Modulation of an essential histone methyltransferase in mouse embryonic stem cells

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

Gustavo Pereira de Almeida

aus Uberaba, Brasilien

München, 2018

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

Betreuer: Prof. Dr. rer. nat. Gunnar Schotta

Zweitgutachter: Prof. Dr. Andreas Ladurner

Dekan: Prof. Dr. med. dent. Reinhard Hickel

Tag der mündlichen Prüfung: 25.07.2018

Eidesstattliche Versicherung

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

"Modulation of an essential histone methyltransferase in mouse embryonic stem cells"

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.

München, 31.07.2018

_____

Ort, Datum

Unterschrift Gustavo Pereira de Almeida

The work presented in this thesis is being assembled into a manuscript for publication in a peer-reviewed journal. While carrying out my PhD thesis I collaborated with colleagues to support other scientific projects, which led to the following publications:

Dambacher, S., **de Almeida, G.P.**, and Schotta, G. (2013). Dynamic changes of the epigenetic landscape during cellular differentiation. Epigenomics 5, 701-713.

Pasquarella, A., Ebert, A.*, **Pereira de Almeida, G.***, Hinterberger, M., Kazerani, M., et al. (2016). Retrotransposon derepression leads to activation of the unfolded protein response and apoptosis in pro-B cells. Development 143, 1788-1799.

* These authors contributed equally to this work.

I. TABLE OF CONTENTS

II. ABSTRACT

The H3K9me3 is one of the major modifications characteristic of repressed chromatin. Its accumulation is linked to silencing of transcription and compaction of chromatin. Among the methyltransferases able to catalyze H3K9me3, SETDB1 leads to the earliest lethal phenotype in knockout mice embryos. *In vitro* studies showed that its interactor ATF7IP is capable of modulating SETDB1-dependent catalysis of H3K9me3. The implications of this interaction in a cellular system still remains an open question. Thus, to understand the modulation of SETDB1 methyltransferase activity during mouse early development was the main goal of this work. Using ChIP-Seq the genome-wide occupancy of both SETDB1 and ATF7IP was identified in FLAG knockin mouse embryonic stem cells (mESC). To investigate the epigenetic outcome of this interaction, the genome-wide enrichment for different H3K9 marks was characterized in control mESC and cells where *Atf7ip* was deleted by the CRISPR/Cas system. Afterwards, these data were coupled with transcriptome profiles to address whether the epigenetic changes implicated in transcriptional deregulation. In contrast to *Setdb1* knockout mESC, cells lacking ATF7IP survive and grow normally. However, several families of endogenous retrovirus (ERV) belonging to classes I and II known to be controlled by SETDB1 were bound by both SETDB1 and ATF7IP and became derepressed in *Atf7ip* knockout mESC. This phenotype is further enhanced when cells are devoid of DNA methylation. Interestingly, while mutant cells are able to differentiate and repress repetitive sequences during differentiation as in control cells, DNA methylation-depleted mutant mESC could not properly repress those sequences. Unexpectedly, H3K9me3 levels in *Atf7ip* knockout mESC were reduced neither globally nor site specifically at transposable elements targets of SETDB1-ATF7IP complex. On the contrary, H3K9me3 deposition increased at those regions and was followed by increase in H3K9me2. To identify other proteins that might be involved in SETDB1-ATF7IP silencing mechanism mass spectrometry of FLAG-ATF7IP immune-complexes was performed. Then, taking advantage of an exogenous retrotransposon repression reporter system, which is impaired in *Atf7ip*-depleted mESC, ATF7IP partner proteins and other known repressors were screened for genetic interaction with *Atf7ip*. Some factors were found to act within the same pathway, while others had synergistic effects and probably belong to independent pathways. Lastly, exchanging the expression of the endogenous locus by that of different *Atf7ip* mutants by using a Bxb1-mediated recombination system demonstrated that nuclear localization, as well as the conserved Domains 1 and 2, are essential for proper ERV repression. Altogether, this work provides a better understanding of the molecular mechanism of repression by SETDB1 and its modulation by the co-factor ATF7IP.

