA sequences that were enriched in the EGFP-positive cell fraction. PCR was performed using Q5 polymerase according to the manufacturer's protocol using pLKO1-specific primer

sequences (5'-cgagactagcctcgagc-3' and 5'-ctgcgagggtactagtgag-3' - PCR protocol: 98°C 1 min, 32x [98°C for 20 sec, 63°C for 20 sec, 72°C for 23 sec], 72°C for 5 min). PCR products were gel-purified and cloned into pLKO1 using *InfusionHD*™ (Clontech). 75 individual lentiviral shRNA vectors were isolated. This small candidate library prepared out of the primary screen was used for a secondary screen.

For secondary screening 293T cells were seeded on a 96-well plate. After one day 293T cells were co-transfected with lentiviral packaging plasmids and the individual shRNA transfer vectors using *Jetprime*™ (Polyplus). Two days after transfection viral supernatants were used to transduce wt26 cells seeded on a 96-well plate in the presence of 8 µg/ml *polybrene*™. Two days after transduction wt26 cells were selected with puromycin for stable integration of the shRNA-cassette. After another two days knockdown cell lines were separated onto two 96-well plates and were either transduced with a GAG2-EGFP reporter virus or an EGFP control virus. Silencing capability was calculated by normalizing the number of EGFP-positive cells from the GAG2-EGFP reporter virus to the EGFP control virus. ShRNA vectors resulting in a strong derepression of GAG2-silencing were sequenced using Sanger sequencing.

4.4.7. Luciferase Assay

For luciferase experiments MEF cells were seeded on a 96-well plate one day before transfection. 150 ng pGL3-5xUAS-firefly plasmid, 7.5 ng Gal4-fusion plasmid and 7.5 ng renilla plasmid were transfected into MEF cells using Lipofectamin2000 (Life Technologies) according to the manufacturer's protocol and luciferase activity was monitored using the *Dual-Luciferase® Reporter Assay System*™ (Promega). Luciferase experiments were performed in triplicates.

4.4.8. FACS counting and FACS sorting

The number of EGFP-positive cells was measured using a *Canto*™ FACS counter (Becton Dickinson) normally two days after lentiviral reporter transduction or transfection of EGFP plasmids. Cells were washed once with 1xPBS and trypsinized using 1xPBS containing Trypsin/EDTA (PAA or Sigma Aldrich). Cells were resuspended, stored on wet ice and immediately used for FACS analysis.

For genome-wide shRNA screening, cells were stored in growth medium on ice until being sorted using a *MoFlo*™ FACS sorter (Becton Dickinson). Cell sorting was performed by Dr. Joachim W. Ellwart at Helmholtz Zentrum Munich.

4.4.9. Measurement of DNA-content by PI-FACS

Cells were washed once with 1xPBS, trypsinized using 1xPBS containing Trypsin/EDTA (PAA or Sigma Aldrich), resuspended in growth medium and counted using a *Casy*

Counter™ (Roche). An equal number of cells of every cell line was washed twice in 1xPBS and resuspended in 1xPBS. -20°C cold methanol was added slowly to the cell suspension under very gentle vortexing until reaching a final concentration of 70 % methanol (v/v). Cells were fixed for 1 h at 4°C or over night at -20°C. After fixation, cells were washed once with 1xPBS and were resuspended in PI staining buffer (40 µg/ml propidium iodide, 3.8 mM sodium citrate in 1xPBS, 10 µg/ml RNase A). After incubation for 1 h at 37°C, cells were analyzed using a *Canto*™ FACS counter (Becton Dickinson) using the PE-channel in linear mode. Data analysis was performed using *FlowJo*™ software (TreeStar).

4.4.10. Generating knockout cells using CRISPR/Cas9

Knockout of genes in mouse ES cells was performed by RNA-guided Cas9 nucleases using the pX330 plasmid (Cong et al., 2013; Hsu et al., 2013). Guide RNA sequences were manually identified close to PAM sequences (5'-NGG-3') in the genome or designed using an online bioinformatic webinterface (http://crispr.mit.edu/) (Hsu et al., 2013). DNA oligonucleotides corresponding to the guide RNA sequence were annealed and ligated into Bbs1 cut pX330 plasmid as described elsewhere (Cong et al., 2013).

300.000 mouse ES were seeded one day before transfection on 6-wells and were transfected using *Jetprime*™ (Polyplus) reagent.

Transfection protocol: 200 µl *Jetprime*™ buffer, 3.1 µg of pX330 plasmid, 0.3 µg of a plasmid with puromycin resistance (pLenti6-EF1a-FLAG IRES PURO) and 6.6 µl *Jetprime*™ reagent were mixed and vortexed shortly. After incubation for 15 min at room temperature the transfection mix was added to the cells and the medium was changed after 8 to 16 h.

Two days after transfection highly transfected cells were selected with 1.5 µg/ml puromycin for 16 to 24 h. Cell were then washed with medium without puromycin and cell cultivation was continued without puromycin. This short selection step significantly enriches for cells carrying mutations (Wang et al., 2013).

Transfected cells were seeded sparsely to allow picking of individual colonies. Clonal cell lines were analyzed by western blotting or by PCR and Sanger sequencing.

4.4.11. Recombinase-mediated cassette exchange (RMCE) and inducible expression

HA36 mouse ES cells that carry a selection cassette allowing positive and negative selection were provided by Prof. Dr. Dirk Schübeler (FMI Basel) and have been described elsewhere (Baubec et al., 2013; Lienert et al., 2011b). Recombinase-mediated cassette exchange was performed as described earlier (Baubec et al., 2013; Lienert et al., 2011b), except for minor changes in the protocol.

HA36 cells were maintained in 150 µg/ml hygromycinB (PAA) until transfection and transfected using *Jetprime*™ transfection reagent (Polyplus).

Transfection protocol: 500 µl *Jetprime*™ buffer, 8 µg of L1-poly-1L/TetO-EGFP-T2A-Zeo-GAG2.22-pA, 1.5 µg of pIC-Cre, 0.5 µg pCAG-Cre and 20 µl *Jetprime*™ reagent were mixed and vortexed shortly. After incubation for 15 min at room temperature the transfection mix was added to the cells and the medium was changed after additional 16 h. 70 h after transfection 3 µM Ganciclovir (Invivogen) was added to the medium for the next six days in order to select for recombined cell clones. Clonal cell lines were tested for successful integration of the reporter by PCR. Inducible expression of the EGFP reporter was tested by doxycyclin treatment and FACS analysis after transducing the cells with a lentivirus containing a reverse tet-transactivator (plasmid # 1308).

5. Abbreviations

aa	amino acids
Atf7ip	Activating transcription factor 7-interacting protein 1 (different name for mAM, Mcaf1)
Atrx	Alpha thalassemia/mental retardation syndrome X-linked
bp	base pair
BS	binding site
Cas9	CRISPR-associated endonuclease 9
Cbx	chromobox protein homolog (a family of proteins)
cDNA	complementary DNA
ChIP	chromatin immunoprecipitation
CMV/LTR→	cytomegalovirus/long terminal repeat hybrid promoter (a DNA sequence element of lentiviral vectors)
Co-IP	co-immunoprecipitation
CpG	a dinucleotide of cytosine and guanine (a substrate for DNA methylation)
CRISPR	clustered regularly interspaced short palindromic repeats
Daxx	Death domain-associated protein protein 6
dko	double knockout
DNA	deoxyribonucleic acid
eco M-MLV env	ecotropic envelope protein of the Moloney-Murine-Leukemia-Virus
EGFP	enhanced green fluorescent protein (a recombinant protein)
EGFP+	enhanced green fluorescent protein positive
FACS	fluorescence-activated cell sorting/counting
FITC channel	Fluorescein isothiocyanate channel (FACS channel detecting green fluorescence)
G418	Geneticin (a selection antibiotic)
GAG	group-specific antigen
GAG2.22	sequence element of the *group specific antigen* coding region of IAP-Ez retrotransposons that is capable of recruiting heterochromatic repression in mES cells
Gal4-DBD	DNA-binding domain of the Gal4 transcription factor

gRNA	guide RNA
H3.3	histone H3.3
H3f3a and H3f3b	mouse genes encoding the histone H3.3 protein
H3K27ac	acetylation of Lysine 27 of Histone H3
H3K4me1	mono-methylation of Lysine 4 of Histone H3
H3K9me1,2 or 3	Histone 3 Lysine 9 monomethylation, dimethylation or trimethylation
H4K20me1,2 or 3	Histone 4 Lysine 20 monomethylation, dimethylation or trimethylation
HA	hemagglutinin (a polypeptide tag)
hEF1α →	human elongation factor 1 alpha promoter
HEK293T	human embryonic kidney cell line 293 transformed with the large-T-antigen of SV40 virus
HeLa	a human cervix carcinoma cell line
Hp1	heterochromatin protein 1 (in mammals three genes encode for Hp1 proteins – Cbx1, Cbx3 and Cbx5)
HygR-HSV-TK	a recombinant selection marker generated by the fusion of the hygromycin resistance gene and the thyimidin kinase of Herpes Simplex Virus allowing positive and negative antibiotic selection
IAP	Internal A-type Particle (a mouse LTR retrotransposons subclass)
IAP-E	Internal A-type Particle containing an envelope gene
IP	immunoprecipitation
IRES	internal ribosomal entry site
Kap1	KRAB-associated protein 1 (alternative name for Trim28)
ko	knockout
LC-MS/MS	liquid chromatography coupled to tandem mass spectrometry
LTR	long terminal repeat (a big subclass of retrotransposons) or regulatory sequence element actually flanking the retrotransposon sequence
M18bp1	Mis18-binding protein 1
mAM	murine Atf7a associated modulator (different name for Atf7ip, Mcaf1)
Mcaf1	MBD1-containing chromatin-associated factor 1 (different name for mAM, Atf7ip)
MEF	mouse embryonic fibroblasts

mES cells	mouse embryonic stem cells
M-MLV	Moloney-Mouse Leukemia Virus (a mouse retrovirus)
MOI	multiplicity of infection (value that accounts for the ratio of transduction particles relative to the number of cells)
mPGK	mouse phosphoglycerate kinase
mRNA	messenger-RNA
neoR	neomycin resistance gene (gene encoding an enzyme conferring resistance to neomycin or its derivates like G418)
Oct-4	octamer-binding transcription factor 4 (Pou5f1)
ORF	open reading frame
PAM	protospacer adjacent motif
PCR	polymerase chain reaction
PE channel	Phycoerythrin channel (FACS channel detecting yellow fluorescence)
PI	propidium-iodide
PML	promyelocytic leukemia (a protein that can form distinct nuclear aggregates called “PML nuclear bodies”)
Pol	polymerase
Polrmt	mitochondrial DNA-directed RNA polymerase
Pro	protease
qPCR	quantitative PCR
RBCC domain	N-terminal RING finger/B-box/coiled coil (a protein domain of Trim28)
Rev	Regulator of Expression of Virion Proteins (an HIV-1 protein)
RGN	RNA-guided nucleases
RMCE	recombinase-mediated cassette exchange
RNA	ribonucleic acid
RRE	rev-response element (a DNA sequence element of lentiviral vectors)
cPPT	central polypurine tract (a DNA sequence element of lentiviral vectors)
RT	reverse transcriptase
RT-qPCR	reverse transcriptase quantitative PCR

rtTA	reverse tet-transactivator (a recombinant protein used for the generation of inducible promoters)
Scr	scrambled
sgRNA	small-guide RNA
shRNA	short hairpin RNA
shRNA-seq	short hairpin RNA- sequencing
SILAC	stable isotope labeling by amino acids in cell culture
SIM	Sumo-interaction motif
SINΔU3 LTR	self-inactivating long terminal repeat promoter lacking the U3 sequence (a DNA sequence element of lentiviral vectors)
SV40ori	origin of replication of the SV40 virus
TAC	thesis advisory committee
T2A	2a-type self-cleavage peptide of the Thosea asigna virus
Tat	Trans-Activator of Transcription (an HIV-1 protein)
Tia1	T-cell-restricted intracellular antigen-1
Tif1b	Transcription intermediary factor 1-beta (alternative name for Trim28)
TRE	Tet response element: Tet operator sequences upstream of a minimal cytomegalovirus promoter (an inducible promoter sequence)
Trim28	Tripartite motif containing 28 (also known as Kap1, Tif1β)
UAS	upstream activating sequence (binding site of Gal4)
Uniprot-ID	identification code of a protein in the uniprot database
UTR	untranslated region
zeoR	resistance gene conferring resistance to Zeocin™
Zfp809	Zinc finger Protein 809

6. Acknowledgements

I'd like to thank Prof. Dr. Gunnar Schotta for supervision of my thesis and allowing me such a high degree of freedom in my research. I'd like to thank my other TAC-committee members Prof. Dr. Peter B. Becker and Dr. Tobias Straub for their time and their valuable advice on my project.

Special thanks go to Katharina Schmidt for her outstanding collaboration in this project. I would also like to thank Dr. Ignasi Forné and Dr. Falk Butter who collaborated with me on mass-spectrometry-based screening experiments and Dr. Joachim Ellwart for cell sorting. I'd like to thank Dr. Peer-Hendrik Kuhn, Prof. Dr. Dr. h.c. Christian Haass and Prof. Dr. Dieter Edbauer for plasmids and introduction into Biosafety Level 2 regulations.

I would like to thank PIs, Postdocs and PhD students of the molecular biology department here at the Adolf-Butenandt-Institute, who made it such a beautiful place and the time being a graduate student so unforgettable. I would like to thank all Socca5 players for the outstanding football matches and Prof. Dr. Axel Imhof and Dr. Lars Israel for organizing the events.

Special thanks go to all Schotta lab members for scientific discussions and for standing at my side during the easy and hard times of my PhD work. Especially intensive discussions and precious moments with Dr. Matthias Hahn, Dr. Silvia Dambacher, Alessandra Pasquarella, Katharina Schmidt and Alexander Nuber will always have a place in my memories.

Thanks to Mutsch, Vatsch, Lejla, Markus, Nando, Bärbel, Matthias, Thomas and Rodi for always being the outstanding family support I need and making me the person I am! I would like to deeply thank Maria for being the love of my life and standing at my side! Your support cannot be expressed in words.

7. Curriculum Vitae

The CV is not accessible in the public version.

8. Appendix

Genes targeted by two to four independently scoring shRNAs in the 1xZfp809/MEF shRNA screen (genes targeted by 1 shRNA are available upon request)

Sequences encoded by the shRNA code can be obtained at www.sigmaaldrich.com (genes are sorted by the number of scoring shRNAs and in alphabetical order)

shRNA code	targeted gene	# of scoring shRNAs
TRCN0000071364	Trim28	3
TRCN0000071366	Trim28	3
TRCN0000071363	Trim28	3
TRCN0000027605	Gak	3
TRCN0000027663	Gak	3
TRCN0000027649	Gak	3
TRCN0000070170	Asna1	2
TRCN0000070172	Asna1	2
TRCN0000174364	Atl2	2
TRCN0000174949	Atl2	2
TRCN0000070086	Atp5g3	2
TRCN0000070083	Atp5g3	2
TRCN0000100102	BC053393	2
TRCN0000100100	BC053393	2
TRCN0000081436	Bpnt1	2
TRCN0000081433	Bpnt1	2
TRCN0000030679	Capn9	2
TRCN0000030681	Capn9	2
TRCN0000069880	Catsper1	2
TRCN0000069879	Catsper1	2
TRCN0000200374	Ccdc53	2
TRCN0000176446	Ccdc53	2
TRCN0000124012	Coq2	2
TRCN0000124010	Coq2	2
TRCN0000120845	Csn3	2
TRCN0000120842	Csn3	2
TRCN0000068246	Cxcl11	2
TRCN0000068245	Cxcl11	2
TRCN0000028750	Cxcr4	2
TRCN0000028678	Cxcr4	2
TRCN0000176612	Fam72a	2
TRCN0000178544	Fam72a	2
TRCN0000119660	Fgb	2
TRCN0000119658	Fgb	2
TRCN0000067181	Fgf18	2
TRCN0000067179	Fgf18	2
TRCN0000065926	Icam1	2
TRCN0000065924	Icam1	2
TRCN0000112345	Itm2a	2
TRCN0000112348	Itm2a	2
TRCN0000076642	Lpo	2
TRCN0000076640	Lpo	2
TRCN0000194167	Mios	2

shRNA code	targeted gene	# of scoring shRNAs
TRCN0000175108	Mios	2
TRCN0000125029	Mlec	2
TRCN0000125030	Mlec	2
TRCN0000124857	Mrgpra8	2
TRCN0000124855	Mrgpra8	2
TRCN0000075816	Oas2	2
TRCN0000075817	Oas2	2
TRCN0000111781	Pcif1	2
TRCN0000111784	Pcif1	2
TRCN0000091414	Plek2	2
TRCN0000091417	Plek2	2
TRCN0000078752	Rsph1	2
TRCN0000078751	Rsph1	2
TRCN0000120986	Rufy3	2
TRCN0000120982	Rufy3	2
TRCN0000123596	Safb2	2
TRCN0000123598	Safb2	2
TRCN0000070316	Slc14a2	2
TRCN0000070314	Slc14a2	2
TRCN0000080166	Smoc2	2
TRCN0000080163	Smoc2	2
TRCN0000110485	St6galnac1	2
TRCN0000110487	St6galnac1	2
TRCN0000119932	Tceanc	2
TRCN0000119935	Tceanc	2
TRCN0000102619	Tial1	2
TRCN0000102615	Tial1	2
TRCN0000124538	Tm7sf2	2
TRCN0000124537	Tm7sf2	2
TRCN0000108812	Tpm4	2
TRCN0000108813	Tpm4	2
TRCN0000124572	Vangl2	2
TRCN0000124570	Vangl2	2
TRCN0000084307	Zfp131	2
TRCN0000084303	Zfp131	2
TRCN0000084687	Zfp646	2
TRCN0000084686	Zfp646	2
TRCN0000096114	Zfp668	2
TRCN0000096115	Zfp668	2
TRCN0000088677		2
TRCN0000120396		2
TRCN0000023983		2
TRCN0000120394		2
TRCN0000023980		2
TRCN0000174122		2
TRCN0000088676		2
TRCN0000067570		2

Genes targeted by two to four independently scoring shRNAs in the GAG2/mES-cell shRNA screens (genes targeted by 1 shRNA are available upon request)

Sequences encoded by the shRNA code can be obtained at www.sigmaaldrich.com (genes are sorted by the number of scoring shRNAs and in alphabetical order)

shRNA code	targeted gene	# of scoring shRNAs
TRCN0000070989	Top2b	4
TRCN0000070992	Top2b	4
TRCN0000070991	Top2b	4
TRCN0000070990	Top2b	4
TRCN0000091125	Actrt2	3
TRCN0000091126	Actrt2	3
TRCN0000091127	Actrt2	3
TRCN0000108577	Dach2	3
TRCN0000108576	Dach2	3
TRCN0000108575	Dach2	3
TRCN0000070846	Hoxb3	3
TRCN0000070844	Hoxb3	3
TRCN0000070843	Hoxb3	3
TRCN0000106089	Rundc3b	3
TRCN0000106087	Rundc3b	3
TRCN0000106088	Rundc3b	3
TRCN0000086331	Zbtb20	3
TRCN0000086328	Zbtb20	3
TRCN0000086329	Zbtb20	3
TRCN0000189745	1110012L19Rik	2
TRCN0000201569	1110012L19Rik	2
TRCN0000034371	A430107D22Rik	2
TRCN0000034373	A430107D22Rik	2
TRCN0000113480	Abcf3	2
TRCN0000113482	Abcf3	2
TRCN0000110442	Abo	2
TRCN0000110444	Abo	2
TRCN0000124737	Acbd3	2
TRCN0000124736	Acbd3	2
TRCN0000113267	Acot5	2
TRCN0000113268	Acot5	2
TRCN0000085625	Aff1	2
TRCN0000085623	Aff1	2
TRCN0000095170	AI854703	2
TRCN0000095173	AI854703	2
TRCN0000090848	Ank2	2
TRCN0000090852	Ank2	2

shRNA code	targeted gene	# of scoring shRNAs
TRCN0000082088	Ankrd6	2
TRCN0000082090	Ankrd6	2
TRCN0000113120	Ap4s1	2
TRCN0000113124	Ap4s1	2
TRCN0000100371	Arf1	2
TRCN0000100373	Arf1	2
TRCN0000100776	Arfgap3	2
TRCN0000100779	Arfgap3	2
TRCN0000110025	Arhgef7	2
TRCN0000110027	Arhgef7	2
TRCN0000124730	Asah2	2
TRCN0000124731	Asah2	2
TRCN0000086218	Asb8	2
TRCN0000086219	Asb8	2
TRCN0000110635	Aspm	2
TRCN0000110636	Aspm	2
TRCN0000112870	Atp1b2	2
TRCN0000112872	Atp1b2	2
TRCN0000197685	Bend3	2
TRCN0000197391	Bend3	2
TRCN0000096898	Bzw1	2
TRCN0000096895	Bzw1	2
TRCN0000103598	Caskin2	2
TRCN0000103597	Caskin2	2
TRCN0000071051	Cbx5	2
TRCN0000071050	Cbx5	2
TRCN0000112846	Chac2	2
TRCN0000112847	Chac2	2
TRCN0000089301	Cklf	2
TRCN0000089300	Cklf	2
TRCN0000068673	Cnga2	2
TRCN0000068674	Cnga2	2
TRCN0000089943	Col24a1	2
TRCN0000089944	Col24a1	2
TRCN0000125698	Coro7	2
TRCN0000125696	Coro7	2
TRCN0000032880	Cpb2	2
TRCN0000032881	Cpb2	2
TRCN0000110597	Cpt1a	2
TRCN0000110599	Cpt1a	2
TRCN0000109565	Creld2	2
TRCN0000109568	Creld2	2
TRCN0000106327	Crispld1	2
TRCN0000106325	Crispld1	2
TRCN0000080152	Cst8	2

shRNA code	targeted gene	# of scoring shRNAs
TRCN0000080148	Cst8	2
TRCN0000032651	Ctsg	2
TRCN0000032650	Ctsg	2
TRCN0000108776	Diap3	2
TRCN0000108775	Diap3	2
TRCN0000024848	Dlg3	2
TRCN0000024844	Dlg3	2
TRCN0000008543	Dnajc1	2
TRCN0000008541	Dnajc1	2
TRCN0000120976	Dnajc17	2
TRCN0000120974	Dnajc17	2
TRCN0000115464	Dnajc9	2
TRCN0000115462	Dnajc9	2
TRCN0000190448	Dos	2
TRCN0000189591	Dos	2
TRCN0000105874	Efs	2
TRCN0000105870	Efs	2
TRCN0000031369	Egfbp2	2
TRCN0000031372	Egfbp2	2
TRCN0000201234	Eif4e2	2
TRCN0000190491	Eif4e2	2
TRCN0000009806	Eif4g2	2
TRCN0000009810	Eif4g2	2
TRCN0000181801	Fam123b	2
TRCN0000198974	Fam123b	2
TRCN0000076322	Farsb	2
TRCN0000076318	Farsb	2
TRCN0000191859	Fermt2	2
TRCN0000191858	Fermt2	2
TRCN0000100110	Gab1	2
TRCN0000100111	Gab1	2
TRCN0000041476	Gapdhs	2
TRCN0000041477	Gapdhs	2
TRCN0000194469	Gp1ba	2
TRCN0000175201	Gp1ba	2
TRCN0000102902	Grid2	2
TRCN0000102900	Grid2	2
TRCN0000092857	Hist2h3c1	2
TRCN0000092853	Hist2h3c1	2
TRCN0000085718	Homez	2
TRCN0000085721	Homez	2
TRCN0000066856	Il17d	2
TRCN0000066854	Il17d	2
TRCN0000124699	Inadl	2
TRCN0000124700	Inadl	2

shRNA code	targeted gene	# of scoring shRNAs
TRCN0000023475	Jak3	2
TRCN0000023477	Jak3	2
TRCN0000125024	Kcnc2	2
TRCN0000125026	Kcnc2	2
TRCN0000069278	Kcnj14	2
TRCN0000069280	Kcnj14	2
TRCN0000068796	Kctd4	2
TRCN0000068797	Kctd4	2
TRCN0000032583	Kif24	2
TRCN0000032582	Kif24	2
TRCN0000032297	Klk10	2
TRCN0000032294	Klk10	2
TRCN0000080263	Kng1	2
TRCN0000080266	Kng1	2
TRCN0000089735	Krt79	2
TRCN0000089737	Krt79	2
TRCN0000098438	Krtap9-1	2
TRCN0000098436	Krtap9-1	2
TRCN0000105615	Lcn9	2
TRCN0000105619	Lcn9	2
TRCN0000097226	Lpxn	2
TRCN0000097227	Lpxn	2
TRCN0000106459	Lrrc66	2
TRCN0000106455	Lrrc66	2
TRCN0000087550	Lxn	2
TRCN0000087551	Lxn	2
TRCN0000042656	Mcf2	2
TRCN0000174066	Mcf2	2
TRCN0000124920	Mlxip	2
TRCN0000124919	Mlxip	2
TRCN0000102521	Mrpl12	2
TRCN0000102520	Mrpl12	2
TRCN0000084582	Mxd1	2
TRCN0000084581	Mxd1	2
TRCN0000089824	Myo18b	2
TRCN0000089825	Myo18b	2
TRCN0000039343	Myst4	2
TRCN0000039341	Myst4	2
TRCN0000012021	Nceh1	2
TRCN0000012018	Nceh1	2
TRCN0000173546	Ncoa6	2
TRCN0000173430	Ncoa6	2
TRCN0000097694	Ndst4	2
TRCN0000097691	Ndst4	2
TRCN0000113665	Negr1	2

shRNA code	targeted gene	# of scoring shRNAs
TRCN0000113667	Negr1	2
TRCN0000119701	Nlrp14	2
TRCN0000119697	Nlrp14	2
TRCN0000188932	Olfr202	2
TRCN0000185356	Olfr202	2
TRCN0000202878	Olfr849	2
TRCN0000202754	Olfr849	2
TRCN0000174685	Orc5l	2
TRCN0000175246	Orc5l	2
TRCN0000125497	Ormdl2	2
TRCN0000125495	Ormdl2	2
TRCN0000085366	Ovol1	2
TRCN0000085367	Ovol1	2
TRCN0000193830	Parn	2
TRCN0000173297	Parn	2
TRCN0000112665	Parva	2
TRCN0000112668	Parva	2
TRCN0000111771	Pdia6	2
TRCN0000111772	Pdia6	2
TRCN0000197570	Piwil4	2
TRCN0000198921	Piwil4	2
TRCN0000098776	Pla2g12a	2
TRCN0000098775	Pla2g12a	2
TRCN0000097292	Pla2g4c	2
TRCN0000097289	Pla2g4c	2
TRCN0000027596	Plk3	2
TRCN0000027594	Plk3	2
TRCN0000198533	Pof1b	2
TRCN0000197655	Pof1b	2
TRCN0000101207	Ppil4	2
TRCN0000101205	Ppil4	2
TRCN0000012626	Ppp2r1a	2
TRCN0000012623	Ppp2r1a	2
TRCN0000089083	Prg4	2
TRCN0000089084	Prg4	2
TRCN0000088570	Prkcdbp	2
TRCN0000088572	Prkcdbp	2
TRCN0000032537	Prss23	2
TRCN0000032534	Prss23	2
TRCN0000032159	Psma5	2
TRCN0000032163	Psma5	2
TRCN0000054532	Psmb5	2
TRCN0000031790	Psmb5	2
TRCN0000120273	Psmc3ip	2
TRCN0000120274	Psmc3ip	2

shRNA code	targeted gene	# of scoring shRNAs
TRCN0000071253	Pttg1	2
TRCN0000071254	Pttg1	2
TRCN0000040757	Rmnd5a	2
TRCN0000040755	Rmnd5a	2
TRCN0000041188	Rnf5	2
TRCN0000041189	Rnf5	2
TRCN0000039431	Rnf8	2
TRCN0000039432	Rnf8	2
TRCN0000096450	Rsl1	2
TRCN0000096451	Rsl1	2
TRCN0000119367	Rtn1	2
TRCN0000119371	Rtn1	2
TRCN0000121384	Sash1	2
TRCN0000121385	Sash1	2
TRCN0000032466	Scpep1	2
TRCN0000032468	Scpep1	2
TRCN0000119389	Serinc1	2
TRCN0000119387	Serinc1	2
TRCN0000105563	Sfxn3	2
TRCN0000105561	Sfxn3	2
TRCN0000098355	Sgcz	2
TRCN0000098358	Sgcz	2
TRCN0000094527	Siglece	2
TRCN0000094526	Siglece	2
TRCN0000068085	Siglech	2
TRCN0000068087	Siglech	2
TRCN0000112081	Skiv2l	2
TRCN0000112084	Skiv2l	2
TRCN0000176761	Slain1	2
TRCN0000198389	Slain1	2
TRCN0000079319	Slc15a2	2
TRCN0000079320	Slc15a2	2
TRCN0000069017	Slc35a5	2
TRCN0000069014	Slc35a5	2
TRCN0000109151	Snrpa1	2
TRCN0000109154	Snrpa1	2
TRCN0000080182	Sparcl1	2
TRCN0000080178	Sparcl1	2
TRCN0000079971	Spock1	2
TRCN0000079970	Spock1	2
TRCN0000103288	Sptlc3	2
TRCN0000103285	Sptlc3	2
TRCN0000085871	Srxn1	2
TRCN0000085868	Srxn1	2
TRCN0000110384	St8sia2	2

shRNA code	targeted gene	# of scoring shRNAs
TRCN0000110382	St8sia2	2
TRCN0000079998	Stfa3	2
TRCN0000079999	Stfa3	2
TRCN0000115527	Stmn4	2
TRCN0000115530	Stmn4	2
TRCN0000108876	Syne1	2
TRCN0000108878	Syne1	2
TRCN0000054843	Tank	2
TRCN0000054844	Tank	2
TRCN0000106072	Tbc1d2b	2
TRCN0000106073	Tbc1d2b	2
TRCN0000114127	Tcl1b5	2
TRCN0000114129	Tcl1b5	2
TRCN0000085951	Tead3	2
TRCN0000085950	Tead3	2
TRCN0000089893	Tekt3	2
TRCN0000089894	Tekt3	2
TRCN0000071302	Terf1	2
TRCN0000071299	Terf1	2
TRCN0000119341	Tmem165	2
TRCN0000119338	Tmem165	2
TRCN0000124294	Tmem54	2
TRCN0000124298	Tmem54	2
TRCN0000177998	Tmem85	2
TRCN0000178319	Tmem85	2
TRCN0000101965	Tnpo3	2
TRCN0000101968	Tnpo3	2
TRCN0000081817	Tox3	2
TRCN0000081815	Tox3	2
TRCN0000098646	Ttll13	2
TRCN0000098647	Ttll13	2
TRCN0000040914	Ube2v2	2
TRCN0000040915	Ube2v2	2
TRCN0000092675	Ubqln4	2
TRCN0000092676	Ubqln4	2
TRCN0000030819	Usp33	2
TRCN0000030821	Usp33	2
TRCN0000030739	Usp4	2
TRCN0000030742	Usp4	2
TRCN0000104975	Vmn2r-ps105	2
TRCN0000104976	Vmn2r-ps105	2
TRCN0000183172	Wipf1	2
TRCN0000195856	Wipf1	2
TRCN0000081652	Zfp14	2
TRCN0000081651	Zfp14	2

shRNA code	targeted gene	# of scoring shRNAs
TRCN0000096581	Zfp688	2
TRCN0000096582	Zfp688	2
TRCN0000088182		2
TRCN0000088181		2
TRCN0000087404		2
TRCN0000087406		2

Top enriched proteins in DNA pull-down by Dr. Falk Butter (Spectionmycin vs. GAG2.22)

Gene Names	Uniprot	Ratio for	Ratio rev
Tcfap2d	Q91ZK0		22,1
Gm1103;Zbtb2;mKIAA1483	Q3V1Q5;Q3V3W4;Q505G7;Q69ZI2	0,0	19,0
Tcfap2c;Tfap2c;	Q61312;A2APA8;Q3ULB3;Q3URU7;Q99L72	0,0	18,5
Cnbp;Cnbp1;Znf9;mCG_130858	P53996-3;P53996;Q3ULK8;P53996-2;Q3U935;P53996-1;Q5QJQ9;Q3U5V2	0,0	17,5
Zbtb25;Zfp50	Q6NV93;Q9CSB1	0,1	17,1
Ap2tf;Tcfap2a;Tfap2a;mCG_5075;AP-2	P34056-3;P34056;Q8BPN4;P34056-4;P34056-1;P34056-2;Q3UL09;Q6LCW3	0,0	13,9
Zfp296;mCG_4870	Q4FZI6;Q9D7X0;Q9EPM0	0,0	13,3
Zic5;Opr	Q7TQ40;Q9CVS3;Q9EQW1	0,0	13,0
Zasc1;Zfp639;Znf639	Q99KZ6;Q80UZ3	0,1	12,8
Sfrs2;Pr264;Sfrs10;mCG_6836;RP23-468A19.7-001;IREBF2	Q8C671;Q62093;A2AA29;Q99MY4;Q99MY5;Q06477	0,1	9,7
L3mbtl3;Mbt1	Q8BLB7-1;Q8BLB7;Q8BLB7-2;Q8BMN8	0,1	9,4
		0,1	9,3
Msh6;Gtmbp	Q61061;Q6GTK8;Q8C2N9;Q9CRH0;P54276	0,1	8,5
Lin28;Lin28a;Tex17	Q8K3Y3		7,6
Ncl;Nuc	P09405;Q3TGR3;Q3TL52;Q3TT41;Q8CD23;Q8CE30;Q9CT46	0,1	7,0
Mta1	Q8K4B0;Q2KHS8	0,1	7,0
Mta1l1;Mta2	Q9R190;Q3TZP3;Q3UDZ8	0,1	6,9
		0,1	6,5
Rbap46;Rbbp7;RP23-436I3.1-002;RP23-436I3.1-003	Q60973;Q3UJI2;Q3UX53;Q8C5H3;A2AFI9;A2AFJ1	0,1	6,4
RP23-199A2.1-001;Vezf1;RP23-199A2.1-002	Q5SXC4;Q9Z162;Q5SXC3	0,1	5,7
Pura	P42669;Q8C6E9	0,2	5,6
Ezf;Gklf;Klf4;Zie;RP23-322L22.2-003	Q60793;B7ZCH2	0,1	5,6
Ctcf	Q61164;Q05CK6;Q3USR8;Q3UYZ8;Q3UZH8	0,2	5,4
Purb	O35295;Q3UTJ8;Q8BQK8	0,1	5,3
Mtf2;Pcl2;mPcl2	Q02395-1;Q02395;Q05C61;Q02395-2;Q924U2	0,2	5,2
Klf2;Lklf	B2RS60;Q3V293;Q9JLV7;Q60843	0,1	5,2
mCG_12245;Rbbp4;RP23-391E6.3-001;Rbap48	A2A875;Q60972	0,2	4,7
Gatad2b;mKIAA1150	Q8VHR5-1;Q8VHR5;A1L3S7;Q69ZQ5;Q6PAH8;Q8VHR5-2	0,1	4,5

Gene Names	Uniprot	Ratio for	Ratio rev
Gatad2a	Q8CHY6	0,1	4,3
Dppa4	Q8CCG4;Q9CWG7	0,3	4,2
Atf2;mCG_19373;	P16951-1;P16951;A2AQF2;P70299;Q8BN75;Q8CGB4;P16951-2;Q640L6;Q8BQY2;Q3UKS5;P16951-3;Q543G2;Q68FE3;Q8CBR9;A2AQF0	0,2	4,2
Sfrs3;Srp20;X16;mCG_21131	P84104-1;P84104;B2KF41;Q9D6W4;P84104-2;Q3U781;A2A4X6		4,1
Chd4;mKIAA4075	Q5DTP7;Q6PDQ2;Q8BM83;Q3U582;Q3UFM8;Q3ULG0;Q3V265;Q6IQY9;Q99JM0;Q9CTT2	0,2	4,0
Ddx36;Dhx36;Kiaa1488;Mlel1	Q8VHK9;A0JLR3;B2RQS6	0,1	4,0
Hmgts;Tfam	P40630-1;P40630;Q3TSW9;P40630-2;P97907	0,4	3,9
Zic2	Q62520;Q8BQC1		3,8
Hdac1;RP23-209C6.7-001	O09106;Q58E49	0,2	3,8
Hdac2;Yy1bp	P70288;Q3TMT1;Q3URA2;Q3UXH8;Q8BQ10	0,2	3,8
Eed	Q921E6-2;Q921E6;Q921E6-1;Q921E6-3	0,4	3,7
Atf7;mCG_17588	Q3TZR9;Q3US59;Q8R0S1;Q3U2X8	0,1	3,7
Bteb2;Iklf;Klf5	Q9Z0Z7;Q923C0	0,1	3,6
D11Ertd530e;Kiaa0160;Suz12	Q80U70;Q3URR6;Q3UX45;Q5XG68	0,3	3,6
Zbed6	Q3UMD3;Q8C1U4;Q8CBM9	0,2	3,4
Mta3;mKIAA1266	A4FTZ3;Q3UII8;Q3UKM9;Q6ZPV1;Q924K8-2;Q924K8;Q3U3A7;Q924K8-1;Q3TY62	0,2	3,4
Enx2;Ezh1	P70351-2;P70351;P70351-1	0,2	3,2
Enx1h;Ezh2;mKIAA4065;mCG_2028	Q61188-1;Q61188;Q3TZH6;Q571L5;Q99L74;Q61188-2;Q3U575;Q8C2I5;Q6AXH7	0,3	3,1
ENSMUSG00000053178;Gm9897;mCG_113542;Mterf	B9EJ57;Q05C75;Q8CHZ9	0,5	3,0
Zfp71-rs1;Zfp738	Q8BY64;Q91W94;B8JJX9;B8JJY0;Q60915;Q4QQP3;Q7M6X0;Q8BLX6;B8JJX8;Q8BSG2	0,3	2,9
Ing5	Q9D8Y8-1;Q9D8Y8;Q9D8Y8-2;Q9D8Y8-3		2,8
Myst4	Q8BRB7-1;Q8BRB7;Q3UH94;Q501M5;Q8BRB7-2	0,4	2,8
Hrs;Sfrs5;mCG_7614	O35326;Q9D8S5;Q5U448;Q640L9	0,3	2,8
Rbm28	Q8CGC6;Q8R2W6	0,4	2,7
Zfx;RP23-269L6.2-001;Zfa;RP23-269L6.2-003	P17012;B1ASD1;Q3UM56;Q3URA1;Q99NJ4;B1ASD2;B7ZN32;P23607;Q8CDV8	0,7	2,7
Dppa2;Phsecrg1;mCG_127782	Q9CWH0;B2RQ54	0,5	2,6
Blm	O88700	0,3	2,6
Zfp281	Q3U063;Q3V3S9;Q4FK52;Q99LI5	0,3	2,6
Cxxc5	Q91WA4	0,2	2,5
Moz;Myst3	Q8BZ21;A0PJC5;Q3UFW4;Q3UPM9;Q3V1G6;Q8C6L0	0,3	2,5
C80731;Zfp568	Q3UPK4;Q0VGV0;Q3TLJ8		2,5

Top enriched proteins in DNA pull-down by Dr. Falk Butter (GAG2.22 vs. point mutant)

Gene Names	Uniprot	Ratio for	Ratio rev
Ubtf	A2AWT5;P25976;A2AWT7;A2AWT6;P25976-2	2,55	0,32

Proteomics interaction list of Trim28 (experiment performed with Ignasi Forné)

Peptides were identified with *Scaffold3*™ - Not enriched peptides and peptides corresponding to splicing and ribosomal proteins have been removed from the list and are available upon request.