III. ZUSAMENFASSUNG

Die Histonmodifikation H3K9me3 ist eine der charakteristischsten für reprimiertes Chromatin. Ihre Ansammlung ist mit Stilllegung der Transkription und Verdichtung von Chromatin verbunden. Unter den Methyltransferasen, die in der Lage sind H3K9me3 zu katalysieren, führt SETDB1 zu dem frühesten letalen Phänotyp bei Knockout-Maus-Embryonen. In-vitro-Studien haben gezeigt, dass sein Interaktionspartner ATF7IP die SETDB1-abhängige Katalyse von H3K9me3 modulieren kann. Die Auswirkungen dieser Interaktion in einem zellulären System stellen nach wie vor eine offene Frage dar. Daher war das Hauptziel dieser Arbeit die Modulation der SETDB1-Methyltransferase-Aktivität während der frühen Mausentwicklung zu verstehen. Mittels ChIP-Seq wurden die genomweiten Bindungsorte, sowohl von SETDB1 als auch von ATF7IP, in FLAG-Knockin-Mausembryonalen Stammzellen (mESC) identifiziert. Um die epigenetischen Folgen dieser Interaktion zu untersuchen, wurde die genomweite Anreicherung für verschiedene H3K9-Modifikationen in Kontroll-mESC und Zellen, in denen *Atf7ip* durch das CRISPR/Cas-System deletiert wurde, charakterisiert. Anschließend wurden diese Daten mit Transkriptomprofilen gekoppelt, um zu untersuchen ob die epigenetischen Veränderungen bei der transkriptionellen Deregulierung beteiligt sind. *Atf7ip*-Knockout-mESC überleben und wachsen normal. Jedoch wurden mehrere Familien der Endogenen Retroviren (ERV), die zu den Klassen I und II gehören und sowohl von SETDB1 als auch von ATF7IP gebunden werden, in *Atf7ip*-Knockout-mESC dereprimiert. Dieser Phänotyp wird weiter verstärkt, wenn die Zellen keine DNA-Methylierung aufweisen. Während mutierte Zellen, sowie Kontrollzellen, in der Lage sind zu differenzieren und repetitive Sequenzen während der Differenzierung zu unterdrücken, konnten interessanterweise die DNA-Methylierungs-reduzierte Mutanten diese Sequenzen nicht richtig unterdrücken. Unerwarteter Weise wurden die H3K9me3-Spiegel in *Atf7ip*-Knockout-mESC weder global noch ortspezifisch an Transposons, die Ziele des SETDB1-ATF7IP-Komplexes sind, reduziert. Im Gegenteil kam es vielmehr zu einer H3K9me3 Anreicherung an diesen Regionen die von einer Zunahme von H3K9me2 gefolgt war. Um andere Proteine zu identifizieren, die an dem SETDB1-ATF7IP-Silencing-Mechanismus beteiligt sein könnten, wurde Massenspektrometrie von FLAG-ATF7IP-Immunkomplexen durchgeführt. Anschließend wurden ATF7IP- Interaktionspartner und andere bekannte Repressoren unter Verwendung eines exogenen Retrotransposon-Repressions-Reportersystems, das in *Atf7ip*-depletierten mESC beeinträchtigt ist, auf eine genetische Interaktion mit *Atf7ip* überprüft. Es wurde festgestellt, dass einige Faktoren auf demselben Weg agieren, während andere synergistische Wirkungen haben und wahrscheinlich zu unabhängigen Signalwegen gehören. Schließlich zeigte der Austausch der Expression des endogenen Locus durch die von verschiedenen *Atf7ip*-Mutanten unter Verwendung eines Bxb1-vermittelten Rekombinationssystems, dass sowohl die Kernlokalisierung als auch die konservierten Domänen 1 und 2 für eine korrekte TE-Repression essentiell sind. Insgesamt liefert diese Arbeit ein besseres Verständnis für den molekularen Mechanismus der Repression durch SETDB1 und dessen Modulation durch den Cofaktor ATF7IP.

1. INTRODUCTION

1.1. From naked DNA to highly organized chromatin

Eukaryotic organisms organize the DNA in their nucleus in structures of different orders of magnitude to form the chromatin template (Flemming, 1882). The nucleosome is the most basic structural unit of the chromatin (Kornberg, 1974), which is constituted of a 147 bp long-DNA wrapped around a histone octamer. Two molecules of each of the highly conserved canonical histones H2A, H2B, H3 and H4 form the octamer. The arrays of nucleosomes form a more compact structure following the recruitment of linker histone H1. Association of this compacted array of nucleosomes to the nuclear periphery or other nuclear regions results in a larger domain organization, characteristic of interphase and metaphase chromatin. During nuclear division, further compaction occurs to form the chromosomes (Luger et al., 1997).

The histone tails are targets of a myriad of post-translational modifications (PTM), responsible for changes in the charges (Dou and Gorovsky, 2000) and packaging to form the higher-order chromatin structure (Wei et al., 1999). These modifications of histone residues, together with DNA methylation (DNAme) and nucleosome structural components, are subjected to changes in response to intrinsic and extrinsic stimuli (Jaenisch and Bird, 2003). In turn, the transcriptional activity of the underlying DNA also changes accordingly (Sims et al., 2004). These signatures at the genome-wide level constitute the so-called epigenetic states and are responsible for defining the global transcriptional program of a cell (Henikoff and Greally, 2016). In this way, multicellular organisms can be composed of many distinct cell types differing in which regions of the genome are active or repressed, even though they carry the same genetic material (Strahl and Allis, 2000; Turner, 2000; Dambacher et al., 2013). Basically, active chromatin exists in a decompacted form, prone for transcriptional activity due to higher permissibility to the transcriptional machinery and is termed euchromatin. On the other hand, the chromatin in a silent state, which adopts a compacted conformation to restrict the underlying information, is termed heterochromatin.

1.2. Features of the heterochromatic DNA

The classical definition of the term heterochromatin is based on the cytological observation of nuclear structures densely stained due to higher DNA compaction (Heitz, 1928). This repressed chromatin state is observed in pericentric regions of chromosomes as well as in telomeres. Further molecular and biochemical characterization led to an expansion of the definition. Of note, despite the broad conservation, not all the defined characteristics are ubiquitous of heterochromatin, especially when considering intermediate states. Probing the heterochromatin by nuclease digestion revealed an organization in large arrays regularly spaced and devoid of hyper-sensitive sites characteristic of nucleosome-free regions. The underlying DNA sequence is mainly represented by repetitive sequences (satellite DNA and transposons), with lower density

of genes. Besides, replication occurs during late S phase and a low frequency of meiotic recombination is observed (Richards and Elgin, 2002).

A covalent modification of the DNA in the form of cytosine methylation is the prevalent nucleotide modification in eukaryotes and consolidates pericentric heterochromatin (Bachman et al., 2001). All sites of modification in the four core histones are hypoacetylated in telomeric regions and heterochromatic loci in yeast (Suka et al., 2001). The deposition of distinct non-histone proteins like the heterochromatin protein 1 (HP1) is also detected as a major component in telomeres and pericentric heterochromatin (James et al., 1989). Another hallmark of heterochromatin is the methylation of amino acid residues in the tail of histones H3 and H4. The methylation of histone H3 at the lysine 9 (H3K9me) is evolutionarily conserved throughout most eukaryotes. The first enzyme described to catalyze this mark was first identified in *Drosophila*, encoded by the suppressor of position-effect variegation gene *Su(var)3-9* (Tschiersch et al., 1994). This protein is part of a large family of histone lysine methyltransferases (HKMT), whose catalytic activity is exerted by the SET domain (Jenuwein et al., 1998). Orthologs of this enzyme were found from fission yeast (Nakayama et al., 2001) to humans (Aagaard et al., 1999).