Accession Number	Molecular Weight	Peptides in Mock	Peptides (Trim28-RBCC)	Peptides Trim28 FL mESC	Peptides FL (MEF cells)
ABT1_MOUSE	31 kDa	2	3	7	5
ADT1_MOUSE	33 kDa	2	5	4	5
ADT2_MOUSE	33 kDa	9	14	18	11
AAAS_MOUSE	59 kDa	0	2	2	2
ATPB_MOUSE	56 kDa	0	3	1	1
ATAD2_MOUSE	118 kDa	0	4	1	12
DHX9_MOUSE	149 kDa	20	29	28	27
BOREA_MOUSE	32 kDa	2	4	5	3
BRD2_MOUSE	88 kDa	1	4	5	6
CBX1_MOUSE	21 kDa	2	5	8	7
CBX3_MOUSE	21 kDa	7	9	10	8
CBX5_MOUSE	22 kDa	3	12	10	12
CHD1_MOUSE	196 kDa	13	26	17	26
H2AY_MOUSE	40 kDa	1	6	12	10
H2AW_MOUSE	40 kDa	0	1	1	6
CUL4B_MOUSE	?	1	4	2	3
QCR2_MOUSE	48 kDa	2	4	3	5
CY1_MOUSE	35 kDa	1	4	3	4
DNJA2_MOUSE	46 kDa	0	3	4	1
RPN2_MOUSE	69 kDa	0	2	1	2
RING2_MOUSE	38 kDa	0	2	1	2
EF1A1_MOUSE	50 kDa	6	7	7	8
EF1G_MOUSE	50 kDa	1	7	7	9
SP16H_MOUSE	120 kDa	2	6	13	3
FXR1_MOUSE	76 kDa	0	1	1	9
GNAI2_MOUSE	40 kDa	8	10	13	15
GNAI3_MOUSE	41 kDa	1	6	3	4
GNAS1_MOUSE (+1)	122 kDa	0	2	1	8
GBLP_MOUSE	35 kDa	0	2	3	10
GNL3_MOUSE	61 kDa	0	3	7	1
HDGR2_MOUSE	74 kDa	3	4	10	5
SAP18_MOUSE	18 kDa	10	12	17	11
H2A1F_MOUSE	14 kDa	3	4	5	6
H2B1C_MOUSE (+1)	14 kDa	10	11	12	11
H33_MOUSE	15 kDa	8	11	10	9
H4_MOUSE	11 kDa	12	13	13	13
EHMT2_MOUSE	138 kDa	1	3	3	2

Accession Number	Molecular Weight	Peptides in Mock	Peptides (Trim28-RBCC)	Peptides Trim28 FL mESC	Peptides FL (MEF cells)
M2OM_MOUSE	34 kDa	0	2	2	2
NAT10_MOUSE	115 kDa	15	22	21	17
NDUA9_MOUSE	43 kDa	0	2	2	3
NOC4L_MOUSE	59 kDa	3	7	8	4
NOL12_MOUSE	25 kDa	0	1	3	1
DDX21_MOUSE	94 kDa	29	36	41	36
UBF1_MOUSE	90 kDa	8	12	22	30
PPHLN_MOUSE	44 kDa	0	2	1	1
DDX49_MOUSE	54 kDa	0	1	4	1
PML_MOUSE	98 kDa	10	16	15	13
YES_MOUSE	61 kDa	0	1	2	1
K0020_MOUSE	73 kDa	10	13	19	13
CTR9_MOUSE	133 kDa	0	1	1	14
RRP12_MOUSE	143 kDa	3	15	17	7
RUVB2_MOUSE	51 kDa	2	9	13	3
S30BP_MOUSE	34 kDa	3	8	9	4
SMC1A_MOUSE	143 kDa	1	7	6	4
SSF1_MOUSE	53 kDa	0	1	7	1
SMCA5_MOUSE	122 kDa	37	40	43	41
THOC2_MOUSE	?	0	2	2	2
SPT6H_MOUSE	199 kDa	2	9	24	4
TIF1A_MOUSE	117 kDa	1	28	30	22
TIF1B_MOUSE	89 kDa	19	56	62	57
TBA1B_MOUSE	50 kDa	6	11	9	10
BAZ1B_MOUSE	171 kDa	17	26	34	49
WDR18_MOUSE	47 kDa	2	9	3	5
WDR61_MOUSE	34 kDa	0	1	4	10
ZCH18_MOUSE	106 kDa	6	10	19	11
ZFP1_MOUSE	47 kDa	0	5	2	1
ZNF12_MOUSE	?	0	3	2	3
ZN120_MOUSE	51 kDa	0	7	6	5
ZFP2_MOUSE	53 kDa	0	2	4	2
ZN235_MOUSE	73 kDa	1	4	3	3
ZFP57_MOUSE	48 kDa	0	23	19	11
1433T_MOUSE	28 kDa	0	2	1	0
ACL6A_MOUSE	47 kDa	2	6	4	1
AROS_MOUSE	16 kDa	1	2	2	1
DDX24_MOUSE	96 kDa	5	7	10	0
BAZ2A_MOUSE	210 kDa	2	7	11	0
CSK2B_MOUSE	25 kDa	0	3	2	0
CEBPZ_MOUSE	120 kDa	3	6	8	3
CENPV_MOUSE	28 kDa	3	6	6	2
CHD4_MOUSE	218 kDa	15	20	25	5

Accession Number	Molecular Weight	Peptides in Mock	Peptides (Trim28-RBCC)	Peptides Trim28 FL mESC	Peptides FL (MEF cells)
STAG1_MOUSE	145 kDa	0	5	2	0
COR2B_MOUSE	55 kDa	0	1	2	0
CTBP2_MOUSE	49 kDa	0	4	7	0
CDK11_MOUSE	?	14	15	31	13
DNMT1_MOUSE	183 kDa	3	7	11	3
MCM2_MOUSE	102 kDa	3	4	6	1
MCM4_MOUSE	97 kDa	3	5	12	3
TOP2A_MOUSE	173 kDa	20	22	30	14
RPB1_MOUSE	217 kDa	0	2	20	0
RPB2_MOUSE	134 kDa	0	1	7	0
RPAB3_MOUSE	17 kDa	0	1	2	0
RAD21_MOUSE	72 kDa	0	1	2	0
UHRF1_MOUSE	88 kDa	8	15	21	2
ELAV1_MOUSE	36 kDa	1	7	4	0
ENY2_MOUSE	12 kDa	2	3	4	1
ESF1_MOUSE	98 kDa	1	2	3	0
GRWD1_MOUSE	49 kDa	0	3	3	0
GRTP1_MOUSE	41 kDa	0	3	2	0
HSP7C_MOUSE	71 kDa	22	23	26	22
MYST4_MOUSE	209 kDa	0	2	3	0
RBBP4_MOUSE	48 kDa	3	5	6	2
EZH2_MOUSE	85 kDa	1	3	3	0
IMA2_MOUSE	58 kDa	5	6	7	2
IMDH2_MOUSE	56 kDa	0	6	3	0
ILF2_MOUSE	43 kDa	0	4	2	0
LRWD1_MOUSE	72 kDa	3	5	4	0
LITD1_MOUSE	88 kDa	22	37	39	0
TDH_MOUSE	41 kDa	0	4	1	0
KDM5B_MOUSE	176 kDa	6	16	11	0
MATR3_MOUSE	95 kDa	19	24	29	16
MTCH2_MOUSE	33 kDa	5	7	7	0
TOM40_MOUSE	38 kDa	0	5	3	0
MK67I_MOUSE	36 kDa	3	6	6	2
MBB1A_MOUSE	152 kDa	45	54	58	39
NDUAA_MOUSE	41 kDa	1	2	2	1
NDUA4_MOUSE	9 kDa	0	2	1	0
NU133_MOUSE	129 kDa	11	19	12	2
NU155_MOUSE	155 kDa	9	15	12	7
NU214_MOUSE	213 kDa	3	7	4	0
NR0B1_MOUSE	53 kDa	0	4	3	0
NPA1P_MOUSE	255 kDa	6	19	30	5
NOL8_MOUSE	129 kDa	5	8	9	2
NUCL_MOUSE	77 kDa	11	15	14	1

Accession Number	Molecular Weight	Peptides in Mock	Peptides (Trim28-RBCC)	Peptides Trim28 FL mESC	Peptides FL (MEF cells)
NU188_MOUSE	197 kDa	3	6	4	1
ORC1_MOUSE	95 kDa	1	6	9	0
ORC2_MOUSE	66 kDa	2	5	4	0
ORC4_MOUSE	50 kDa	2	5	5	1
ORC5_MOUSE	50 kDa	1	5	3	0
PAPD5_MOUSE	70 kDa	0	1	3	0
PCID2_MOUSE	46 kDa	4	10	5	4
PHF11_MOUSE	?	2	22	16	0
MPCP_MOUSE	40 kDa	6	9	10	1
PINX1_MOUSE	37 kDa	0	1	2	0
EED_MOUSE	50 kDa	0	4	7	0
DDX47_MOUSE	51 kDa	0	1	2	0
RBM19_MOUSE	106 kDa	2	4	5	2
ELYS_MOUSE	?	8	16	16	4
FA60A_MOUSE	25 kDa	1	3	2	0
JARD2_MOUSE	137 kDa	1	8	6	0
MAK16_MOUSE	35 kDa	0	3	3	0
REST_MOUSE	118 kDa	0	3	1	0
RFC2_MOUSE	39 kDa	1	3	3	1
RRP7A_MOUSE	32 kDa	7	12	12	7
RBM7_MOUSE	30 kDa	0	2	5	0
PNO1_MOUSE	27 kDa	2	7	7	2
RRP15_MOUSE	31 kDa	0	2	2	0
SAFB2_MOUSE	112 kDa	1	6	2	0
AURKB_MOUSE	39 kDa	5	6	7	1
PP1G_MOUSE	37 kDa	5	7	6	2
SFXN1_MOUSE	36 kDa	0	2	2	0
PDS5B_MOUSE	164 kDa	2	6	7	1
UTP20_MOUSE	318 kDa	20	34	41	8
SON_MOUSE	261 kDa	0	1	2	0
ERR2_MOUSE	48 kDa	0	2	1	0
SMC3_MOUSE	142 kDa	0	2	2	0
SMRC1_MOUSE	123 kDa	3	7	6	1
RIF1_MOUSE	266 kDa	5	11	11	2
TERF1_MOUSE	48 kDa	0	2	1	0
ZO2_MOUSE	131 kDa	0	2	1	0
SMCA4_MOUSE	?	10	12	13	10
TRI39_MOUSE	56 kDa	0	7	2	0
TBB5_MOUSE	50 kDa	6	17	17	6
CC063_MOUSE	181 kDa	0	4	5	0
UTF1_MOUSE	36 kDa	0	4	3	0
VDAC1_MOUSE	32 kDa	0	1	2	0
VDAC2_MOUSE	32 kDa	4	9	9	2

Accession Number	Molecular Weight	Peptides in Mock	Peptides (Trim28-RBCC)	Peptides Trim28 FL mESC	Peptides FL (MEF cells)
VDAC3_MOUSE	31 kDa	4	8	5	4
WDR83_MOUSE	?	0	1	5	0
WDR82_MOUSE	35 kDa	0	3	1	0
ZMYM4_MOUSE	172 kDa	1	3	3	0
ARPC2_MOUSE	34 kDa	2	3	2	10
ADNP_MOUSE	92 kDa	3	4	1	6
CLH_MOUSE	192 kDa	7	12	6	24
OST48_MOUSE	49 kDa	0	2	0	1
DPYD_MOUSE-R	?	0	1	0	2
NEDD4_MOUSE	103 kDa	2	3	1	6
EIF3F_MOUSE	38 kDa	0	1	0	2
FINC_MOUSE	272 kDa	1	3	0	19
GPC4_MOUSE	63 kDa	0	1	0	4
AT1A1_MOUSE	113 kDa	0	2	0	1
TCPZ_MOUSE	58 kDa	0	3	0	1
ATRX_MOUSE	279 kDa	13	20	11	35
TMOD3_MOUSE	40 kDa	23	24	18	24
CX056_MOUSE	26 kDa	0	1	0	2
USP9X_MOUSE-R	?	0	1	0	2
VIME_MOUSE	54 kDa	2	5	0	23
ZFP60_MOUSE	82 kDa	0	4	0	1
ZFP90_MOUSE	72 kDa	0	1	0	7
TRI27_MOUSE	59 kDa	0	4	0	1
MCM3A_MOUSE	217 kDa	2	3	2	1
ASSY_MOUSE	47 kDa	0	3	0	0
ATPA_MOUSE	60 kDa	5	6	4	0
ATPG_MOUSE	33 kDa	1	2	1	1
CSK21_MOUSE	45 kDa	11	14	10	7
CENPQ_MOUSE	31 kDa	0	2	0	0
C1QBP_MOUSE	31 kDa	0	2	0	0
CDK1_MOUSE	?	3	5	3	0
CYTSA_MOUSE	124 kDa	0	15	0	0
ZDBF2_MOUSE	274 kDa	0	2	0	0
MSH6_MOUSE	151 kDa	1	2	1	0
MCM6_MOUSE	93 kDa	10	13	10	5
TRI33_MOUSE	124 kDa	0	2	0	0
EFTU_MOUSE	50 kDa	0	4	0	0
EDC4_MOUSE	152 kDa	1	2	1	0
FRIL1_MOUSE	21 kDa	1	2	1	0
G3P_MOUSE	36 kDa	7	8	6	4
HXK2_MOUSE	103 kDa	1	4	1	1
H2A2A_MOUSE	14 kDa	9	11	9	8
ING4_MOUSE	29 kDa	1	2	1	0

Accession Number	Molecular Weight	Peptides in Mock	Peptides (Trim28-RBCC)	Peptides Trim28 FL mESC	Peptides FL (MEF cells)
ING5_MOUSE	28 kDa	5	6	4	2
ODB2_MOUSE	53 kDa	0	10	0	0
MBD3_MOUSE	32 kDa	3	5	3	1
MK01_MOUSE	41 kDa	0	2	0	0
NID2_MOUSE	154 kDa	0	2	0	0
NU107_MOUSE	107 kDa	10	12	10	4
NU160_MOUSE	158 kDa	27	30	25	23
NVL_MOUSE	94 kDa	3	5	3	0
NOC3L_MOUSE	93 kDa	1	2	0	0
NUP37_MOUSE	37 kDa	5	7	5	5
NUP43_MOUSE	42 kDa	4	6	2	2
NUP53_MOUSE	35 kDa	1	2	1	0
PCNA_MOUSE	29 kDa	3	4	3	1
ANM1_MOUSE	42 kDa	7	12	3	1
P5CR2_MOUSE	34 kDa	1	2	1	0
GDIB_MOUSE	51 kDa	0	2	0	0
RFC3_MOUSE	41 kDa	2	3	2	1
RFC4_MOUSE	40 kDa	0	3	0	0
RBM14_MOUSE	69 kDa	1	2	1	0
STK38_MOUSE	54 kDa	4	7	2	1
SPTA2_MOUSE	285 kDa	14	19	4	2
SPIN1_MOUSE	30 kDa	2	3	2	0
SUN1_MOUSE	?	1	2	0	0
SMRCD_MOUSE	116 kDa	0	2	0	0
TYSY_MOUSE	35 kDa	5	6	5	1
UBP7_MOUSE	128 kDa	0	2	0	0
CI114_MOUSE	43 kDa	11	15	10	2
K1797_MOUSE	199 kDa	1	2	0	1
WDR5_MOUSE	37 kDa	1	3	1	1
ZF106_MOUSE	209 kDa	0	3	0	0
ZN250_MOUSE	60 kDa	0	4	0	0
ZFP59_MOUSE	73 kDa	0	4	0	0
ZFP62_MOUSE	105 kDa	1	3	1	1
ZN667_MOUSE	70 kDa	0	2	0	0
PRS4_MOUSE	49 kDa	0	0	1	4
PRS7_MOUSE	49 kDa	1	1	2	4
RT16_MOUSE	15 kDa	0	0	1	3
RT31_MOUSE	44 kDa	1	1	3	4
RM23_MOUSE	17 kDa	0	0	3	1
RM03_MOUSE	39 kDa	0	0	3	1
RS15_MOUSE	17 kDa	3	2	4	8
RS17_MOUSE	16 kDa	3	3	4	7
RS25_MOUSE	14 kDa	3	3	4	6

Accession Number	Molecular Weight	Peptides in Mock	Peptides (Trim28-RBCC)	Peptides Trim28 FL mESC	Peptides FL (MEF cells)
RS8_MOUSE	24 kDa	5	5	9	11
RL13_MOUSE	24 kDa	6	5	8	11
RL18A_MOUSE	21 kDa	3	0	4	7
RL36A_MOUSE	12 kDa	7	6	8	8
IF4A3_MOUSE	47 kDa	23	21	37	26
IGH1M_MOUSE (+1)	43 kDa	19	17	20	20
GCAA_MOUSE (+1)	36 kDa	0	0	1	4
INT1_MOUSE	245 kDa	0	0	1	2
KRR1_MOUSE	44 kDa	2	2	3	3
NCBP1_MOUSE	92 kDa	0	0	3	1
PININ_MOUSE	82 kDa	20	9	38	27
DDX41_MOUSE	70 kDa	4	0	9	11
RNPS1_MOUSE	34 kDa	2	2	7	6
SRSF9_MOUSE	?	1	1	7	3
THOC1_MOUSE	75 kDa	0	0	1	2
TRA2B_MOUSE	34 kDa	11	10	19	12
UTP6_MOUSE	70 kDa	8	8	19	10
RM37_MOUSE	48 kDa	2	2	3	2
RM44_MOUSE	38 kDa	0	0	3	0
RL14_MOUSE	24 kDa	2	1	3	1
RL19_MOUSE	23 kDa	2	2	4	2
RL32_MOUSE	16 kDa	2	2	3	2
RL38_MOUSE	8 kDa	2	2	3	2
DDX51_MOUSE	70 kDa	4	2	8	0
DDX54_MOUSE	98 kDa	1	1	2	0
CIR1_MOUSE	?	0	0	2	0
CRNL1_MOUSE	83 kDa	8	2	17	3
NHP2_MOUSE	17 kDa	1	1	2	1
HNRPL_MOUSE	64 kDa	5	5	8	5
HNRPM_MOUSE	78 kDa	14	4	18	9
LV1A_MOUSE (+1)	12 kDa	1	1	3	1
K1C14_MOUSE	53 kDa	3	1	5	3
K22O_MOUSE	63 kDa	1	1	2	1
NOP14_MOUSE	99 kDa	6	6	8	3
NOP56_MOUSE	64 kDa	27	20	28	18
NOP58_MOUSE	60 kDa	18	10	20	10
NO66_MOUSE	68 kDa	2	1	3	0
PESC_MOUSE	68 kDa	5	3	7	4
PR40A_MOUSE	108 kDa	4	4	5	4
PRP6_MOUSE	107 kDa	10	10	12	10
ISY1_MOUSE	33 kDa	4	1	8	2
SLU7_MOUSE	68 kDa	0	0	2	0
SPF27_MOUSE	26 kDa	0	0	2	0
DDX5_MOUSE	69 kDa	7	6	8	2

Accession Number	Molecular Weight	Peptides in Mock	Peptides (Trim28-RBCC)	Peptides Trim28 FL mESC	Peptides FL (MEF cells)
DDX56_MOUSE	61 kDa	4	2	6	3
CWC15_MOUSE	27 kDa	2	2	8	1
SDA1_MOUSE	80 kDa	1	0	2	0
VIR_MOUSE	201 kDa	5	4	6	3
RRP1B_MOUSE	81 kDa	7	5	13	2
RBM39_MOUSE	59 kDa	4	4	5	0
RBM8A_MOUSE	20 kDa	3	1	6	3
SRRM1_MOUSE	107 kDa	0	0	2	0
SRRM2_MOUSE	295 kDa	27	16	40	11
SRSF1_MOUSE	?	10	9	18	9
SRSF2_MOUSE	?	3	2	5	3
SRSF4_MOUSE	?	2	2	4	2
SRSF7_MOUSE	?	8	7	12	6
SMD2_MOUSE	14 kDa	1	1	2	0
SNW1_MOUSE	61 kDa	5	1	11	0
SK2L2_MOUSE	118 kDa	1	1	6	0
RU17_MOUSE	52 kDa	6	4	12	5
PSMD2_MOUSE	100 kDa	1	1	1	9
RS12_MOUSE	15 kDa	2	2	2	4
RS24_MOUSE	15 kDa	1	1	1	2
RS29_MOUSE	7 kDa	2	1	2	4
RLA2_MOUSE	12 kDa	0	0	0	2
RL11_MOUSE	20 kDa	5	5	5	9
RL13A_MOUSE	23 kDa	0	0	0	2
RL18_MOUSE	22 kDa	1	1	1	2
RL21_MOUSE	19 kDa	3	3	3	5
RL23_MOUSE	15 kDa	3	3	3	4
RL27A_MOUSE	17 kDa	3	3	2	5
RL34_MOUSE	13 kDa	0	0	0	2
RL37A_MOUSE	10 kDa	4	3	4	7
RL8_MOUSE	28 kDa	6	6	6	11
IGKC_MOUSE	12 kDa	17	17	17	18
KV2A7_MOUSE	12 kDa	8	8	8	9
KV3AA_MOUSE (+1)	12 kDa	1	1	1	2
KV3AJ_MOUSE	12 kDa	3	3	3	4
KV5AB_MOUSE (+2)	12 kDa	2	1	1	3
K1C17_MOUSE	48 kDa	10	6	6	12
K2C6A_MOUSE	59 kDa	10	7	10	13
SRSF5_MOUSE	?	2	2	2	3
GRP78_MOUSE	72 kDa	3	1	4	12
ACINU_MOUSE	151 kDa	23	13	38	24
DDX50_MOUSE	82 kDa	0	0	1	6
MCM7_MOUSE	81 kDa	6	2	15	9
HNRPU_MOUSE	?	24	24	33	26

Accession Number	Molecular Weight	Peptides in Mock	Peptides (Trim28-RBCC)	Peptides Trim28 FL mESC	Peptides FL (MEF cells)
IMA4_MOUSE	58 kDa	0	0	1	2
LMNA_MOUSE	74 kDa	5	1	6	48
NIPBL_MOUSE	315 kDa	1	0	3	3
NOL9_MOUSE	81 kDa	3	3	6	9
CDC73_MOUSE	61 kDa	0	0	1	10
PTCD3_MOUSE	78 kDa	1	0	6	5
PELP1_MOUSE	118 kDa	7	6	10	8
LAS1L_MOUSE	?	7	7	11	13
SENP3_MOUSE	64 kDa	4	3	7	5
SET1B_MOUSE-R	?	0	0	2	1
WDR65_MOUSE	?	0	0	2	1
DDX18_MOUSE	74 kDa	4	2	6	4
CSK22_MOUSE	41 kDa	5	5	6	2
CIR1A_MOUSE	77 kDa	4	3	6	0
CCDC9_MOUSE	61 kDa	0	0	6	0
CCNL1_MOUSE	60 kDa	2	2	10	2
CCNL2_MOUSE	58 kDa	1	1	4	0
DOCK6_MOUSE	233 kDa	1	1	3	0
TDIF2_MOUSE	84 kDa	14	12	15	10
DPPA2_MOUSE	34 kDa	3	3	4	0
DPPA4_MOUSE	33 kDa	3	2	4	0
RPB3_MOUSE	31 kDa	0	0	3	0
RPB9_MOUSE	15 kDa	0	0	2	0
DMD_MOUSE	426 kDa	1	0	2	1
RBP2_MOUSE	341 kDa	19	18	22	7
FANCI_MOUSE	149 kDa	0	0	2	0
GPTC8_MOUSE	165 kDa	0	0	2	0
H13_MOUSE	22 kDa	3	2	6	3
H15_MOUSE	23 kDa	0	0	2	0
H2AX_MOUSE	15 kDa	3	2	4	3
IP3KA_MOUSE	51 kDa	0	0	2	0
LMNB1_MOUSE	67 kDa	23	23	30	11
HELLS_MOUSE	95 kDa	11	6	15	0
MEN1_MOUSE	67 kDa	1	1	2	0
MFAP1_MOUSE	52 kDa	2	1	3	2
MINT_MOUSE	399 kDa	0	0	2	0
NAA40_MOUSE	?	2	2	3	0
NGDN_MOUSE	36 kDa	3	2	5	3
NUP93_MOUSE	93 kDa	12	3	17	12
NXF1_MOUSE	70 kDa	1	1	4	1
NOL6_MOUSE	129 kDa	23	23	25	18
NUP85_MOUSE	75 kDa	14	6	19	13
ORC3_MOUSE	82 kDa	3	0	8	0
PPIG_MOUSE	88 kDa	7	5	16	7

Accession Number	Molecular Weight	Peptides in Mock	Peptides (Trim28-RBCC)	Peptides Trim28 FL mESC	Peptides FL (MEF cells)
PWP1_MOUSE	56 kDa	7	2	8	1
DDX10_MOUSE	101 kDa	18	16	23	8
DDX27_MOUSE	86 kDa	0	0	2	0
DDX52_MOUSE	67 kDa	4	1	5	0
IWS1_MOUSE	85 kDa	4	2	7	2
MGNR_MOUSE (+1)	17 kDa	8	6	14	7
CN021_MOUSE	70 kDa	5	2	6	2
RBM34_MOUSE	41 kDa	0	0	2	0
RUVB1_MOUSE	50 kDa	6	5	7	6
SAS10_MOUSE	53 kDa	2	1	4	1
SPT5H_MOUSE	121 kDa	0	0	2	0
TCOF_MOUSE	135 kDa	4	3	5	0
WDR43_MOUSE	75 kDa	9	6	13	5
WDR76_MOUSE	69 kDa	0	0	2	0
ZCH10_MOUSE	19 kDa	1	1	3	1
ACTA_MOUSE	42 kDa	17	17	16	31
ARP2_MOUSE	45 kDa	4	4	0	9
ARC1B_MOUSE	41 kDa	2	0	0	3
ARPC4_MOUSE	20 kDa	3	2	3	4
ARP3_MOUSE	47 kDa	3	2	1	13
AKAP9_MOUSE-R	?	1	0	1	2
ANXA1_MOUSE	39 kDa	0	0	0	4
ANXA2_MOUSE	39 kDa	0	0	0	9
ATAD3_MOUSE	67 kDa	1	0	0	4
PGBM_MOUSE	398 kDa	0	0	0	4
BCLF1_MOUSE	106 kDa	26	10	19	46
ACTBL_MOUSE	42 kDa	1	0	1	2
PGFRB_MOUSE	123 kDa	0	0	0	2
PGS1_MOUSE	42 kDa	0	0	0	5
BAZ1A_MOUSE	178 kDa	1	1	1	5
CRTAP_MOUSE	46 kDa	0	0	0	6
CAV1_MOUSE	21 kDa	0	0	0	4
CHM4B_MOUSE	25 kDa	0	0	0	2
CHCH3_MOUSE	26 kDa	2	1	2	3
F120A_MOUSE	122 kDa	0	0	0	4
CUL4A_MOUSE	88 kDa	0	0	0	2
DDB1_MOUSE	127 kDa	2	0	0	9
MCM5_MOUSE	82 kDa	11	2	10	16
TOP2B_MOUSE	182 kDa	1	0	1	5
RPN1_MOUSE	69 kDa	1	0	0	3
ERH_MOUSE	12 kDa	7	7	6	8
ERLN1_MOUSE (+1)	39 kDa	1	1	1	2
EIF3C_MOUSE	106 kDa	0	0	0	5
EIF3I_MOUSE	36 kDa	0	0	0	2

Accession Number	Molecular Weight	Peptides in Mock	Peptides (Trim28-RBCC)	Peptides Trim28 FL mESC	Peptides FL (MEF cells)
EIF3L_MOUSE	67 kDa	0	0	0	2
SSRP1_MOUSE	81 kDa	3	0	3	7
FBLN2_MOUSE	132 kDa	0	0	0	3
FLNA_MOUSE	281 kDa	1	0	0	7
FMR1_MOUSE	69 kDa	0	0	0	26
FXR2_MOUSE	74 kDa	0	0	0	7
LEG1_MOUSE	15 kDa	0	0	0	2
GELS_MOUSE	86 kDa	5	0	0	22
GNA1_MOUSE	21 kDa	0	0	0	2
GUF1_MOUSE	72 kDa	1	1	1	2
GNAO_MOUSE	40 kDa	0	0	0	2
HP1B3_MOUSE	61 kDa	0	0	0	5
HDAC2_MOUSE	55 kDa	0	0	0	2
SUV91_MOUSE	48 kDa	0	0	0	3
IF2B2_MOUSE	66 kDa	1	1	0	7
ITIH3_MOUSE	99 kDa	0	0	0	2
IFM3_MOUSE	?	0	0	0	2
K1211_MOUSE-R	?	1	1	1	2
LARP4_MOUSE	80 kDa	0	0	0	2
NFIP1_MOUSE	25 kDa	0	0	0	3
NCOA5_MOUSE	65 kDa	0	0	0	2
PALLD_MOUSE-R	?	1	1	1	2
PHIP_MOUSE	207 kDa	0	0	0	15
PGAM5_MOUSE	32 kDa	9	6	8	10
PLEC_MOUSE	?	2	1	0	66
PTRF_MOUSE	44 kDa	0	0	0	11
PLOD3_MOUSE	85 kDa	1	0	1	20
LRP1_MOUSE	505 kDa	0	0	0	2
P4HA1_MOUSE	61 kDa	0	0	0	2
KPCA_MOUSE	77 kDa	0	0	0	3
S61A1_MOUSE	52 kDa	0	0	0	5
ODPA_MOUSE	43 kDa	0	0	0	3
ODPB_MOUSE	39 kDa	0	0	0	2
RP9_MOUSE	25 kDa	0	0	0	4
WDR12_MOUSE	47 kDa	0	0	0	2
PAF1_MOUSE	61 kDa	2	0	2	10
LEO1_MOUSE	76 kDa	0	0	0	6
SQSTM_MOUSE	48 kDa	0	0	0	2
SDPR_MOUSE	47 kDa	0	0	0	7
GTR1_MOUSE	54 kDa	1	0	0	3
SMHD1_MOUSE	226 kDa	0	0	0	3
SUN2_MOUSE	?	1	0	1	21
TEX10_MOUSE	?	6	2	5	7

Accession Number	Molecular Weight	Peptides in Mock	Peptides (Trim28-RBCC)	Peptides Trim28 FL mESC	Peptides FL (MEF cells)
TSP1_MOUSE	130 kDa	0	0	0	3
TR150_MOUSE	108 kDa	37	19	29	47
TTHY_MOUSE	16 kDa	3	1	2	4
JAK1_MOUSE	133 kDa	0	0	0	5
ZN326_MOUSE	65 kDa	0	0	0	10
ZRAB2_MOUSE	37 kDa	0	0	0	2

Used IAP-GAG sequences that are not listed in the methods section (in 5' to 3' direction)

gag (used and cloned in antisense)

tgaatttcccttgctgtccttcagaggattcccacgtcaaggcaacggagcttatagtgggacatgattgatatcctaggccctgtaatgaatgagatgactctgtggtgggccatgctttg
ggccaccaatgtgtagagattatacttttatctgctccggtatcaaggatgccttcaaactcttttccattaatcttaaggcggagcttaggtctatcatttaaagatacaaccaaataggca
gaatcatttcctgaggagcccatttttctttatctcaggtcctgcagatttctccctggtattatcagggaggagcagcagctgagctatcctatctcctttactaatagaaaaaacgcccttag
ggcttgagcacaggacctgtatttcaggggaatgttgacaatccataactccagggtggactactaagccctgcaaggtgagtgaaccccggccgagaataaggcccatggttccc
aggggcaaggatggtataggctccactggcaccggctgaatactcatttgaggcattaataggaagtcggaggcggcacgcaggtccacccttgtgggtcttcctgggtcgcctctct
gactgcttcttgggtcctgacaaaccggttcccatatctttgagggccctgggaccgagggcccgatgacccgtttttggcacataaagctgattgactatcaggtgggggaaggact
ctgccctttatatccctcacagagcgacactggtcagctctatgataacccttgccacacttagagcaaagagtgagagtccctccctgtttatctggagctctgcaatctttcttaaaatgc
ccaggctttccgcaattaaaacatgtcctctgatcatttctgctcatggagcggttctgagattggaggatggcggccgctaagcctgcattggtgagaggtcccccaagctctcgacag
accctgagccagtcttgtaagcctttgttctttcttggagctatggccgctcggcactcctttgtggcttgctcatagatgagctgttctatcagaggcgcagcttgctctgactctccaaaaat
acgctctgctgcctctgtcattctggccacaaaatctgagaaggattcctgaggtccctggattatctttgttaactgcccagtggtttcacctgctcgggagagcgccttccaggccctaat
agccgtggaagaaatctgggcataagctccccaatggtagtttgtctgatcagcagaataagctccctgacccgttaacaagtcaaaagtccaatctctctgctctggagtcaaagca
gctgcgtttgctcgggcctgcgcttgtgcagcttcatgccaaagcgctctccattccatatatttgcccatactagggagagcggcttttacaatcgtttgccagtcagcaggagttagtgc
catgccggcgagcctgtctaactgcaccaaggtaaaattagcattggttccgtatttacggaccgactcggcaatttctttaatctgtacgtattctaccggagcgtggacacgcccacc
ctcggctccttcaaagactggaaatgcctgttgtattttccttgttcctctctgggaatgaatgagtctgcacattgcctctctgcgcattgtctctctgcgcagggctgaccactacgcaggg
cggggactccgcatagggcggaccttgaagccgactgccctgaggccaatcagcaaactggccttcgccagccgcttttggcttccttaactgattagctagcactttacctggctggt
acccttttttctcataatgagctgcttcttcctcccagtctgtttcttcagaggagaattcatcatctgattcagagctaccaagagctggctccccgagcccatctagcgatgagcacaggc
tccttttcctagagacctccgctaattgatctttcttcttttccctttttcctctttt

gag2

aaggaaaatacaacaggcatttccagtctttgaaggagccgagggtgggcgtgtccacgctccggtagaatacgtacagattaaagaaattgccgagtcggtccgtaaatacgga
accaatgctaattttaccttggtgcagttagacaggctcgccggcatggcactaactcctgctgactggcaaacgattgtaaaagccgctctccctagtatgggcaaatatatggaatg
gagagcgctttggcatgaagctgcacaagcgcaggcccgagcaaacgcagctgctttgactccagagcagagagattggacttttgacttgttaacgggtcagggagcttattctgct
gatcagacaaactaccattggggagcttatgcccagatttcttccacggctattagggcct

9. Citations

Adam, S., Polo, S.E., and Almouzni, G. (2013). Transcription recovery after DNA damage requires chromatin priming by the H3.3 histone chaperone HIRA. Cell *155*, 94-106.

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

Akhtar, W., de Jong, J., Pindyurin, A.V., Pagie, L., Meuleman, W., de Ridder, J., Berns, A., Wessels, L.F., van Lohuizen, M., and van Steensel, B. (2013). Chromatin position effects assayed by thousands of reporters integrated in parallel. Cell *154*, 914-927.

Albritton, L.M., Tseng, L., Scadden, D., and Cunningham, J.M. (1989). A putative murine ecotropic retrovirus receptor gene encodes a multiple membrane-spanning protein and confers susceptibility to virus infection. Cell *57*, 659-666.

Allis, C.D., Jenuwein, T., and Reinberg, D. (2007). Epigenetics (Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press).

Arand, J., Spieler, D., Karius, T., Branco, M.R., Meilinger, D., Meissner, A., Jenuwein, T., Xu, G., Leonhardt, H., Wolf, V.*, et al.* (2012). In vivo control of CpG and non-CpG DNA methylation by DNA methyltransferases. PLoS genetics *8*, e1002750.

Arnold, C.D., Gerlach, D., Stelzer, C., Boryn, L.M., Rath, M., and Stark, A. (2013). Genome-wide quantitative enhancer activity maps identified by STARR-seq. Science *339*, 1074-1077.

Ashe, A., Morgan, D.K., Whitelaw, N.C., Bruxner, T.J., Vickaryous, N.K., Cox, L.L., Butterfield, N.C., Wicking, C., Blewitt, M.E., Wilkins, S.J.*, et al.* (2008). A genome-wide screen for modifiers of transgene variegation identifies genes with critical roles in development. Genome biology *9*, R182.