Euchromatic regions can undergo heterochromatinization to form facultative heterochromatin. This subtype occurs in a cell type-dependent manner and just a fraction of the cells will display such compaction (Mozzetta et al., 2015). It can also be stablished in only one of the homolog chromosomes (Richards and Elgin, 2002), as seen for the female X chromosome inactivation in mammals (Lyon, 1961). Interestingly, the repressed state is maintained by the daughter cells after mitosis. The occurrence of H3K9me in such euchromatic regions is dependent on enzymes other than the SU(VAR)3-9, but belonging to the same family (Huisinga et al., 2006).

1.3. The SUV39 family of histone lysine methyltransferases

All members of the SUV39 family of methyltransferases are able to methylate their lysine substrate to one or more of the three states (mono-, di- or trimethylation) in the presence of the cofactor S-adenosyl-L-methyonine (SAM). The SET domain association to two cysteine-rich domains (pre-SET and post-SET), essential for the catalytic activity *in vitro* (Schultz et al., 2002), distinguishes this group from other SET domain-containing proteins (Rea et al., 2000). SUV39H1 was the first of six members to be characterized in mammalian (Mozzetta et al., 2015). SUV39H2 is a second homolog of the fruit fly Su(var)3-9 with 59% identity to SUV39H1 (O'Carroll et al., 2000). SETDB1 and SETDB2, whose SET domain is bifurcated by an insertion of a few hundred amino acids (Schultz et al., 2002; Falandry et al., 2010), are also part of this group. The last two members, G9A and GLP, are not functional alone *in vivo* (Shinkai and Tachibana, 2011) and cannot catalyze trimethylation in this system. Only a few members of the PR-domain containing (PRDM) family share the H3K9 methyltransferase (H3K9MT) ability (Pinheiro et al., 2012). Interestingly, non-histone substrates were also described for many H3K9MT (Herz et al., 2013).

In general, SET containing-HKMT present an additional domain capable of recognizing PTM, specially chromatin marks. This property ensures the HKMT act in the right context and suggests mechanisms of protein crosstalk for establishment and dispersion of histone marks to build up the mature chromatin. SUV39H1/2 target methylated lysine through their chromatin-organization modifier domain (chromodomain), which is essential for the specific binding to pericentric heterochromatin (Melcher et al., 2000). SETDB1/2 contain a canonical methyl-CpG binding domain (MBD), which might selectively bind methylated DNA (Kang, 2015). G9A and GLP ankyrin repeats bind H3K9me1/2 marks and are essential for the H3 tail interaction (Collins et al., 2008).

All SUV39 family members are able to catalyze the three states of H3K9me *in vitro* (Mozzetta et al., 2015), with the exception of SETDB2, which can only trimethylate (Falandry et al., 2010). Thus, the redundancy of these enzymes impedes the assignment of specific functions *in vivo*. However, the bulk H3K9me3 is attributed to SUV39H1/2 enzymes (Peters et al., 2003) and their activity is essential for spreading this mark at pericentric heterochromatin (Peters et al., 2001). At telomeres, these enzymes mark chromatin with both H3K9me2/3 (Garcia-Cao et al., 2004). Besides that, SUV39H-dependent H3K9me during S-phase of differentiating cells is linked to transcriptional silencing (Ait-Si-Ali et al., 2004). SETDB2 is recruited to heterochromatic regions *in vivo* for trimethylation of centromere-associated repeats (Falandry et al., 2010). G9A and GLP are implicated in regulation of pericentric heterochomatin (Dong et al., 2008; Fritsch et al., 2010). On the other hand, they bind euchromatic regions as well and act in concert to establish global H3K9me1/2 (Tachibana et al., 2005). SETDB1 is also shown to act over both heterochromatin (Loyola et al., 2009) and euchromatin (Schultz et al., 2002) and will be further discussed below.

1.4. The bifurcated SET domain protein SETDB1

1.4.a. Structural features

SETDB1 coding sequence was first predicted in an analysis of cDNA clones from a human cell line (Nomura et al., 1994). Later on, in a search for SET domain proteins, *SETDB1* was found containing a conserved motif at its C-terminus. However, its SET domain is peculiarly interrupted by an insertion of 347 amino acids in a less conserved region in the middle of two highly conserved ones (Harte et al., 1999). In mouse, SETDB1 was first identified in a screening for interaction partners of ERG, a transcription factor related to cell growth and differentiation. It was termed ERG-associated protein with SET domain (ESET) and revealed an identity of 92 % to the human homolog (Yang et al., 2002). The fruit fly homolog, known as *Egg* or *Eggless*, also contains an insertion in the bifurcated SET domain and shows 63 %, 44 % and 39 % identity to the SET, tudor and MBD domains of the mouse version, respectively (Stabell et al., 2006). An homolog is also well characterized in *C. elegans*, where it is called *met-2* (Poulin et al., 2005) and shows 19 % identity to hSETDB1 overall and a 50 % identity at the SET domain (Andersen and Horvitz, 2007).

Although the calculated molecular weight for SETDB1 is 145 kDa, it is resolved on SDS-PAGE at approximately 180 kDa due to PTM (Yang et al., 2002). The presence of a second band with higher molecular weight is due to ubiquitination of the SET domain, at K867 in humans (Ishimoto et al., 2016). SETDB1 contains a MBD (Yang et al., 2002), characteristic of some proteins related to silencing of methylated DNA (Bird and Wolffe, 1999). Recombinant hSETDB1 is not able to bind methylated DNA by itself and mutations at the MBD did not impair hSETDB1 localization to an endogenous locus (Matsumura et al., 2015). In *Drosophila*, dSETDB1 preferentially binds DNA bearing one or multiple methylated CpA motifs *in vitro*, like *Rb* gene and the Rt1b retrotransposon (Gou et al., 2010). SETDB1 bears also a tudor domain (Yang et al., 2002), which is known to be involved in protein-protein interaction (Ponting, 1997). At the N-terminal region a functional SUMO interacting motif (SIM) is present (Ivanov et al., 2007; Tanaka and Saitoh, 2010; Cho et al., 2013) nearby two nuclear export (NES) and two nuclear localization (NLS) signals (Cho et al., 2013). Interestingly, the N-terminus of dSETDB1 is SUMOylated in S2 cells and colocalizes to SUMO in oocytes (Koch et al., 2009). SUMOylation has been observed in mammals too (Yeap et al., 2009).
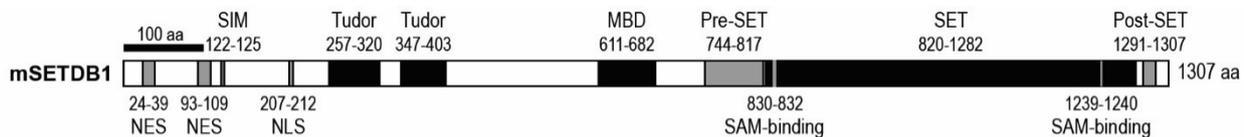
Figure 1.1 | Schematic representation of the domain structure of mouse SETDB1.
SETDB1 bears two nuclear export signals (NES) and one nuclear localization signal (NLS) at the N-terminal region. Between them there is a SUMO interaction motif (SIM). Two tudor domains are involved in protein interaction. At the center of the protein there is a putative methyl-DNA binding domain (MBD). The Pre-SET, SET and Post-SET domains are involved in the methyltransferase activity. The residues responsible for SAM binding are indicated.

1.4.b. Expression and localization patterns

Setdb1 promoter in mice has an Ets (avian erythroblastosis virus oncogene-E twenty-six) binding site and its activity is under control of the transcription factor Ets-2 (Lee et al., 2008). The full length transcript is expressed in several mouse cell lines and tissues (liver, brain, thymus, heart, lung, spleen, testis, ovary, kidney, liver, skeletal muscle). A shorter splicing variant lacking the SET domain and MBD is expressed in many immortalized cell lines, though it only occurs in brain, testis and ovary (Yang et al., 2002; Blackburn et al., 2003).

SETDB1, but not G9a or SUV39H1, is detected in vesicle-stage oocytes. After fecundation, at pronucleus stage zygotes, SETDB1 accumulates around the satellite DNA-rich nucleolus. It is very abundant as small foci in the male pronucleus, which is devoid of H3K9me2/3, differently from the female (Cho et al., 2012). During embryogenesis, expression in the zygote only begins at blastocyst stage and there is maternal contribution during the preimplantation period (Dodge et al., 2004). The maternal stock decreases from 2-cell stage to complete absence in morula, although the protein level is normal until 4-cell stage, with a diffuse nuclear pattern. During 8-cell and morula stages, the levels stay low only to reappear in the blastula as several foci in both inner

cell mass (ICM) and trophectoderm cells (Cho et al., 2011; Cho et al., 2012). In blastocyst outgrowths, SETDB1 was present only in Oct4-expressing cells derived from the ICM in several foci similarly to Pml (Cho et al., 2012). Expression is ubiquitous from embryonic day E7.5 to E9.5 (Dodge et al., 2004). In NIH/3T3 murine embryonic fibroblasts (Schultz et al., 2002) and mouse embryonic stem cells (mESC), SETDB1 is predominant in euchromatin (Kourmouli et al., 2005). In HeLa, HEK293, hepatoblastoma, functional liver FLC4, adenocarcinoma and lipossarcoma cells it localizes mainly at the cytoplasm (Tachibana et al., 2015).

In *Drosophila*, dSETDB1 transcripts are not detectable in the first three hours of embryonic development and no maternal contribution is observed (Stabell et al., 2006; Seum et al., 2007; Tzeng et al., 2007). However, it is present at lower levels in larvae and pupae and regularly expressed during other developmental stages (Stabell et al., 2006). Besides expression in the ovary, females showed slightly higher expression than males in adult tissues (Yoon et al., 2008). dSETDB1 localizes to euchromatin, heterochromatic regions and chromocenter, mainly at chromosome 4 (Stabell et al., 2006; Seum et al., 2007). Pericentric localization was also observed in the germarium (Yoon et al., 2008).

1.4.c. Histone lysine methyltransferase activity

When probed for methyltransferase activity, hSETDB1 showed specificity over H3 substrates and did not methylate H2A, H2B and H4. This enzymatic activity was attributed to hSETDB1 itself, as mutations abolished H3 methylation. PTM are required to properly exert its enzymatic activity, as recombinant hSETDB1 (Yang et al., 2002) or dSETDB1 (Stabell et al., 2006) purified from *E. coli* could not methylate core histones. The *in vitro* activity depends on the pre-SET, SET and post-SET domains in human (Schultz et al., 2002) and *Drosophila* cells (Tzeng et al., 2007), not being affected by mutations on MBD or tudor domains. However, complete removal of the MBD abrogates the catalytic ability. SET domain ubiquitination requires the MBD and increases hSETDB1 methyltransferase activity (Ishimoto et al., 2016). This ubiquitination by UBE2E1/2/3 is protected by multiple motifs in hSETDB1 which interact with ubiquitin (Sun and Fang, 2016).

The specificity of SETDB1 for the lysine 9 was demonstrated by using mutated recombinant histone tails. No additional PTM of the substrate is needed, although demethylation of H3K9 and acetylation of different residues inhibits activity (Schultz et al., 2002). H3K27me3 also blocks deposition of H3K9me3, but not of H3K9me1/2 (Fei et al., 2015b). Binding of the C-terminus of hSETDB1 expressed in insect cells to recombined histone H3 is impaired by H3K4me3 and blocks catalysis (Binda et al., 2010). The activity of recombinant hSETDB1 over core histones is lower than the native complex isolated from HeLa cells. Free core histones are a preferred substrate, compared to mono- or oligonucleosomes (Wang et al., 2003). In contrast, dSETDB1 trimethylates preferentially nucleosomal H3 (Gou et al., 2010). All three states of H3K9me are sequentially

catalyzed by dSETDB1 purified from SL2 cells (Tzeng et al., 2007). *In vivo*, dSETDB1 deposits H3K9me1/2 at euchromatin, but not at chromocenter or telomeres (Seum et al., 2007).