Ayyanathan, K., Lechner, M.S., Bell, P., Maul, G.G., Schultz, D.C., Yamada, Y., Tanaka, K., Torigoe, K., and Rauscher, F.J., 3rd (2003). Regulated recruitment of HP1 to a euchromatic gene induces mitotically heritable, epigenetic gene silencing: a mammalian cell culture model of gene variegation. Genes & development *17*, 1855-1869.

Banaszynski, L.A., Wen, D., Dewell, S., Whitcomb, S.J., Lin, M., Diaz, N., Elsasser, S.J., Chapgier, A., Goldberg, A.D., Canaani, E.*, et al.* (2013). Hira-dependent histone H3.3 deposition facilitates PRC2 recruitment at developmental loci in ES cells. Cell *155*, 107-120.

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

Baranello, L., Levens, D., Gupta, A., and Kouzine, F. (2012). The importance of being supercoiled: how DNA mechanics regulate dynamic processes. Biochimica et biophysica acta *1819*, 632-638.

Barde, I., Rauwel, B., Marin-Florez, R.M., Corsinotti, A., Laurenti, E., Verp, S., Offner, S., Marquis, J., Kapopoulou, A., Vanicek, J.*, et al.* (2013). A KRAB/KAP1-miRNA cascade regulates erythropoiesis through stage-specific control of mitophagy. Science *340*, 350-353.

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.

Bartke, T., Vermeulen, M., Xhemalce, B., Robson, S.C., Mann, M., and Kouzarides, T. (2010). Nucleosome-interacting proteins regulated by DNA and histone methylation. Cell *143*, 470-484.

Baubec, T., Ivanek, R., Lienert, F., and Schubeler, D. (2013). Methylation-dependent and -independent genomic targeting principles of the MBD protein family. Cell *153*, 480-492.

Baudat, F., Buard, J., Grey, C., Fledel-Alon, A., Ober, C., Przeworski, M., Coop, G., and de Massy, B. (2010). PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. Science *327*, 836-840.

Beadle, G.W. (1932). A Possible Influence of the Spindle Fibre on Crossing-Over in Drosophila. Proceedings of the National Academy of Sciences of the United States of America *18*, 160-165.

Beck, C.R., Collier, P., Macfarlane, C., Malig, M., Kidd, J.M., Eichler, E.E., Badge, R.M., and Moran, J.V. (2010). LINE-1 retrotransposition activity in human genomes. Cell *141*, 1159-1170.

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

Bellefroid, E.J., Poncelet, D.A., Lecocq, P.J., Revelant, O., and Martial, J.A. (1991). The evolutionarily conserved Kruppel-associated box domain defines a subfamily of eukaryotic multifingered proteins. Proceedings of the National Academy of Sciences of the United States of America *88*, 3608-3612.

Belshaw, R., Dawson, A.L., Woolven-Allen, J., Redding, J., Burt, A., and Tristem, M. (2005). Genomewide screening reveals high levels of insertional polymorphism in the human endogenous retrovirus family HERV-K(HML2): implications for present-day activity. Journal of virology *79*, 12507-12514.

Benetti, R., Gonzalo, S., Jaco, I., Schotta, G., Klatt, P., Jenuwein, T., and Blasco, M.A. (2007). Suv4-20h deficiency results in telomere elongation and derepression of telomere recombination. The Journal of cell biology *178*, 925-936.

Berg, I.L., Neumann, R., Lam, K.W., Sarbajna, S., Odenthal-Hesse, L., May, C.A., and Jeffreys, A.J. (2010). PRDM9 variation strongly influences recombination hot-spot activity and meiotic instability in humans. Nature genetics *42*, 859-863.

Bernstein, E., and Allis, C.D. (2005). RNA meets chromatin. Genes & development *19*, 1635-1655.

Best, S., Le Tissier, P., Towers, G., and Stoye, J.P. (1996). Positional cloning of the mouse retrovirus restriction gene Fv1. Nature *382*, 826-829.

Bilodeau, S., Kagey, M.H., Frampton, G.M., Rahl, P.B., and Young, R.A. (2009). SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state. Genes & development *23*, 2484-2489.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes & development *16*, 6-21.

Blaise, S., de Parseval, N., Benit, L., and Heidmann, T. (2003). Genomewide screening for fusogenic human endogenous retrovirus envelopes identifies syncytin 2, a gene conserved on primate evolution. Proceedings of the National Academy of Sciences of the United States of America *100*, 13013-13018.

Blaise, S., de Parseval, N., and Heidmann, T. (2005). Functional characterization of two newly identified Human Endogenous Retrovirus coding envelope genes. Retrovirology *2*, 19.

Blewitt, M.E., Vickaryous, N.K., Paldi, A., Koseki, H., and Whitelaw, E. (2006). Dynamic reprogramming of DNA methylation at an epigenetically sensitive allele in mice. PLoS genetics *2*, e49.

Blond, J.L., Lavillette, D., Cheynet, V., Bouton, O., Oriol, G., Chapel-Fernandes, S., Mandrand, B., Mallet, F., and Cosset, F.L. (2000). An envelope glycoprotein of the human endogenous retrovirus HERV-W is expressed in the human placenta and fuses cells expressing the type D mammalian retrovirus receptor. Journal of virology *74*, 3321-3329.

Bock, C., Reither, S., Mikeska, T., Paulsen, M., Walter, J., and Lengauer, T. (2005). BiQ Analyzer: visualization and quality control for DNA methylation data from bisulfite sequencing. Bioinformatics *21*, 4067-4068.

Boeke, J., Regnard, C., Cai, W., Johansen, J., Johansen, K.M., Becker, P.B., and Imhof, A. (2010). Phosphorylation of SU(VAR)3-9 by the chromosomal kinase JIL-1. PloS one *5*, e10042.

Bourc'his, D., and Bestor, T.H. (2004). Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. Nature *431*, 96-99.

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

Brick, K., Smagulova, F., Khil, P., Camerini-Otero, R.D., and Petukhova, G.V. (2012). Genetic recombination is directed away from functional genomic elements in mice. Nature *485*, 642-645.

Brown, S.W. (1966). Heterochromatin. Science *151*, 417-425.

Bulut-Karslioglu, A., Perrera, V., Scaranaro, M., de la Rosa-Velazquez, I.A., van de Nobelen, S., Shukeir, N., Popow, J., Gerle, B., Opravil, S., Pagani, M.*, et al.* (2012). A transcription factor-based mechanism for mouse heterochromatin formation. Nature structural & molecular biology *19*, 1023-1030.

Burgess, R.J., and Zhang, Z. (2013). Histone chaperones in nucleosome assembly and human disease. Nature structural & molecular biology *20*, 14-22.

Butter, F., Kappei, D., Buchholz, F., Vermeulen, M., and Mann, M. (2010). A domesticated transposon mediates the effects of a single-nucleotide polymorphism responsible for enhanced muscle growth. EMBO reports *11*, 305-311.

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.

Canzio, D., Chang, E.Y., Shankar, S., Kuchenbecker, K.M., Simon, M.D., Madhani, H.D., Narlikar, G.J., and Al-Sady, B. (2011). Chromodomain-mediated oligomerization of HP1 suggests a nucleosome-bridging mechanism for heterochromatin assembly. Molecular cell *41*, 67-81.

Castel, S.E., and Martienssen, R.A. (2013). RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. Nature reviews Genetics *14*, 100-112.

Ceol, C.J., Houvras, Y., Jane-Valbuena, J., Bilodeau, S., Orlando, D.A., Battisti, V., Fritsch, L., Lin, W.M., Hollmann, T.J., Ferre, F.*, et al.* (2011). The histone methyltransferase SETDB1 is recurrently amplified in melanoma and accelerates its onset. Nature *471*, 513-517.

Chang, C.C., Lin, D.Y., Fang, H.I., Chen, R.H., and Shih, H.M. (2005). Daxx mediates the small ubiquitin-like modifier-dependent transcriptional repression of Smad4. The Journal of biological chemistry *280*, 10164-10173.

Chang, C.C., Naik, M.T., Huang, Y.S., Jeng, J.C., Liao, P.H., Kuo, H.Y., Ho, C.C., Hsieh, Y.L., Lin, C.H., Huang, N.J.*, et al.* (2011). Structural and functional roles of Daxx SIM phosphorylation in SUMO paralog-selective binding and apoptosis modulation. Molecular cell *42*, 62-74.

Cheung, N.K., Zhang, J., Lu, C., Parker, M., Bahrami, A., Tickoo, S.K., Heguy, A., Pappo, A.S., Federico, S., Dalton, J.*, et al.* (2012). Association of age at diagnosis and genetic mutations in patients with neuroblastoma. JAMA : the journal of the American Medical Association *307*, 1062-1071.

Cheutin, T., McNairn, A.J., Jenuwein, T., Gilbert, D.M., Singh, P.B., and Misteli, T. (2003). Maintenance of stable heterochromatin domains by dynamic HP1 binding. Science *299*, 721-725.

Cho, S., Park, J.S., and Kang, Y.K. (2011). Dual functions of histone-lysine N-methyltransferase Setdb1 protein at promyelocytic leukemia-nuclear body (PML-NB): maintaining PML-NB structure and regulating the expression of its associated genes. The Journal of biological chemistry *286*, 41115-41124.

Chong, S., Vickaryous, N., Ashe, A., Zamudio, N., Youngson, N., Hemley, S., Stopka, T., Skoultchi, A., Matthews, J., Scott, H.S.*, et al.* (2007). Modifiers of epigenetic reprogramming show paternal effects in the mouse. Nature genetics *39*, 614-622.

Christman, J.K. (2002). 5-Azacytidine and 5-aza-2'-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy. Oncogene *21*, 5483-5495.

Coffin, J.M., Hughes, S.H., and Varmus, H. (1997). Retroviruses (Plainview, N.Y.: Cold Spring Harbor Laboratory Press,), pp. xv, 843 p.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A.*, et al.* (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819-823.

Consortium, E.P. (2004). The ENCODE (ENCyclopedia Of DNA Elements) Project. Science *306*, 636-640.

Consortium, E.P., Bernstein, B.E., Birney, E., Dunham, I., Green, E.D., Gunter, C., and Snyder, M. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57-74.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science *322*, 1845-1848.

Corish, P., and Tyler-Smith, C. (1999). Attenuation of green fluorescent protein half-life in mammalian cells. Protein engineering *12*, 1035-1040.

Corsinotti, A., Kapopoulou, A., Gubelmann, C., Imbeault, M., Santoni de Sio, F.R., Rowe, H.M., Mouscaz, Y., Deplancke, B., and Trono, D. (2013). Global and stage specific patterns of Kruppel-associated-box zinc finger protein gene expression in murine early embryonic cells. PloS one *8*, e56721.

Coufal, N.G., Garcia-Perez, J.L., Peng, G.E., Yeo, G.W., Mu, Y., Lovci, M.T., Morell, M., O'Shea, K.S., Moran, J.V., and Gage, F.H. (2009). L1 retrotransposition in human neural progenitor cells. Nature *460*, 1127-1131.

Cowell, I.G., Papageorgiou, N., Padget, K., Watters, G.P., and Austin, C.A. (2011). Histone deacetylase inhibition redistributes topoisomerase IIbeta from heterochromatin to euchromatin. Nucleus *2*, 61-71.

Croxton, R., Puto, L.A., de Belle, I., Thomas, M., Torii, S., Hanaii, F., Cuddy, M., and Reed, J.C. (2006). Daxx represses expression of a subset of antiapoptotic genes regulated by nuclear factor-kappaB. Cancer research *66*, 9026-9035.

Dambacher, S. (2013). Regulation of centromeric and pericentric heterochromatin (Ludwig-Maximilians-Universität München).

Dambacher, S., Deng, W., Hahn, M., Sadic, D., Frohlich, J., Nuber, A., Hoischen, C., Diekmann, S., Leonhardt, H., and Schotta, G. (2012). CENP-C facilitates the recruitment of M18BP1 to centromeric chromatin. Nucleus *3*, 101-110.

Dambacher, S., Hahn, M., and Schotta, G. (2010). Epigenetic regulation of development by histone lysine methylation. Heredity *105*, 24-37.

Davis, C.A., Haberland, M., Arnold, M.A., Sutherland, L.B., McDonald, O.G., Richardson, J.A., Childs, G., Harris, S., Owens, G.K., and Olson, E.N. (2006). PRISM/PRDM6, a transcriptional repressor that promotes the proliferative gene program in smooth muscle cells. Molecular and cellular biology *26*, 2626-2636.

Daxinger, L., Harten, S.K., Oey, H., Epp, T., Isbel, L., Huang, E., Whitelaw, N., Apedaile, A., Sorolla, A., Yong, J.*, et al.* (2013). An ENU mutagenesis screen identifies novel and known genes involved in epigenetic processes in the mouse. Genome biology *14*, R96.

de Koning, A.P., Gu, W., Castoe, T.A., Batzer, M.A., and Pollock, D.D. (2011). Repetitive elements may comprise over two-thirds of the human genome. PLoS genetics *7*, e1002384.

De La Fuente, R., Viveiros, M.M., Wigglesworth, K., and Eppig, J.J. (2004). ATRX, a member of the SNF2 family of helicase/ATPases, is required for chromosome alignment and meiotic spindle organization in metaphase II stage mouse oocytes. Developmental biology *272*, 1-14.

Deng, Z., Wan, M., and Sui, G. (2007). PIASy-mediated sumoylation of Yin Yang 1 depends on their interaction but not the RING finger. Molecular and cellular biology *27*, 3780-3792.

Dennis, K., Fan, T., Geiman, T., Yan, Q., and Muegge, K. (2001). Lsh, a member of the SNF2 family, is required for genome-wide methylation. Genes & development *15*, 2940-2944.

Dewannieux, M., Dupressoir, A., Harper, F., Pierron, G., and Heidmann, T. (2004). Identification of autonomous IAP LTR retrotransposons mobile in mammalian cells. Nature genetics *36*, 534-539.

Dhayalan, A., Tamas, R., Bock, I., Tattermusch, A., Dimitrova, E., Kudithipudi, S., Ragozin, S., and Jeltsch, A. (2011). The ATRX-ADD domain binds to H3 tail peptides and reads the combined methylation state of K4 and K9. Human molecular genetics *20*, 2195-2203.

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

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.

Dong, K.B., Maksakova, I.A., Mohn, F., Leung, D., Appanah, R., Lee, S., Yang, H.W., Lam, L.L., Mager, D.L., Schubeler, D.*, et al.* (2008). DNA methylation in ES cells requires the lysine methyltransferase G9a but not its catalytic activity. The EMBO journal *27*, 2691-2701.

Drane, P., Ouararhni, K., Depaux, A., Shuaib, M., and Hamiche, A. (2010). The death-associated protein DAXX is a novel histone chaperone involved in the replication-independent deposition of H3.3. Genes & development *24*, 1253-1265.

Dupressoir, A., Vernochet, C., Bawa, O., Harper, F., Pierron, G., Opolon, P., and Heidmann, T. (2009). Syncytin-A knockout mice demonstrate the critical role in placentation of a fusogenic, endogenous retrovirus-derived, envelope gene. Proceedings of the National Academy of Sciences of the United States of America *106*, 12127-12132.

Elsasser, S.J., Allis, C.D., and Lewis, P.W. (2011). Cancer. New epigenetic drivers of cancers. Science *331*, 1145-1146.

Elsasser, S.J., Huang, H., Lewis, P.W., Chin, J.W., Allis, C.D., and Patel, D.J. (2012). DAXX envelops a histone H3.3-H4 dimer for H3.3-specific recognition. Nature *491*, 560-565.

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

Eustermann, S., Yang, J.C., Law, M.J., Amos, R., Chapman, L.M., Jelinska, C., Garrick, D., Clynes, D., Gibbons, R.J., Rhodes, D.*, et al.* (2011). Combinatorial readout of histone H3 modifications specifies localization of ATRX to heterochromatin. Nature structural & molecular biology *18*, 777-782.

Evsikov, A.V., de Vries, W.N., Peaston, A.E., Radford, E.E., Fancher, K.S., Chen, F.H., Blake, J.A., Bult, C.J., Latham, K.E., Solter, D.*, et al.* (2004). Systems biology of the 2-cell mouse embryo. Cytogenetic and genome research *105*, 240-250.

Falandry, C., Fourel, G., Galy, V., Ristriani, T., Horard, B., Bensimon, E., Salles, G., Gilson, E., and Magdinier, F. (2010). CLLD8/KMT1F is a lysine methyltransferase that is important for chromosome segregation. The Journal of biological chemistry *285*, 20234-20241.

Fazzio, T.G., Huff, J.T., and Panning, B. (2008). An RNAi screen of chromatin proteins identifies Tip60-p400 as a regulator of embryonic stem cell identity. Cell *134*, 162-174.

Feuer, G., Taketo, M., Hanecak, R.C., and Fan, H. (1989). Two blocks in Moloney murine leukemia virus expression in undifferentiated F9 embryonal carcinoma cells as determined by transient expression assays. Journal of virology *63*, 2317-2324.

Filion, G.J., van Bemmel, J.G., Braunschweig, U., Talhout, W., Kind, J., Ward, L.D., Brugman, W., de Castro, I.J., Kerkhoven, R.M., Bussemaker, H.J.*, et al.* (2010). Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. Cell *143*, 212-224.

Flemming, W. (1882). Zellsubstanz, Kern und Zelltheilung (F. C. W. Vogel, Leipzig).

Fodor, B.D., Shukeir, N., Reuter, G., and Jenuwein, T. (2010). Mammalian Su(var) genes in chromatin control. Annual review of cell and developmental biology *26*, 471-501.

Friedman, J.R., Fredericks, W.J., Jensen, D.E., Speicher, D.W., Huang, X.P., Neilson, E.G., and Rauscher, F.J., 3rd (1996). KAP-1, a novel corepressor for the highly conserved KRAB repression domain. Genes & development *10*, 2067-2078.

Fritsch, L., Robin, P., Mathieu, J.R., Souidi, M., Hinaux, H., Rougeulle, C., Harel-Bellan, A., Ameyar-Zazoua, M., and Ait-Si-Ali, S. (2010). A subset of the histone H3 lysine 9 methyltransferases Suv39h1, G9a, GLP, and SETDB1 participate in a multimeric complex. Molecular cell *37*, 46-56.

Gamell, C., Jan Paul, P., Haupt, Y., and Haupt, S. (2014). PML tumour suppression and beyond: Therapeutic implications. FEBS letters.

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

Garcia-Perez, J.L., Morell, M., Scheys, J.O., Kulpa, D.A., Morell, S., Carter, C.C., Hammer, G.D., Collins, K.L., O'Shea, K.S., Menendez, P.*, et al.* (2010). Epigenetic silencing of engineered L1 retrotransposition events in human embryonic carcinoma cells. Nature *466*, 769-773.

Garrick, D., Samara, V., McDowell, T.L., Smith, A.J., Dobbie, L., Higgs, D.R., and Gibbons, R.J. (2004). A conserved truncated isoform of the ATR-X syndrome protein lacking the SWI/SNF-homology domain. Gene *326*, 23-34.