Ribosome complexes translating histone H3 in HeLa cells showed H3K9me1/2 deposition while H3 is still associated with ribosomes. Purified ribosome complexes showed H3K9me1/2 methyltransferase activity and H3K9me1 was detected in nascent polypeptides. This activity is dependent on hSETDB1 and its association with ribosome subunits RPL5 and RPS3a. hSETDB1 dissociates from the substrate before the next catalysis event (Rivera et al., 2015). Interestingly, presence of SUMOylated hKAP1 stimulates baculovirus-expressed hSETDB1 activity towards histone H3 in a dose-dependent-manner (Ivanov et al., 2007). Non-histone proteins can also be used as substrate (Binda et al., 2010; Hwang et al., 2014).

Met-2, the *C. elegans* homolog, cooperates with another H3K9me, *set-25*, to keep chromatin at the periphery of the nucleus. *Met-2* mono- and dimethylates H3K9, while *set-25* trimethylates it. H3K9me3 is not needed for perinuclear localization of chromatin, but is required for silencing. *Met-2* is very abundant in the cytoplasm, while *set-25* is found in the nucleus at perinuclear foci in a H3K9me3-dependent manner (Towbin et al., 2012). In summary, the primary structure of SETDB1 and its expression patterns have been well characterized so far, however, not so much is known about the quaternary structure and about the mechanisms of SETDB1 HKMT activity, especially in an in vivo system.

1.5. SETDB1 roles in development and differentiation

1.5.a. Embryonic stages

Deletion of *Setdb1* leads to peri-implantation lethality between 3.5 and 5.5 (Dodge et al., 2004; Keniry et al., 2016). Null blastocysts do not show global changes in staining for H3K9me2/3. Besides that, no mESC lines can be derived by blastocyst outgrowth, consistent with the impaired HKMT-dependent proliferation and mESC death following depletion (Yuan et al., 2009; Lohmann et al., 2010). *Setdb1*-depleted mESC show no changes in H3K9me2 level and mild or no reduction of global H3K9me3 level (Matsui et al., 2010; Thompson et al., 2015), although strong H3K9me3 loss was reported previously (Yeap et al., 2009). Depletion of hSETDB1 also impairs proliferation in HeLa and HEK293T cells (Wang et al., 2003). Interestingly, SETDB1, HP1γ, G9a and GLP constitute a block for reprogramming and their depletion enhances the reprograming rates, with a more pronounced effect for late steps of reprogramming (Sridharan et al., 2013).

SETDB1 is essential for keeping mESC pluripotent state, the ability to form colonies and capacity to differentiate into embryoid bodies (Bilodeau et al., 2009; Yuan et al., 2009; Lohmann et al., 2010). It binds developmental genes, which may contain SETDB1-depedent H3K9me3. Depletion leads to differentiation mainly into trophectoderm lineage cells, following the repression of pluripotency genes and downregulation of differentiation markers (Bilodeau et al., 2009; Yuan et al., 2009). When modified by SUMO1 (Yeap et al., 2009), SETDB1 is recruited by the stem

cell-specific transcription factor Oct4 to the promoter of the trophectoderm determinant genes and induces H3K9me2/3 deposition to repress transcription (Yuan et al., 2009; Lohmann et al., 2010). Curiously, most of these genes are also enriched for H3K4me3 and some of them are co-occupied by the Polycomb group (PcG) complex subunit Suz12 and H3K27me2/3 (Bilodeau et al., 2009; Lohmann et al., 2010), but not by DNAme (Karimi et al., 2011). Interestingly, SETDB1 interacts with the PcG member RNF2 in murine erythroleukemia cells (Sanchez et al., 2007). Furthermore, SETDB1-depleted two-cell stage embryos upregulates trophectoderm genes and incorporates preferentially into the trophectoderm rather than the blastocysts ICM (Yuan et al., 2009; Lohmann et al., 2010). SETDB1 also binds a set of germline genes which contains DNAme and loses H3K9me3 in knockout mESC to become upregulated in this mutants and in DNMT triple knockout (TKO) mESC (Karimi et al., 2011). Intriguingly, several SETDB1 binding sites have no H3K9me3 10 kb around the peak center, while fewer had H3K9me1/2/3 (Fei et al., 2015b).

Compared to the male X chromosome in mouse embryonic fibroblasts (MEF), the female shows dense regions of H3K9me2/3 and SETDB1 occupancy occurring at low gene density areas, while H3K27me3 is enriched at gene-rich areas. Knockout female MEF and embryos cannot perform proper X chromosome inactivation (XCI) of an X-linked reporter. H3K9me2/3 and H3K27me3 are reduced in all chromosomes and along the X chromosome, which also showed small reduction of DNAme at CpG islands. However, only minimal reactivation of endogenous X-linked genes occurred. To circumvent the redundancy of XCI maintenance, *Setdb1* was depleted in differentiating mESC and impaired silencing of X-linked and autosomal genes, but not *Xist* expression or coating of the X inactive chromosome. Genes which undergo silencing during mESC differentiation were impaired after *Setdb1* depletion, whereas genes which maintained silenced state throughout differentiation showed no reactivation. Thus, Setdb1 is involved in establishment and in early stages of maintenance of silencing (Keniry et al., 2016).

dSETDB1 depletion leads to low viability and early lethality. The progeny shows developmental arrest just before the second larval instar (Stabell et al., 2006) and death at late pupal stage (Seum et al., 2007). Mono-, di- and trimethylation of H3K9 are globally reduced in third instar larvae (Tzeng et al., 2007; Riddle et al., 2012), HP1α is lost at chromosome 4 and genes are derepressed (Riddle et al., 2012). However, A/T motif-rich promoters also bound by HP2 at this chromosome keep a localized H3K9me-independent HP1α binding. The spread of HP1α to the gene bodies, though, happens in a dSETDB1 methylation-dependent manner and in the presence of the painting of fourth protein POF (Figueiredo et al., 2012), recruited by dSETDB1 (Riddle et al., 2012). In first instar larvae, upregulated genes in chromosome 4 overlapped extensively in *dSetdb1*, HP1α and *dSu(var)3-9* mutants (Lundberg et al., 2013). Pupae fail to recruit HP1 and to repress transgenes in the vicinity of chromosome 4 heterochomatic domains (Seum et al., 2007) or in pericentric heterochromatin (Brower-Toland et al., 2009), indicating its requirement for

repression of variegation. Flies reaching adulthood exhibit spread-out wings (Tzeng et al., 2007) and shorter life span, which are more pronounced in males (Brower-Toland et al., 2009).

1.5.b. Germline cells and reproduction

In mice testis, SETDB1 and H3K9me3 levels increase during development. *Setdb1*-depleted spermatogonial stem cells (SSC) show reduced levels of H3K9me3, lower viability due to apoptosis and does not localize properly to the seminiferous tubules to undergo spermatogenesis. The testes showed much lower weight and size. In the spermatogonial stem cell line C18-4, the promoter of the cytochrome oxidase *Cox4i2*, which marks the early onset of apoptotic events, shows enrichment for SETDB1-dependent H3K9me3 and DNAme. Knockdown of *Cox4i2* in *Setdb1* depleted C18-4 cells partially rescues from apoptosis (An et al., 2014). In male E13.5 primordial germ cells (PCG), *Setdb1* deletion driven by *Tnap* promoter, leads to partial loss of H3K9me3 and H3K27me3 and reduction in the number of germ cells in the gonads. The global DNAme level increases at H3K9me3-enriched regions and at gene bodies, though. Impaired gametogenesis is observed in post-natal and adult mutant mice (Liu et al., 2014).

Maternal contribution of SETDB1 is essential for early mouse embryonic development (Eymery et al., 2016) and mouse oocytes show strong staining for SETDB1 (Kim et al., 2016). Deletion in mouse oocytes, driven by *Zp3-Cre*, results in embryo degeneration before morula stage due to defects in the first mitotic cycle, even if it is wild type for zygotic expression (Eymery et al., 2016; Kim et al., 2016). Mutant females show normal folliculogenesis, besides reduced H3K9me2 (Eymery et al., 2016) and mild decrease in H3K9me3, but not in H3K9me1. However, they are infertile and fewer oocytes reach later stages, showing more γ-H2AX foci, characteristic of DNA double-strand breaks. The oocyte requires proper *Setdb1* expression and activity to resume meiosis (Kim et al., 2016), for instance, for meiotic maturation, bipolar spindle formation, stable kinetochore attachment to microtubule and chromatid segregation. Dysregulated genes in *Setdb1* knockout oocytes were enriched for cell cycle, cell division and chromosome organization genes (Eymery et al., 2016). The meiosis-related gene *Cdc14b* is repressed by SETDB1-dependent H3K9me3 in mESC and shows higher expression in mutant oocytes. Depletion of *Cdc14b* in *Setdb1* mutant oocytes ameliorated the meiotic defects (Kim et al., 2016).

While *dSetdb1* is not critical for spermatogenesis and spermiogenesis (Ushijima et al., 2012), mutant females are sterile (Seum et al., 2007), with reduced ovary size, early blockage of oogenesis and no formation of egg chambers (Clough et al., 2007; Yoon et al., 2008). Germ cells and somatic cells show impaired proliferation, strong H3K9me3 reduction and apoptosis (Clough et al., 2007). Maintenance of germline stem cells, their differentiation and survival of escort cells also depends on dSETDB1 (Wang et al., 2011). In mutant adult flies, total levels of the three H3K9 methylation states are reduced (Brower-Toland et al., 2009). In *C. elegans*, the *Setdb1* homolog (*met-2*) is involved in multivulval phenotype (Poulin et al., 2005) when deleted in combination with other multivulval class A genes, like the homologs of *Setd2*, HP1 (Andersen and Horvitz, 2007)

and MBT (Koester-Eiserfunke and Fischle, 2011). These genes work redundantly to repress the vulval cell fate transcription factor *lin-3*. Mutant embryos show decreased levels of H3K9me3 and H3K36me3 (Andersen and Horvitz, 2007), whereas mutant adults have gonadal defects, loss of H3K9me2 in germ cells (Bessler et al., 2010) and sterility (Koester-Eiserfunke and Fischle, 2011). Co-deletion of the H3K4me2 demethylase *spr-5* (Kerr et al., 2014), the *Chd* homolog let-418 or the *Zfp* lin-13 enhances sterility (McMurchy et al., 2017). Interestingly, while the X chromosome is enriched for H3K9me2, in *C. briggsae* it is covered with H3K9me3 (Larson et al., 2016). Co-deletion of another H3K9me *set-25* leads to complete loss of H3K9me and sterility in adults (Zeller et al., 2016). Functions of MET-2 were also shown in terminal differentiation of neurons (Zheng et al., 2013), small-RNA-guided H3K9me3 (Mao et al., 2015; Kalinava et al., 2017) and mitochondrial stress response (Tian et al., 2016).

1.5.c. Neuronal development

Neural progenitor cells (NPC) express SETDB1 at E9.5, but transcription reduces over time until it reaches very low levels at E17.5. SETDB1 is required for appropriate expression of neuronal and non-neuronal genes, like the gliogenesis regulator *Sox9* and the astrocyte marker *Gfap*. Thus, knockout in E11.5 mouse brain impairs neurogenesis and reduces formation of deep layer neurons with increase in apoptotic cells and global H3K9me3 reduction at E14.5 and E18.5. NPC proliferation is compromised and astrocyte formation enhanced. These mice do not survive more than 10 days after birth (Tan et al., 2012). In fly proneural cluster, the phenotype of *dSetdb1* loss is further enhanced by co-deletion of caspase, due to its non-apoptotic functions (Shinoda et al., 2016). In mESC, almost all SETDB1 binding sites at neuronal development regulators and pluripotent state regulators are devoid of H3K9me marks. Several neuronal-related transcription factors bound by SETDB1 are co-occupied by subunits of the PRC2 complex, their interactors and H3K27me3. Loss of *Setdb1* reduces the levels of EZH2 and H3K27me3 and leads to upregulation, while EZH2 knockdown resulted in deposition of H3K9me3. Thus, enhanced neuronal differentiation of mESC is observed in the absence of *Setdb1* (Fei et al., 2015b).