Garrick, D., Sharpe, J.A., Arkell, R., Dobbie, L., Smith, A.J., Wood, W.G., Higgs, D.R., and Gibbons, R.J. (2006). Loss of Atrx affects trophoblast development and the pattern of X-inactivation in extraembryonic tissues. PLoS genetics *2*, e58.

Gaudet, F., Rideout, W.M., 3rd, Meissner, A., Dausman, J., Leonhardt, H., and Jaenisch, R. (2004). Dnmt1 expression in pre- and postimplantation embryogenesis and the maintenance of IAP silencing. Molecular and cellular biology *24*, 1640-1648.

Gerstein, M.B., Kundaje, A., Hariharan, M., Landt, S.G., Yan, K.K., Cheng, C., Mu, X.J., Khurana, E., Rozowsky, J., Alexander, R.*, et al.* (2012). Architecture of the human regulatory network derived from ENCODE data. Nature *489*, 91-100.

Gewies, A., Castineiras-Vilarino, M., Ferch, U., Jahrling, N., Heinrich, K., Hoeckendorf, U., Przemeck, G.K., Munding, M., Gross, O., Schroeder, T.*, et al.* (2013). Prdm6 is essential for cardiovascular development in vivo. PloS one *8*, e81833.

Gibbons, R. (2006). Alpha thalassaemia-mental retardation, X linked. Orphanet journal of rare diseases *1*, 15.

Gibbons, R.J., McDowell, T.L., Raman, S., O'Rourke, D.M., Garrick, D., Ayyub, H., and Higgs, D.R. (2000). Mutations in ATRX, encoding a SWI/SNF-like protein, cause diverse changes in the pattern of DNA methylation. Nature genetics *24*, 368-371.

Gibbons, R.J., Picketts, D.J., Villard, L., and Higgs, D.R. (1995). Mutations in a putative global transcriptional regulator cause X-linked mental retardation with alpha-thalassemia (ATR-X syndrome). Cell *80*, 837-845.

Gibbons, R.J., Wada, T., Fisher, C.A., Malik, N., Mitson, M.J., Steensma, D.P., Fryer, A., Goudie, D.R., Krantz, I.D., and Traeger-Synodinos, J. (2008). Mutations in the chromatin-associated protein ATRX. Human mutation *29*, 796-802.

Gibson, D.G., Smith, H.O., Hutchison, C.A., 3rd, Venter, J.C., and Merryman, C. (2010). Chemical synthesis of the mouse mitochondrial genome. Nature methods *7*, 901-903.

Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., 3rd, and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nature methods *6*, 343-345.

Gifford, R., and Tristem, M. (2003). The evolution, distribution and diversity of endogenous retroviruses. Virus genes *26*, 291-315.

Gifford, W.D., Pfaff, S.L., and Macfarlan, T.S. (2013). Transposable elements as genetic regulatory substrates in early development. Trends in cell biology *23*, 218-226.

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.

Gonzalo, S., Jaco, I., Fraga, M.F., Chen, T., Li, E., Esteller, M., and Blasco, M.A. (2006). DNA methyltransferases control telomere length and telomere recombination in mammalian cells. Nature cell biology *8*, 416-424.

Greener, T., Zhao, X., Nojima, H., Eisenberg, E., and Greene, L.E. (2000). Role of cyclin G-associated kinase in uncoating clathrin-coated vesicles from non-neuronal cells. The Journal of biological chemistry *275*, 1365-1370.

Grewal, S.I., and Jia, S. (2007). Heterochromatin revisited. Nature reviews Genetics *8*, 35-46.

Grey, C., Barthes, P., Chauveau-Le Friec, G., Langa, F., Baudat, F., and de Massy, B. (2011). Mouse PRDM9 DNA-binding specificity determines sites of histone H3 lysine 4 trimethylation for initiation of meiotic recombination. PLoS biology *9*, e1001176.

Groner, A.C., Meylan, S., Ciuffi, A., Zangger, N., Ambrosini, G., Denervaud, N., Bucher, P., and Trono, D. (2010). KRAB-zinc finger proteins and KAP1 can mediate long-range transcriptional repression through heterochromatin spreading. PLoS genetics *6*, e1000869.

Groner, A.C., Tschopp, P., Challet, L., Dietrich, J.E., Verp, S., Offner, S., Barde, I., Rodriguez, I., Hiiragi, T., and Trono, D. (2012). The Kruppel-associated box repressor domain can induce reversible heterochromatization of a mouse locus in vivo. The Journal of biological chemistry *287*, 25361-25369.

Haas, D.L., Lutzko, C., Logan, A.C., Cho, G.J., Skelton, D., Jin Yu, X., Pepper, K.A., and Kohn, D.B. (2003). The Moloney murine leukemia virus repressor binding site represses expression in murine and human hematopoietic stem cells. Journal of virology *77*, 9439-9450.

Hahn, M., Dambacher, S., Dulev, S., Kuznetsova, A.Y., Eck, S., Worz, S., Sadic, D., Schulte, M., Mallm, J.P., Maiser, A.*, et al.* (2013). Suv4-20h2 mediates chromatin compaction and is important for cohesin recruitment to heterochromatin. Genes & development *27*, 859-872.

Hahn, M., Dambacher, S., and Schotta, G. (2010). Heterochromatin dysregulation in human diseases. Journal of applied physiology *109*, 232-242.

Hake, S.B., and Allis, C.D. (2006). Histone H3 variants and their potential role in indexing mammalian genomes: the "H3 barcode hypothesis". Proceedings of the National Academy of Sciences of the United States of America *103*, 6428-6435.

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.

Hamilton, M.D., Nuara, A.A., Gammon, D.B., Buller, R.M., and Evans, D.H. (2007). Duplex strand joining reactions catalyzed by vaccinia virus DNA polymerase. Nucleic acids research *35*, 143-151.

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

Heaphy, C.M., de Wilde, R.F., Jiao, Y., Klein, A.P., Edil, B.H., Shi, C., Bettegowda, C., Rodriguez, F.J., Eberhart, C.G., Hebbar, S.*, et al.* (2011). Altered telomeres in tumors with ATRX and DAXX mutations. Science *333*, 425.

Hediger, F., and Gasser, S.M. (2006). Heterochromatin protein 1: don't judge the book by its cover! Current opinion in genetics & development *16*, 143-150.

Heitz, E. (1928). Das Heterochromatin der Moose, 1. Jahrb Wiss Bot *69*, 762–818.

Helena Mangs, A., and Morris, B.J. (2007). The Human Pseudoautosomal Region (PAR): Origin, Function and Future. Current genomics *8*, 129-136.

Herquel, B., Ouararhni, K., Martianov, I., Le Gras, S., Ye, T., Keime, C., Lerouge, T., Jost, B., Cammas, F., Losson, R.*, et al.* (2013). Trim24-repressed VL30 retrotransposons regulate gene expression by producing noncoding RNA. Nature structural & molecular biology *20*, 339-346.

Hollenbach, A.D., McPherson, C.J., Mientjes, E.J., Iyengar, R., and Grosveld, G. (2002). Daxx and histone deacetylase II associate with chromatin through an interaction with core histones and the chromatin-associated protein Dek. Journal of cell science *115*, 3319-3330.

Hollenbach, A.D., Sublett, J.E., McPherson, C.J., and Grosveld, G. (1999). The Pax3-FKHR oncoprotein is unresponsive to the Pax3-associated repressor hDaxx. The EMBO journal *18*, 3702-3711.

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O.*, et al.* (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nature biotechnology *31*, 827-832.

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 & development *23*, 837-848.

Huang, H.S., Allen, J.A., Mabb, A.M., King, I.F., Miriyala, J., Taylor-Blake, B., Sciaky, N., Dutton, J.W., Jr., Lee, H.M., Chen, X.*, et al.* (2012). Topoisomerase inhibitors unsilence the dormant allele of Ube3a in neurons. Nature *481*, 185-189.

Huang, J., Fan, T., Yan, Q., Zhu, H., Fox, S., Issaq, H.J., Best, L., Gangi, L., Munroe, D., and Muegge, K. (2004). Lsh, an epigenetic guardian of repetitive elements. Nucleic acids research *32*, 5019-5028.

Huang, X., and Miller, W. (1991). A time-efficient, linear-space local similarity algorithm. Advances in Applied Mathematics *12*, 337-357.

Huisinga, K.L., Brower-Toland, B., and Elgin, S.C. (2006). The contradictory definitions of heterochromatin: transcription and silencing. Chromosoma *115*, 110-122.

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.

Ichimura, T., Watanabe, S., Sakamoto, Y., Aoto, T., Fujita, N., and Nakao, M. (2005). Transcriptional repression and heterochromatin formation by MBD1 and MCAF/AM family proteins. The Journal of biological chemistry *280*, 13928-13935.

Ishov, A.M., Sotnikov, A.G., Negorev, D., Vladimirova, O.V., Neff, N., Kamitani, T., Yeh, E.T., Strauss, J.F., 3rd, and Maul, G.G. (1999). PML is critical for ND10 formation and recruits the PML-interacting protein daxx to this nuclear structure when modified by SUMO-1. The Journal of cell biology *147*, 221-234.

Ishov, A.M., Vladimirova, O.V., and Maul, G.G. (2004). Heterochromatin and ND10 are cell-cycle regulated and phosphorylation-dependent alternate nuclear sites of the transcription repressor Daxx and SWI/SNF protein ATRX. Journal of cell science *117*, 3807-3820.

Ivanov, A.V., Peng, H., Yurchenko, V., Yap, K.L., Negorev, D.G., Schultz, D.C., Psulkowski, E., Fredericks, W.J., White, D.E., Maul, G.G*., et al.* (2007). PHD domain-mediated E3 ligase activity directs intramolecular sumoylation of an adjacent bromodomain required for gene silencing. Molecular cell *28*, 823-837.

Iwase, S., Xiang, B., Ghosh, S., Ren, T., Lewis, P.W., Cochrane, J.C., Allis, C.D., Picketts, D.J., Patel, D.J., Li, H*., et al.* (2011). ATRX ADD domain links an atypical histone methylation recognition mechanism to human mental-retardation syndrome. Nature structural & molecular biology *18*, 769-776.

Iyengar, S., and Farnham, P.J. (2011). KAP1 protein: an enigmatic master regulator of the genome. The Journal of biological chemistry *286*, 26267-26276.

Iyengar, S., Ivanov, A.V., Jin, V.X., Rauscher, F.J., 3rd, and Farnham, P.J. (2011). Functional analysis of KAP1 genomic recruitment. Molecular and cellular biology *31*, 1833-1847.

Jackson, A.L., Bartz, S.R., Schelter, J., Kobayashi, S.V., Burchard, J., Mao, M., Li, B., Cavet, G., and Linsley, P.S. (2003). Expression profiling reveals off-target gene regulation by RNAi. Nature biotechnology *21*, 635-637.

Jakobsson, J., Cordero, M.I., Bisaz, R., Groner, A.C., Busskamp, V., Bensadoun, J.C., Cammas, F., Losson, R., Mansuy, I.M., Sandi, C*., et al.* (2008). KAP1-mediated epigenetic repression in the forebrain modulates behavioral vulnerability to stress. Neuron *60*, 818-831.

Jern, P., and Coffin, J.M. (2008). Effects of retroviruses on host genome function. Annual review of genetics *42*, 709-732.

Jiao, Y., Killela, P.J., Reitman, Z.J., Rasheed, A.B., Heaphy, C.M., de Wilde, R.F., Rodriguez, F.J., Rosemberg, S., Oba-Shinjo, S.M., Nagahashi Marie, S.K*., et al.* (2012). Frequent ATRX, CIC, FUBP1 and IDH1 mutations refine the classification of malignant gliomas. Oncotarget *3*, 709-722.

Jiao, Y., Shi, C., Edil, B.H., de Wilde, R.F., Klimstra, D.S., Maitra, A., Schulick, R.D., Tang, L.H., Wolfgang, C.L., Choti, M.A*., et al.* (2011). DAXX/ATRX, MEN1, and mTOR pathway genes are frequently altered in pancreatic neuroendocrine tumors. Science *331*, 1199-1203.

Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nature reviews Genetics *13*, 484-492.

Jones, P.A., and Liang, G. (2009). Rethinking how DNA methylation patterns are maintained. Nature reviews Genetics *10*, 805-811.

Kanaoka, Y., Kimura, S.H., Okazaki, I., Ikeda, M., and Nojima, H. (1997). GAK: a cyclin G associated kinase contains a tensin/auxilin-like domain. FEBS letters *402*, 73-80.

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.

Kato, Y., Kaneda, M., Hata, K., Kumaki, K., Hisano, M., Kohara, Y., Okano, M., Li, E., Nozaki, M., and Sasaki, H. (2007). Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse. Human molecular genetics *16*, 2272-2280.

Kernohan, K.D., Jiang, Y., Tremblay, D.C., Bonvissuto, A.C., Eubanks, J.H., Mann, M.R., and Berube, N.G. (2010). ATRX partners with cohesin and MeCP2 and contributes to developmental silencing of imprinted genes in the brain. Developmental cell *18*, 191-202.

Kharchenko, P.V., Alekseyenko, A.A., Schwartz, Y.B., Minoda, A., Riddle, N.C., Ernst, J., Sabo, P.J., Larschan, E., Gorchakov, A.A., Gu, T.*, et al.* (2011). Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. Nature *471*, 480-485.

Kigami, D., Minami, N., Takayama, H., and Imai, H. (2003). MuERV-L is one of the earliest transcribed genes in mouse one-cell embryos. Biology of reproduction *68*, 651-654.

Kim, E.J., Park, J.S., and Um, S.J. (2003). Identification of Daxx interacting with p73, one of the p53 family, and its regulation of p53 activity by competitive interaction with PML. Nucleic acids research *31*, 5356-5367.

Kim, S.S., Chen, Y.M., O'Leary, E., Witzgall, R., Vidal, M., and Bonventre, J.V. (1996). A novel member of the RING finger family, KRIP-1, associates with the KRAB-A transcriptional repressor domain of zinc finger proteins. Proceedings of the National Academy of Sciences of the United States of America *93*, 15299-15304.

Kinameri, E., Inoue, T., Aruga, J., Imayoshi, I., Kageyama, R., Shimogori, T., and Moore, A.W. (2008). Prdm proto-oncogene transcription factor family expression and interaction with the Notch-Hes pathway in mouse neurogenesis. PloS one *3*, e3859.

King, I.F., Yandava, C.N., Mabb, A.M., Hsiao, J.S., Huang, H.S., Pearson, B.L., Calabrese, J.M., Starmer, J., Parker, J.S., Magnuson, T.*, et al.* (2013). Topoisomerases facilitate transcription of long genes linked to autism. Nature *501*, 58-62.

Kipling, D., Salido, E.C., Shapiro, L.J., and Cooke, H.J. (1996a). High frequency de novo alterations in the long-range genomic structure of the mouse pseudoautosomal region. Nature genetics *13*, 78-80.

Kipling, D., Wilson, H.E., Thomson, E.J., Lee, M., Perry, J., Palmer, S., Ashworth, A., and Cooke, H.J. (1996b). Structural variation of the pseudoautosomal region between and within inbred mouse strains. Proceedings of the National Academy of Sciences of the United States of America *93*, 171-175.

Koike-Yusa, H., Li, Y., Tan, E.P., Velasco-Herrera, M.D., and Yusa, K. (2013). Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. Nature biotechnology.

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

Kourmouli, N., Sun, Y.M., van der Sar, S., Singh, P.B., and Brown, J.P. (2005). Epigenetic regulation of mammalian pericentric heterochromatin in vivo by HP1. Biochemical and biophysical research communications *337*, 901-907.

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

Krebs, C.J., Khan, S., MacDonald, J.W., Sorenson, M., and Robins, D.M. (2009). Regulator of sex-limitation KRAB zinc finger proteins modulate sex-dependent and -independent liver metabolism. Physiological genomics *38*, 16-28.

Krebs, C.J., Larkins, L.K., Price, R., Tullis, K.M., Miller, R.D., and Robins, D.M. (2003). Regulator of sex-limitation (Rsl) encodes a pair of KRAB zinc-finger genes that control sexually dimorphic liver gene expression. Genes & development *17*, 2664-2674.

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.

Kulis, M., and Esteller, M. (2010). DNA methylation and cancer. Advances in genetics *70*, 27-56.

Kuo, H.Y., Chang, C.C., Jeng, J.C., Hu, H.M., Lin, D.Y., Maul, G.G., Kwok, R.P., and Shih, H.M. (2005). SUMO modification negatively modulates the transcriptional activity of CREB-binding protein via the recruitment of Daxx. Proceedings of the National Academy of Sciences of the United States of America *102*, 16973-16978.

Kurth, R., and Bannert, N. (2010). Beneficial and detrimental effects of human endogenous retroviruses. International journal of cancer Journal international du cancer *126*, 306-314.

Kwon, S.H., and Workman, J.L. (2011). The changing faces of HP1: From heterochromatin formation and gene silencing to euchromatic gene expression: HP1 acts as a positive regulator of transcription. BioEssays : news and reviews in molecular, cellular and developmental biology *33*, 280-289.

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.

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

Law, J.A., and Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nature reviews Genetics *11*, 204-220.

Law, M.J., Lower, K.M., Voon, H.P., Hughes, J.R., Garrick, D., Viprakasit, V., Mitson, M., De Gobbi, M., Marra, M., Morris, A.*, et al.* (2010). ATR-X syndrome protein targets tandem repeats and influences allele-specific expression in a size-dependent manner. Cell *143*, 367-378.

Le Douarin, B., Nielsen, A.L., Garnier, J.M., Ichinose, H., Jeanmougin, F., Losson, R., and Chambon, P. (1996). A possible involvement of TIF1 alpha and TIF1 beta in the epigenetic control of transcription by nuclear receptors. The EMBO journal *15*, 6701-6715.

Lechner, M.S., Begg, G.E., Speicher, D.W., and Rauscher, F.J., 3rd (2000). Molecular determinants for targeting heterochromatin protein 1-mediated gene silencing: direct chromoshadow domain-KAP-1 corepressor interaction is essential. Molecular and cellular biology *20*, 6449-6465.

Lehembre, F., Muller, S., Pandolfi, P.P., and Dejean, A. (2001). Regulation of Pax3 transcriptional activity by SUMO-1-modified PML. Oncogene *20*, 1-9.

Lehnertz, B., Ueda, Y., Derijck, A.A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T., Li, E., Jenuwein, T., and Peters, A.H. (2003). Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. Current biology : CB *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. Proceedings of the National Academy of Sciences of the United States of America *108*, 5718-5723.

Leung, D.C., and Lorincz, M.C. (2012). Silencing of endogenous retroviruses: when and why do histone marks predominate? Trends in biochemical sciences *37*, 127-133.

Lewis, P.W., Elsaesser, S.J., Noh, K.M., Stadler, S.C., and Allis, C.D. (2010). Daxx is an H3.3-specific histone chaperone and cooperates with ATRX in replication-independent chromatin assembly at telomeres. Proceedings of the National Academy of Sciences of the United States of America *107*, 14075-14080.

Lewis, P.W., Muller, M.M., Koletsky, M.S., Cordero, F., Lin, S., Banaszynski, L.A., Garcia, B.A., Muir, T.W., Becher, O.J., and Allis, C.D. (2013). Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma. Science *340*, 857-861.