Setdb1 deletion in postnatal forebrain neurons is not lethal, though. Adult mutant mice brains are smaller, even though no cell death or neuronal loss occurs. The general spatial organization of the genome is unchanged as seen by Hi-C chromosome conformation capture. Several long-range loop contacts are lost in mutant neurons. Many topologically associated domains (TAD) were lost at a cluster containing more than 70 genes, most of which encode adhesion molecules grouped in three clusters *Pcdha/b/g* regulating neuronal connectivity. Several regions losing H3K9me3, including this cluster, have CTCF motifs, similarly to mESC and B lymphocytes. Many CTCF binding regions show increased binding in mutant, especially in regions losing H3K9me3. Thus, SETDB1 prevents excess CTCF binding. Most of the new CTCF binding sites close to TAD boundary and to H3K9me3-depleted regions show increased insulation, besides the *Pcdh* cluster. There, insulation was completely lost due to structural collapse, with accumulation of histone

hyperacetylation and loss of DNAme. In contrast, DNAme levels are normally already very low and CTCF levels high at two related enhancer elements. Many of the upregulated genes in mutant neurons were located at the *Pcdh* cluster and acquired histone acetylation with concomitant increase in transcription. This cluster does not show similar effects in prenatal deficiency of *Setdb1*. The enhancer elements form shorter-range promoter contacts and are contained in a H3K9me3-marked chromatin in normal neurons, inside repressive loops involving multiple KRAB-ZFP, like ZFP143 (Jiang et al., 2017).

1.5.d. Hematopoiesis

Induction of *Setdb1* deletion in hematopoietic stem and progenitor cells (HSPC) via tamoxifen-responsive *Rosa:Cre-ERT* driver leads to host mice death 3 weeks after transplantation due to strong hematopoietic failure. Levels of bone marrow and primitive hematopoietic cells, including LSK hematopoietic stem cells (HSC), LSK multipotent progenitor cell (MPC), common myeloid progenitors (CMP), granulocyte/macrophage progenitors (GMP), except for erythrocyte progenitors, are strikingly lower. Apoptosis increases in LSK HSC and MPC and cell proliferation is lower. Global H3K9me3 level and DNAme in mutant GMP were unchanged but several non-hematopoietic genes lost H3K9me3 and tended to be derepressed. The gluconeogenesis enzymes *Fbp1/2* were identified as SETDB1 targets which get upregulated in mutant GMP and LSK cells greatly reducing ATP levels and impairing metabolic homeostasis (Koide et al., 2016).

Deletion of *Setdb1* in early stages of B-cell development using the *Mb1-Cre* driver completely suppresses late-stage B-cells at bone marrow and spleen, but not in the thymus. Expression of the anti-apoptotic *Bcl2* rescues pro-B and pre-B cell compartments, but not late stages B-cells. *Setdb1* deletion upregulates several innate immunity or non-hematopoietic lineage genes (Collins et al., 2015; Pasquarella et al., 2016), none of which showing SETDB1 binding. Transplantation assays shows cell intrinsic defects and mutant pro-B cells do not differentiate into B cells *in vitro*. Unfolded protein response (UPR) genes are upregulated in mutant pro-B cells. Knockdown of the pro-apoptotic gene *Bcl2l11*, which regulates apoptosis in the UPR context, could rescue the ability of mutant pro-B cells to form colonies. UPR gets triggered due to high levels of the MLV envelop protein Env detected in mutant pro-B cells. Indeed, ectopic overexpression of the Env protein in B cells led to apoptosis and upregulation of UPR pathway (Pasquarella et al., 2016).

Thymocyte-specific deletion with *Lck-Cre* driver reduces cellularity in the thymus, lymph nodes and spleen, including late-stage thymocytes (Martin et al., 2015; Takikita et al., 2016). Mutant thymocytes are more prone to undergo cell death due to TCR agonism. The IgG inhibitory receptor gene *Fcgr2b*, involved in T-cell development, loses SETDB1-dependent H3K9me3 in mutants leading to exacerbated signaling through the TCR complex and apoptosis (Martin et al., 2015). It was also suggested that activation of ERK by phosphorylation is decreased by FcγRIIb in mutant thymocytes and this disturbs positive selection (Takikita et al., 2016).

Lipid A stimulation after *LysM-Cre* induction of *Setdb1* loss in macrophage further upregulates genes compared to wild type. These genes are enriched for interleukin and chemokine pathways, for instance IL6, which lost H3K9me3. When used as reporter, the IL6 promoter is suppressed in a HKMT-dependent manner and silencing requires its NF-κB binding site. Binding of NF-κB p65 to the promoter is further increased in mutant macrophages. The knockout mice are more prone to endotoxin shock after LPS treatment, due to enhanced response. Thus, *Setdb1* is able to suppress *in vivo* inflammatory responses mediated by TLR4 (Hachiya et al., 2016).