Li, H., Leo, C., Zhu, J., Wu, X., O'Neil, J., Park, E.J., and Chen, J.D. (2000a). Sequestration and inhibition of Daxx-mediated transcriptional repression by PML. Molecular and cellular biology *20*, 1784-1796.

Li, H., Rauch, T., Chen, Z.X., Szabo, P.E., Riggs, A.D., and Pfeifer, G.P. (2006). The histone methyltransferase SETDB1 and the DNA methyltransferase DNMT3A interact directly and localize to promoters silenced in cancer cells. The Journal of biological chemistry *281*, 19489-19500.

Li, L.C., and Dahiya, R. (2002). MethPrimer: designing primers for methylation PCRs. Bioinformatics *18*, 1427-1431.

Li, R., Pei, H., Watson, D.K., and Papas, T.S. (2000b). EAP1/Daxx interacts with ETS1 and represses transcriptional activation of ETS1 target genes. Oncogene *19*, 745-753.

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.

Li, X., Lee, Y.K., Jeng, J.C., Yen, Y., Schultz, D.C., Shih, H.M., and Ann, D.K. (2007). Role for KAP1 serine 824 phosphorylation and sumoylation/desumoylation switch in regulating KAP1-mediated transcriptional repression. The Journal of biological chemistry *282*, 36177-36189.

Li, Z., Wang, D., Na, X., Schoen, S.R., Messing, E.M., and Wu, G. (2003). The VHL protein recruits a novel KRAB-A domain protein to repress HIF-1alpha transcriptional activity. The EMBO journal *22*, 1857-1867.

Lienert, F., Mohn, F., Tiwari, V.K., Baubec, T., Roloff, T.C., Gaidatzis, D., Stadler, M.B., and Schubeler, D. (2011a). Genomic prevalence of heterochromatic H3K9me2 and transcription do not discriminate pluripotent from terminally differentiated cells. PLoS genetics *7*, e1002090.

Lienert, F., Wirbelauer, C., Som, I., Dean, A., Mohn, F., and Schubeler, D. (2011b). Identification of genetic elements that autonomously determine DNA methylation states. Nature genetics *43*, 1091-1097.

Lin, D.Y., Fang, H.I., Ma, A.H., Huang, Y.S., Pu, Y.S., Jenster, G., Kung, H.J., and Shih, H.M. (2004). Negative modulation of androgen receptor transcriptional activity by Daxx. Molecular and cellular biology *24*, 10529-10541.

Lin, D.Y., Huang, Y.S., Jeng, J.C., Kuo, H.Y., Chang, C.C., Chao, T.T., Ho, C.C., Chen, Y.C., Lin, T.P., Fang, H.I., *et al.* (2006). Role of SUMO-interacting motif in Daxx SUMO modification, subnuclear localization, and repression of sumoylated transcription factors. Molecular cell *24*, 341-354.

Lin, D.Y., Lai, M.Z., Ann, D.K., and Shih, H.M. (2003). Promyelocytic leukemia protein (PML) functions as a glucocorticoid receptor co-activator by sequestering Daxx to the PML oncogenic domains (PODs) to enhance its transactivation potential. The Journal of biological chemistry *278*, 15958-15965.

Lin, D.Y., and Shih, H.M. (2002). Essential role of the 58-kDa microspherule protein in the modulation of Daxx-dependent transcriptional repression as revealed by nucleolar sequestration. The Journal of biological chemistry *277*, 25446-25456.

Lindsay, C.R., Giovinazzi, S., and Ishov, A.M. (2009). Daxx is a predominately nuclear protein that does not translocate to the cytoplasm in response to cell stress. Cell cycle *8*, 1544-1551.

Liu, C.P., Xiong, C., Wang, M., Yu, Z., Yang, N., Chen, P., Zhang, Z., Li, G., and Xu, R.M. (2012). Structure of the variant histone H3.3-H4 heterodimer in complex with its chaperone DAXX. Nature structural & molecular biology *19*, 1287-1292.

Loh, T.P., Sievert, L.L., and Scott, R.W. (1987). Proviral sequences that restrict retroviral expression in mouse embryonal carcinoma cells. Molecular and cellular biology *7*, 3775-3784.

Lohmann, F., Loureiro, J., Su, H., Fang, Q., Lei, H., Lewis, T., Yang, Y., Labow, M., Li, E., Chen, T., *et al.* (2010). KMT1E mediated H3K9 methylation is required for the maintenance of embryonic stem cells by repressing trophectoderm differentiation. Stem cells *28*, 201-212.

Lovejoy, C.A., Li, W., Reisenweber, S., Thongthip, S., Bruno, J., de Lange, T., De, S., Petrini, J.H., Sung, P.A., Jasin, M., *et al.* (2012). Loss of ATRX, genome instability, and an altered DNA damage response are hallmarks of the alternative lengthening of telomeres pathway. PLoS genetics *8*, e1002772.

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 reports *10*, 769-775.

Lueders, K.K., and Kuff, E.L. (1977). Sequences associated with intracisternal A particles are reiterated in the mouse genome. Cell *12*, 963-972.

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.

Lyst, M.J., Nan, X., and Stancheva, I. (2006). Regulation of MBD1-mediated transcriptional repression by SUMO and PIAS proteins. The EMBO journal *25*, 5317-5328.

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.* (2011). Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A. Genes & development *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.

Macgregor, S., Montgomery, G.W., Liu, J.Z., Zhao, Z.Z., Henders, A.K., Stark, M., Schmid, H., Holland, E.A., Duffy, D.L., Zhang, M.*, et al.* (2011). Genome-wide association study identifies a new melanoma susceptibility locus at 1q21.3. Nature genetics *43*, 1114-1118.

Magklara, A., Yen, A., Colquitt, B.M., Clowney, E.J., Allen, W., Markenscoff-Papadimitriou, E., Evans, Z.A., Kheradpour, P., Mountoufaris, G., Carey, C.*, et al.* (2011). An epigenetic signature for monoallelic olfactory receptor expression. Cell *145*, 555-570.

Mahtani, M.M., and Willard, H.F. (1998). Physical and genetic mapping of the human X chromosome centromere: repression of recombination. Genome research *8*, 100-110.

Maksakova, I.A., Goyal, P., Bullwinkel, J., Brown, J.P., Bilenky, M., Mager, D.L., Singh, P.B., and Lorincz, M.C. (2011). H3K9me3-binding proteins are dispensable for SETDB1/H3K9me3-dependent retroviral silencing. Epigenetics & chromatin *4*, 12.

Maksakova, I.A., Romanish, M.T., Gagnier, L., Dunn, C.A., van de Lagemaat, L.N., and Mager, D.L. (2006). Retroviral elements and their hosts: insertional mutagenesis in the mouse germ line. PLoS genetics *2*, e2.

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.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. Science *339*, 823-826.

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 research *11*, 1187-1197.

Margolin, J.F., Friedman, J.R., Meyer, W.K., Vissing, H., Thiesen, H.J., and Rauscher, F.J., 3rd (1994). Kruppel-associated boxes are potent transcriptional repression domains. Proceedings of the National Academy of Sciences of the United States of America *91*, 4509-4513.

Mather, K. (1939). Crossing over and Heterochromatin in the X Chromosome of Drosophila Melanogaster. Genetics *24*, 413-435.

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.

McCarthy, E.M., and McDonald, J.F. (2004). Long terminal repeat retrotransposons of Mus musculus. Genome biology *5*, R14.

McDowell, T.L., Gibbons, R.J., Sutherland, H., O'Rourke, D.M., Bickmore, W.A., Pombo, A., Turley, H., Gatter, K., Picketts, D.J., Buckle, V.J.*, et al.* (1999). Localization of a putative transcriptional regulator (ATRX) at pericentromeric heterochromatin and the short arms of acrocentric chromosomes. Proceedings of the National Academy of Sciences of the United States of America *96*, 13983-13988.

Medstrand, P., and Mager, D.L. (1998). Human-specific integrations of the HERV-K endogenous retrovirus family. Journal of virology *72*, 9782-9787.

Meshorer, E., Yellajoshula, D., George, E., Scambler, P.J., Brown, D.T., and Misteli, T. (2006). Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. Developmental cell *10*, 105-116.

Messerschmidt, D.M., de Vries, W., Ito, M., Solter, D., Ferguson-Smith, A., and Knowles, B.B. (2012). Trim28 is required for epigenetic stability during mouse oocyte to embryo transition. Science *335*, 1499-1502.

Michaelson, J.S., Bader, D., Kuo, F., Kozak, C., and Leder, P. (1999). Loss of Daxx, a promiscuously interacting protein, results in extensive apoptosis in early mouse development. Genes & development *13*, 1918-1923.

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.

Moffat, J., Grueneberg, D.A., Yang, X., Kim, S.Y., Kloepfer, A.M., Hinkle, G., Piqani, B., Eisenhaure, T.M., Luo, B., Grenier, J.K., *et al.* (2006). A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. Cell *124*, 1283-1298.

Molenaar, J.J., Koster, J., Zwijnenburg, D.A., van Sluis, P., Valentijn, L.J., van der Ploeg, I., Hamdi, M., van Nes, J., Westerman, B.A., van Arkel, J., *et al.* (2012). Sequencing of neuroblastoma identifies chromothripsis and defects in neuritogenesis genes. Nature *483*, 589-593.

Moosmann, P., Georgiev, O., Le Douarin, B., Bourquin, J.P., and Schaffner, W. (1996). Transcriptional repression by RING finger protein TIF1 beta that interacts with the KRAB repressor domain of KOX1. Nucleic acids research *24*, 4859-4867.

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.

Muotri, A.R., Chu, V.T., Marchetto, M.C., Deng, W., Moran, J.V., and Gage, F.H. (2005). Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. Nature *435*, 903-910.

Muromoto, R., Sugiyama, K., Takachi, A., Imoto, S., Sato, N., Yamamoto, T., Oritani, K., Shimoda, K., and Matsuda, T. (2004). Physical and functional interactions between Daxx and DNA methyltransferase 1-associated protein, DMAP1. Journal of immunology *172*, 2985-2993.

Nakamura, T., Arai, Y., Umehara, H., Masuhara, M., Kimura, T., Taniguchi, H., Sekimoto, T., Ikawa, M., Yoneda, Y., Okabe, M., *et al.* (2007). PGC7/Stella protects against DNA demethylation in early embryogenesis. Nature cell biology *9*, 64-71.

Nakatake, Y., Fujii, S., Masui, S., Sugimoto, T., Torikai-Nishikawa, S., Adachi, K., and Niwa, H. (2013). Kinetics of drug selection systems in mouse embryonic stem cells. BMC biotechnology *13*, 64.

Naldini, L., Blomer, U., Gallay, P., Ory, D., Mulligan, R., Gage, F.H., Verma, I.M., and Trono, D. (1996). In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. Science *272*, 263-267.

Nan, X., Hou, J., Maclean, A., Nasir, J., Lafuente, M.J., Shu, X., Kriaucionis, S., and Bird, A. (2007). Interaction between chromatin proteins MECP2 and ATRX is disrupted by mutations that cause inherited mental retardation. Proceedings of the National Academy of Sciences of the United States of America *104*, 2709-2714.

Ngo, V.N., Davis, R.E., Lamy, L., Yu, X., Zhao, H., Lenz, G., Lam, L.T., Dave, S., Yang, L., Powell, J., *et al.* (2006). A loss-of-function RNA interference screen for molecular targets in cancer. Nature *441*, 106-110.

Nicetto, D., Hahn, M., Jung, J., Schneider, T.D., Straub, T., David, R., Schotta, G., and Rupp, R.A. (2013). Suv4-20h histone methyltransferases promote neuroectodermal differentiation by silencing the pluripotency-associated Oct-25 gene. PLoS genetics *9*, e1003188.

Nightingale, S.J., Hollis, R.P., Pepper, K.A., Petersen, D., Yu, X.J., Yang, C., Bahner, I., and Kohn, D.B. (2006). Transient gene expression by nonintegrating lentiviral vectors. Molecular therapy : the journal of the American Society of Gene Therapy *13*, 1121-1132.

Noe, L., and Kucherov, G. (2005). YASS: enhancing the sensitivity of DNA similarity search. Nucleic acids research *33*, W540-543.

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

Obradovic, D., Tirard, M., Nemethy, Z., Hirsch, O., Gronemeyer, H., and Almeida, O.F. (2004). DAXX, FLASH, and FAF-1 modulate mineralocorticoid and glucocorticoid receptor-mediated transcription in hippocampal cells--toward a basis for the opposite actions elicited by two nuclear receptors? Molecular pharmacology *65*, 761-769.

Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell *99*, 247-257.

Palmer, S., Perry, J., Kipling, D., and Ashworth, A. (1997). A gene spans the pseudoautosomal boundary in mice. Proceedings of the National Academy of Sciences of the United States of America *94*, 12030-12035.

Pannell, D., Osborne, C.S., Yao, S., Sukonnik, T., Pasceri, P., Karaiskakis, A., Okano, M., Li, E., Lipshitz, H.D., and Ellis, J. (2000). Retrovirus vector silencing is de novo methylase independent and marked by a repressive histone code. The EMBO journal *19*, 5884-5894.

Park, J., Lee, J.H., La, M., Jang, M.J., Chae, G.W., Kim, S.B., Tak, H., Jung, Y., Byun, B., Ahn, J.K.*, et al.* (2007). Inhibition of NF-kappaB acetylation and its transcriptional activity by Daxx. Journal of molecular biology *368*, 388-397.

Parvanov, E.D., Petkov, P.M., and Paigen, K. (2010). Prdm9 controls activation of mammalian recombination hotspots. Science *327*, 835.

Peaston, A.E., Evsikov, A.V., Graber, J.H., de Vries, W.N., Holbrook, A.E., Solter, D., and Knowles, B.B. (2004). Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. Developmental cell *7*, 597-606.

Peng, H., Begg, G.E., Harper, S.L., Friedman, J.R., Speicher, D.W., and Rauscher, F.J., 3rd (2000a). Biochemical analysis of the Kruppel-associated box (KRAB) transcriptional repression domain. The Journal of biological chemistry *275*, 18000-18010.

Peng, H., Begg, G.E., Schultz, D.C., Friedman, J.R., Jensen, D.E., Speicher, D.W., and Rauscher, F.J., 3rd (2000b). Reconstitution of the KRAB-KAP-1 repressor complex: a model system for defining the molecular anatomy of RING-B box-coiled-coil domain-mediated protein-protein interactions. Journal of molecular biology *295*, 1139-1162.

Peng, H., Feldman, I., and Rauscher, F.J., 3rd (2002). Hetero-oligomerization among the TIF family of RBCC/TRIM domain-containing nuclear cofactors: a potential mechanism for regulating the switch between coactivation and corepression. Journal of molecular biology *320*, 629-644.

Peng, H., Ivanov, A.V., Oh, H.J., Lau, Y.F., and Rauscher, F.J., 3rd (2009). Epigenetic gene silencing by the SRY protein is mediated by a KRAB-O protein that recruits the KAP1 co-repressor machinery. The Journal of biological chemistry *284*, 35670-35680.

Perrat, P.N., DasGupta, S., Wang, J., Theurkauf, W., Weng, Z., Rosbash, M., and Waddell, S. (2013). Transposition-driven genomic heterogeneity in the Drosophila brain. Science *340*, 91-95.

Perry, J., Palmer, S., Gabriel, A., and Ashworth, A. (2001). A short pseudoautosomal region in laboratory mice. Genome research *11*, 1826-1832.

Peters, A.H., Kubicek, S., Mechtler, K., O'Sullivan, R.J., Derijck, A.A., Perez-Burgos, L., Kohlmaier, A., Opravil, S., Tachibana, M., Shinkai, Y.*, et al.* (2003). Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. Molecular cell *12*, 1577-1589.

Peters, A.H., Mermoud, J.E., O'Carroll, D., Pagani, M., Schweizer, D., Brockdorff, N., and Jenuwein, T. (2002). Histone H3 lysine 9 methylation is an epigenetic imprint of facultative heterochromatin. Nature genetics *30*, 77-80.

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.

Picketts, D.J., Higgs, D.R., Bachoo, S., Blake, D.J., Quarrell, O.W., and Gibbons, R.J. (1996). ATRX encodes a novel member of the SNF2 family of proteins: mutations point to a common mechanism underlying the ATR-X syndrome. Human molecular genetics *5*, 1899-1907.

Pinheiro, I., Margueron, R., Shukeir, N., Eisold, M., Fritzsch, C., Richter, F.M., Mittler, G., Genoud, C., Goyama, S., Kurokawa, M.*, et al.* (2012). Prdm3 and Prdm16 are H3K9me1 methyltransferases required for mammalian heterochromatin integrity. Cell *150*, 948-960.

Poznanski, A.A., and Calarco, P.G. (1991). The expression of intracisternal A particle genes in the preimplantation mouse embryo. Developmental biology *143*, 271-281.

Quenneville, S., Turelli, P., Bojkowska, K., Raclot, C., Offner, S., Kapopoulou, A., and Trono, D. (2012). The KRAB-ZFP/KAP1 system contributes to the early embryonic establishment of site-specific DNA methylation patterns maintained during development. Cell reports *2*, 766-773.

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. Molecular cell *44*, 361-372.

Ratnakumar, K., and Bernstein, E. (2013). ATRX: the case of a peculiar chromatin remodeler. Epigenetics : official journal of the DNA Methylation Society *8*, 3-9.

Ratnakumar, K., Duarte, L.F., LeRoy, G., Hasson, D., Smeets, D., Vardabasso, C., Bonisch, C., Zeng, T., Xiang, B., Zhang, D.Y*., et al.* (2012). ATRX-mediated chromatin association of histone variant macroH2A1 regulates alpha-globin expression. Genes & development *26*, 433-438.

Ribet, D., Dewannieux, M., and Heidmann, T. (2004). An active murine transposon family pair: retrotransposition of "master" MusD copies and ETn trans-mobilization. Genome research *14*, 2261-2267.

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 research *18*, 597-609.

Ritchie, K., Seah, C., Moulin, J., Isaac, C., Dick, F., and Berube, N.G. (2008). Loss of ATRX leads to chromosome cohesion and congression defects. The Journal of cell biology *180*, 315-324.

Rodriguez-Paredes, M., Martinez de Paz, A., Simo-Riudalbas, L., Sayols, S., Moutinho, C., Moran, S., Villanueva, A., Vazquez-Cedeira, M., Lazo, P.A., Carneiro, F*., et al.* (2013). Gene amplification of the histone methyltransferase SETDB1 contributes to human lung tumorigenesis. Oncogene.

Ross, R.J., Weiner, M.M., and Lin, H. (2014). PIWI proteins and PIWI-interacting RNAs in the soma. Nature *505*, 353-359.

Rowe, H.M., Friedli, M., Offner, S., Verp, S., Mesnard, D., Marquis, J., Aktas, T., and Trono, D. (2013a). De novo DNA methylation of endogenous retroviruses is shaped by KRAB-ZFPs/KAP1 and ESET. Development *140*, 519-529.

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. (2013b). TRIM28 repression of retrotransposon-based enhancers is necessary to preserve transcriptional dynamics in embryonic stem cells. Genome research *23*, 452-461.

Rowe, H.M., and Trono, D. (2011). Dynamic control of endogenous retroviruses during development. Virology *411*, 273-287.

Ryan, R.F., Schultz, D.C., Ayyanathan, K., Singh, P.B., Friedman, J.R., Fredericks, W.J., and Rauscher, F.J., 3rd (1999). KAP-1 corepressor protein interacts and colocalizes with heterochromatic and euchromatic HP1 proteins: a potential role for Kruppel-associated box-zinc finger proteins in heterochromatin-mediated gene silencing. Molecular and cellular biology *19*, 4366-4378.

Saitou, M., Kagiwada, S., and Kurimoto, K. (2012). Epigenetic reprogramming in mouse pre-implantation development and primordial germ cells. Development *139*, 15-31.

Sambrook, J., Russell, D.W., Fritsch, E.F., and Maniatis, T. (2001). Molecular cloning : a laboratory manual, 3rd ed edn (Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press).

Santoni de Sio, F.R., Barde, I., Offner, S., Kapopoulou, A., Corsinotti, A., Bojkowska, K., Genolet, R., Thomas, J.H., Luescher, I.F., Pinschewer, D*., et al.* (2012a). KAP1 regulates gene networks controlling T-cell development and responsiveness. FASEB journal : official publication of the Federation of American Societies for Experimental Biology *26*, 4561-4575.

Santoni de Sio, F.R., Massacand, J., Barde, I., Offner, S., Corsinotti, A., Kapopoulou, A., Bojkowska, K., Dagklis, A., Fernandez, M., Ghia, P*., et al.* (2012b). KAP1 regulates gene networks controlling mouse B-lymphoid cell differentiation and function. Blood *119*, 4675-4685.

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. Molecular cell *15*, 595-605.

Sato, J., Shimizu, H., Kasama, T., Yabuta, N., and Nojima, H. (2009). GAK, a regulator of clathrin-mediated membrane trafficking, localizes not only in the cytoplasm but also in the nucleus. Genes to cells : devoted to molecular & cellular mechanisms *14*, 627-641.

Sawatsubashi, S., Murata, T., Lim, J., Fujiki, R., Ito, S., Suzuki, E., Tanabe, M., Zhao, Y., Kimura, S., Fujiyama, S.*, et al.* (2010). A histone chaperone, DEK, transcriptionally coactivates a nuclear receptor. Genes & development *24*, 159-170.

Schambach, A., Galla, M., Modlich, U., Will, E., Chandra, S., Reeves, L., Colbert, M., Williams, D.A., von Kalle, C., and Baum, C. (2006). Lentiviral vectors pseudotyped with murine ecotropic envelope: increased biosafety and convenience in preclinical research. Experimental hematology *34*, 588-592.

Schlesinger, S., Lee, A.H., Wang, G.Z., Green, L., and Goff, S.P. (2013). Proviral silencing in embryonic cells is regulated by Yin Yang 1. Cell reports *4*, 50-58.

Schmidt, K. (2014). thesis in preparation (Ludwig-Maximilians Universität München).

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 & development *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 & development *22*, 2048-2061.

Schreiner, S., and Wodrich, H. (2013). Virion factors that target Daxx to overcome intrinsic immunity. Journal of virology *87*, 10412-10422.

Schroder, A.R., Shinn, P., Chen, H., Berry, C., Ecker, J.R., and Bushman, F. (2002). HIV-1 integration in the human genome favors active genes and local hotspots. Cell *110*, 521-529.

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 & development *16*, 919-932.

Schultz, D.C., Friedman, J.R., and Rauscher, F.J., 3rd (2001). Targeting histone deacetylase complexes via KRAB-zinc finger proteins: the PHD and bromodomains of KAP-1 form a cooperative unit that recruits a novel isoform of the Mi-2alpha subunit of NuRD. Genes & development *15*, 428-443.

Schwartzentruber, J., Korshunov, A., Liu, X.Y., Jones, D.T., Pfaff, E., Jacob, K., Sturm, D., Fontebasso, A.M., Quang, D.A., Tonjes, M.*, et al.* (2012). Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. Nature *482*, 226-231.

Seila, A.C., Calabrese, J.M., Levine, S.S., Yeo, G.W., Rahl, P.B., Flynn, R.A., Young, R.A., and Sharp, P.A. (2008). Divergent transcription from active promoters. Science *322*, 1849-1851.

Sekiyama, N., Ikegami, T., Yamane, T., Ikeguchi, M., Uchimura, Y., Baba, D., Ariyoshi, M., Tochio, H., Saitoh, H., and Shirakawa, M. (2008). Structure of the small ubiquitin-like modifier (SUMO)-interacting motif of MBD1-containing chromatin-associated factor 1 bound to SUMO-3. The Journal of biological chemistry *283*, 35966-35975.

Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelsen, T.S., Heckl, D., Ebert, B.L., Root, D.E., Doench, J.G.*, et al.* (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. Science *343*, 84-87.

Shen, J.C., Rideout, W.M., 3rd, and Jones, P.A. (1992). High frequency mutagenesis by a DNA methyltransferase. Cell *71*, 1073-1080.

Shibata, M., Blauvelt, K.E., Liem, K.F., Jr., and Garcia-Garcia, M.J. (2011). TRIM28 is required by the mouse KRAB domain protein ZFP568 to control convergent extension and morphogenesis of extra-embryonic tissues. Development *138*, 5333-5343.

Shih, H.M., Chang, C.C., Kuo, H.Y., and Lin, D.Y. (2007). Daxx mediates SUMO-dependent transcriptional control and subnuclear compartmentalization. Biochemical Society transactions *35*, 1397-1400.

Shimizu, H., Nagamori, I., Yabuta, N., and Nojima, H. (2009). GAK, a regulator of clathrin-mediated membrane traffic, also controls centrosome integrity and chromosome congression. Journal of cell science *122*, 3145-3152.

Silva, F.P., Hamamoto, R., Furukawa, Y., and Nakamura, Y. (2006). TIPUH1 encodes a novel KRAB zinc-finger protein highly expressed in human hepatocellular carcinomas. Oncogene *25*, 5063-5070.

Silva, J.M., Marran, K., Parker, J.S., Silva, J., Golding, M., Schlabach, M.R., Elledge, S.J., Hannon, G.J., and Chang, K. (2008). Profiling essential genes in human mammary cells by multiplex RNAi screening. Science *319*, 617-620.

Smothers, J.F., and Henikoff, S. (2000). The HP1 chromo shadow domain binds a consensus peptide pentamer. Current biology : CB *10*, 27-30.

Sripathy, S.P., Stevens, J., and Schultz, D.C. (2006). The KAP1 corepressor functions to coordinate the assembly of de novo HP1-demarcated microenvironments of heterochromatin required for KRAB zinc finger protein-mediated transcriptional repression. Molecular and cellular biology *26*, 8623-8638.

Stetson, D.B., Ko, J.S., Heidmann, T., and Medzhitov, R. (2008). Trex1 prevents cell-intrinsic initiation of autoimmunity. Cell *134*, 587-598.

Stielow, B., Sapetschnig, A., Wink, C., Kruger, I., and Suske, G. (2008). SUMO-modified Sp3 represses transcription by provoking local heterochromatic gene silencing. EMBO reports *9*, 899-906.

Stocking, C., and Kozak, C.A. (2008). Murine endogenous retroviruses. Cellular and molecular life sciences : CMLS *65*, 3383-3398.

Svoboda, P., Stein, P., Anger, M., Bernstein, E., Hannon, G.J., and Schultz, R.M. (2004). RNAi and expression of retrotransposons MuERV-L and IAP in preimplantation mouse embryos. Developmental biology *269*, 276-285.

Tachibana, M., Matsumura, Y., Fukuda, M., Kimura, H., and Shinkai, Y. (2008). G9a/GLP complexes independently mediate H3K9 and DNA methylation to silence transcription. The EMBO journal *27*, 2681-2690.

Tadepally, H.D., Burger, G., and Aubry, M. (2008). Evolution of C2H2-zinc finger genes and subfamilies in mammals: species-specific duplication and loss of clusters, genes and effector domains. BMC evolutionary biology *8*, 176.

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

Tan, S.L., Nishi, M., Ohtsuka, T., Matsui, T., Takemoto, K., Kamio-Miura, A., Aburatani, H., Shinkai, Y., and Kageyama, R. (2012). Essential roles of the histone methyltransferase ESET in the epigenetic control of neural progenitor cells during development. Development *139*, 3806-3816.

Tanaka, N., and Saitoh, H. (2010). A real-time SUMO-binding assay for the analysis of the SUMO-SIM protein interaction network. Bioscience, biotechnology, and biochemistry *74*, 1302-1305.

Tanenbaum, M.E., Vallenius, T., Geers, E.F., Greene, L., Makela, T.P., and Medema, R.H. (2010). Cyclin G-associated kinase promotes microtubule outgrowth from chromosomes during spindle assembly. Chromosoma *119*, 415-424.

Tang, J., Wu, S., Liu, H., Stratt, R., Barak, O.G., Shiekhattar, R., Picketts, D.J., and Yang, X. (2004). A novel transcription regulatory complex containing death domain-associated protein and the ATR-X syndrome protein. The Journal of biological chemistry *279*, 20369-20377.

Teich, N.M., Weiss, R.A., Martin, G.R., and Lowy, D.R. (1977). Virus infection of murine teratocarcinoma stem cell lines. Cell *12*, 973-982.

Thomas, J.H., and Emerson, R.O. (2009). Evolution of C2H2-zinc finger genes revisited. BMC evolutionary biology *9*, 51.

Thomas, J.H., and Schneider, S. (2011). Coevolution of retroelements and tandem zinc finger genes. Genome research *21*, 1800-1812.

Ting, C.N., Rosenberg, M.P., Snow, C.M., Samuelson, L.C., and Meisler, M.H. (1992). Endogenous retroviral sequences are required for tissue-specific expression of a human salivary amylase gene. Genes & development *6*, 1457-1465.

Ting, D.T., Lipson, D., Paul, S., Brannigan, B.W., Akhavanfard, S., Coffman, E.J., Contino, G., Deshpande, V., Iafrate, A.J., Letovsky, S*., et al.* (2011). Aberrant overexpression of satellite repeats in pancreatic and other epithelial cancers. Science *331*, 593-596.

Torii, S., Egan, D.A., Evans, R.A., and Reed, J.C. (1999). Human Daxx regulates Fas-induced apoptosis from nuclear PML oncogenic domains (PODs). The EMBO journal *18*, 6037-6049.

Uchimura, Y., Ichimura, T., Uwada, J., Tachibana, T., Sugahara, S., Nakao, M., and Saitoh, H. (2006). Involvement of SUMO modification in MBD1- and MCAF1-mediated heterochromatin formation. The Journal of biological chemistry *281*, 23180-23190.

Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., and Leunissen, J.A. (2007). Primer3Plus, an enhanced web interface to Primer3. Nucleic acids research *35*, W71-74.

van de Lagemaat, L.N., Landry, J.R., Mager, D.L., and Medstrand, P. (2003). Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. Trends in genetics : TIG *19*, 530-536.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., and Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome biology *3*, RESEARCH0034.

Varga-Weisz, P.D., Wilm, M., Bonte, E., Dumas, K., Mann, M., and Becker, P.B. (1997). Chromatin-remodelling factor CHRAC contains the ATPases ISWI and topoisomerase II. Nature *388*, 598-602.

Vermeulen, M., Eberl, H.C., Matarese, F., Marks, H., Denissov, S., Butter, F., Lee, K.K., Olsen, J.V., Hyman, A.A., Stunnenberg, H.G*., et al.* (2010). Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. Cell *142*, 967-980.

Verschure, P.J., van der Kraan, I., de Leeuw, W., van der Vlag, J., Carpenter, A.E., Belmont, A.S., and van Driel, R. (2005). In vivo HP1 targeting causes large-scale chromatin condensation and enhanced histone lysine methylation. Molecular and cellular biology *25*, 4552-4564.

Vos, S.M., Tretter, E.M., Schmidt, B.H., and Berger, J.M. (2011). All tangled up: how cells direct, manage and exploit topoisomerase function. Nature reviews Molecular cell biology *12*, 827-841.

Walsh, C.P., Chaillet, J.R., and Bestor, T.H. (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. Nature genetics *20*, 116-117.

Wang, G.Z., Wolf, D., and Goff, S.P. (2014a). EBP1, a novel host factor involved in primer binding site-dependent restriction of moloney murine leukemia virus in embryonic cells. Journal of virology *88*, 1825-1829.

Wang, H., An, W., Cao, R., Xia, L., Erdjument-Bromage, H., Chatton, B., Tempst, P., Roeder, R.G., and Zhang, Y. (2003). mAM facilitates conversion by ESET of dimethyl to trimethyl lysine 9 of histone H3 to cause transcriptional repression. Molecular cell *12*, 475-487.

Wang, H., Yang, H., Shivalila, C.S., Dawlaty, M.M., Cheng, A.W., Zhang, F., and Jaenisch, R. (2013). One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell *153*, 910-918.

Wang, T., Wei, J.J., Sabatini, D.M., and Lander, E.S. (2014b). Genetic screens in human cells using the CRISPR-Cas9 system. Science *343*, 80-84.

Ward, M.C., Wilson, M.D., Barbosa-Morais, N.L., Schmidt, D., Stark, R., Pan, Q., Schwalie, P.C., Menon, S., Lukk, M., Watt, S*., et al.* (2013). Latent regulatory potential of human-specific repetitive elements. Molecular cell *49*, 262-272.

Watanabe, H., Soejima, K., Yasuda, H., Kawada, I., Nakachi, I., Yoda, S., Naoki, K., and Ishizaka, A. (2008). Deregulation of histone lysine methyltransferases contributes to oncogenic transformation of human bronchoepithelial cells. Cancer cell international *8*, 15.

Westman, B.J., Verheggen, C., Hutten, S., Lam, Y.W., Bertrand, E., and Lamond, A.I. (2010). A proteomic screen for nucleolar SUMO targets shows SUMOylation modulates the function of Nop5/Nop58. Molecular cell *39*, 618-631.

Westphal, T., and Reuter, G. (2002). Recombinogenic effects of suppressors of position-effect variegation in Drosophila. Genetics *160*, 609-621.

White, D.E., Negorev, D., Peng, H., Ivanov, A.V., Maul, G.G., and Rauscher, F.J., 3rd (2006). KAP1, a novel substrate for PIKK family members, colocalizes with numerous damage response factors at DNA lesions. Cancer research *66*, 11594-11599.

Wolf, D., Cammas, F., Losson, R., and Goff, S.P. (2008). Primer binding site-dependent restriction of murine leukemia virus requires HP1 binding by TRIM28. Journal of virology *82*, 4675-4679.

Wolf, D., and Goff, S.P. (2007). TRIM28 mediates primer binding site-targeted silencing of murine leukemia virus in embryonic cells. Cell *131*, 46-57.

Wolf, D., and Goff, S.P. (2009). Embryonic stem cells use ZFP809 to silence retroviral DNAs. Nature *458*, 1201-1204.

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 research *20*, 351-360.

Wu, Y., Ferguson, J.E., 3rd, Wang, H., Kelley, R., Ren, R., McDonough, H., Meeker, J., Charles, P.C., Wang, H., and Patterson, C. (2008). PRDM6 is enriched in vascular precursors during development and inhibits endothelial cell proliferation, survival, and differentiation. Journal of molecular and cellular cardiology *44*, 47-58.

Xue, Y., Gibbons, R., Yan, Z., Yang, D., McDowell, T.L., Sechi, S., Qin, J., Zhou, S., Higgs, D., and Wang, W. (2003). The ATRX syndrome protein forms a chromatin-remodeling complex with Daxx and localizes in promyelocytic leukemia nuclear bodies. Proceedings of the National Academy of Sciences of the United States of America *100*, 10635-10640.

Yamauchi, M., Freitag, B., Khan, C., Berwin, B., and Barklis, E. (1995). Stem cell factor binding to retrovirus primer binding site silencers. Journal of virology *69*, 1142-1149.

Yan, Q., Huang, J., Fan, T., Zhu, H., and Muegge, K. (2003). Lsh, a modulator of CpG methylation, is crucial for normal histone methylation. The EMBO journal *22*, 5154-5162.

Yang, L., Xia, L., Wu, D.Y., Wang, H., Chansky, H.A., Schubach, W.H., Hickstein, D.D., and Zhang, Y. (2002). Molecular cloning of ESET, a novel histone H3-specific methyltransferase that interacts with ERG transcription factor. Oncogene *21*, 148-152.

Yang, X., Khosravi-Far, R., Chang, H.Y., and Baltimore, D. (1997). Daxx, a novel Fas-binding protein that activates JNK and apoptosis. Cell *89*, 1067-1076.

Yoder, J.A., Walsh, C.P., and Bestor, T.H. (1997). Cytosine methylation and the ecology of intragenomic parasites. Trends in genetics : TIG *13*, 335-340.

Yokoe, T., Toiyama, Y., Okugawa, Y., Tanaka, K., Ohi, M., Inoue, Y., Mohri, Y., Miki, C., and Kusunoki, M. (2010). KAP1 is associated with peritoneal carcinomatosis in gastric cancer. Annals of surgical oncology *17*, 821-828.

Young, G.R., Eksmond, U., Salcedo, R., Alexopoulou, L., Stoye, J.P., and Kassiotis, G. (2012). Resurrection of endogenous retroviruses in antibody-deficient mice. Nature *491*, 774-778.

Yuan, P., Han, J., Guo, G., Orlov, Y.L., Huss, M., Loh, Y.H., Yaw, L.P., Robson, P., Lim, B., and Ng, H.H. (2009). Eset partners with Oct4 to restrict extraembryonic trophoblast lineage potential in embryonic stem cells. Genes & development *23*, 2507-2520.

Zeng, L., Yap, K.L., Ivanov, A.V., Wang, X., Mujtaba, S., Plotnikova, O., Rauscher, F.J., 3rd, and Zhou, M.M. (2008). Structural insights into human KAP1 PHD finger-bromodomain and its role in gene silencing. Nature structural & molecular biology *15*, 626-633.

Zhao, L.Y., Liu, J., Sidhu, G.S., Niu, Y., Liu, Y., Wang, R., and Liao, D. (2004). Negative regulation of p53 functions by Daxx and the involvement of MDM2. The Journal of biological chemistry *279*, 50566-50579.

Zhong, S., Salomoni, P., Ronchetti, S., Guo, A., Ruggero, D., and Pandolfi, P.P. (2000). Promyelocytic leukemia protein (PML) and Daxx participate in a novel nuclear pathway for apoptosis. The Journal of experimental medicine *191*, 631-640.

Zhu, Q., Pao, G.M., Huynh, A.M., Suh, H., Tonnu, N., Nederlof, P.M., Gage, F.H., and Verma, I.M. (2011). BRCA1 tumour suppression occurs via heterochromatin-mediated silencing. Nature *477*, 179-184.

Zufferey, R., Nagy, D., Mandel, R.J., Naldini, L., and Trono, D. (1997). Multiply attenuated lentiviral vector achieves efficient gene delivery in vivo. Nature biotechnology *15*, 871-875.