1.5.e. Musculoskeletal system

At E14.5, when the growth plate is assembled in the mouse forelimb bone, all H3K9MT are expressed in proliferating chondrocytes. H3K9me1/2/3 are modestly detected in the more differentiated prehypertrophic, hypertrophic chondrocytes and also in trabecular bones at E16.5, when osteogenesis can be observed. H3K9 acetylation (H3K9ac) was evident in prehypertrophic chondrocytes and is lost over time in hypertrophic chondrocytes. Higher expression of H3K9MT is observed in the trabecular bone compared to other bone regions (Ideno et al., 2013). *Setdb1* deletion in mice bone mesenchymal cells driven by *Prx1* leads to improper formation of skeleton in newborn pups. Femoral growth plates are disorganized and chondrocyte hypertrophy is enhanced (Yang et al., 2013). The trabecular network in tibia already shows impairment 7 days after birth, with decrease in proliferating osteoblasts and no formation of epiphyseal plates. Mesenchymal stem cell differentiation into osteoblasts is compromised. Basically, the osteoblast specific transcription factor Runx2 interacts with SETDB1 and HDAC4 to keep proper levels of its target gene osteocalcin (Lawson et al., 2013a; Yang et al., 2013). Regarding the articular cartilage, one month-old mutant mice accumulates proliferating hypertrophic chondrocytes undergoing apoptosis. In three-month old mutant mice, the chondrocytes reach terminal differentiation near the articular surface, where type-II collagen and proteoglycan are lost. In twelve-month old, degeneration of the articular cartilage is enhanced (Lawson et al., 2013b). In Meckel's cartilage, that supports embryonic formation of mandible, *Setdb1* deletion induced by *Wnt1-Cre* reduces craniofacial size. Mutant mice show larger cartilage with mineralization instead of losing it. Hypertrophy and cell proliferation is enhanced in chondrocytes. BMP signaling is strongly active due to high pSMAD1/5/8 level (Yahiro et al., 2017).

In adult mouse skeletal muscle satellite cells, *Setdb1* expression strongly increases after activation from quiescence (Beyer et al., 2016). It is expressed in proliferating mouse skeletal myoblast C2C12 cells, increases during early differentiation and drops afterwards (Song et al., 2015; Beyer et al., 2016). Depletion in proliferating myoblasts impairs proliferation and differentiation. Expression of muscle differentiation genes, like *MyoD* and myogenin are reduced, even though SETDB1 neither interacts with MyoD nor binds its promoter (Song et al., 2015). In contrast, depletion at the onset of terminal differentiation decreases self-renewal, induces late muscle differentiation markers and enhances commitment. Most SETDB1 binding sites have

H3K9me3, including many genes upregulated during myoblast differentiation. The homogenous distribution of SETDB1 is converted to mainly cytoplasmic during differentiation and the enhancer of the myoblast-specific gene *Ankrd1* loses SETDB1 binding and H3K9me3 to get upregulated. Similarly, ectopic expression of *Wnt3a*, which is essential for embryonic myogenesis, decreases SETDB1 binding and redistributes it preferentially to the cytoplasm in myoblasts, specially of a post-translationally modified form. Similar effects were seen in HeLa cells. (Beyer et al., 2016).

During early stages of adipogenesis in 3T3-L1 preadipocytes, most SETDB1 binding sites are intragenic, occur 1 kb downstream from transcription start sites (TSS) and overlap with H3K9me3, MBD1 and lineage-specific DNAme. Several of those sites are developmental genes and also showed enrichment for H3K4me3 at the proximal promoter. In mESC, these genes are enriched for H3K4me3 and H3K27me3 instead. SETDB1 enrichment decreases during preadipocyte differentiation and its depletion in 3T3-L1 preadipocytes and lineage-committed mesenchymal stem cells leads to differentiation. H3K9me3 deposition by SETDB1 prevents the binding of an early adipogenic transcription factor to the adipogenic master regulator *Cebpa*, the deposition of H3K27ac and H3K4me3 at the gene body and keeps Pol II in a paused state. At later stages, the adipocytes still keep the DNAme levels on *Cebpa* gene body. Hence, reduction of SETDB1 recruitment is due to the decrease in expression during differentiation (Matsumura et al., 2015). It is also helped by enhanced recruitment of LSD1, which leads to decrease in H3K9me2 basal levels in favor of H3K4me2 followed by H3K4me3 to activate transcription (Musri et al., 2010).

Lastly, although not directly related to development, hSETDB1 has also a role in DNA damage response, where it depends on the DNA damage sensor ATM, HP1 and hCAF-1 to be recruited to double strand break sites. Depletion of *SETDB1*, *HP1* or *SUV39H1/2* in human immortalized fibroblasts leads to homologous recombination defects during G2 phase and reduces sister chromatin association, but has no effects on non-homologous end-joining. Even though those proteins are not required for resection initiation, they are important for the extension step, for repositioning the damage response mediator protein h53BP1 peripherally (Alagoz et al., 2015).

All in all, SETDB1 is regarded as an important factor for the development of several tissues and for the proper differentiation of many specific cell types. Besides that, this protein is required to keep the pluripotent state in stem cells. Those facts demonstrate its importance as a transcriptional regulator and its essentiality to keep the cell identity. Repression by SETDB1 has been described in different contexts, for instance at repetitive genomic regions and also at euchromatic genes. The next two chapters cover what is known about this silencing by SETDB1.

1.6. Silencing of euchromatic genes

1.6.a. Artificial promoters

hSETDB1 was linked for the first time to transcriptional repression when its interaction with the bromodomain of the KRAB-associated protein hTRIM28 was described. This protein is recruited

to chromatin via Krüppel-associated box domain-containing zinc finger transcription factors (KRAB-ZFP) to regulate transcriptional repression via chromatin compaction. Targeting a KRAB repressor protein to a reporter promoter in NIH/3T3 fibroblasts, hTRIM28 accumulation is concomitant to that of hSETDB1 in cells stably silencing the reporter. H3K9me and HP1α deposition is also observed (Schultz et al., 2002). This response does not spread to a nearby promoter and is strikingly localized to the targeted region, which adopts a compact chromatin structure resistant to nuclease treatment (Ayyanathan et al., 2003). Silencing is followed by spatial relocalization to condensed chromatin islands. A subpopulation of cells kept a stable mitotically heritable repression over several generations. The locus had higher CpG methylation, suggesting a role of DNAme on heritable silencing. Similar picture was seen for the endogenous gene *Col11a2* under the control of a natural KRAB-ZFP.

Targeting SETDB1 to an artificial promoter in NIH/3T3 fibroblasts strongly represses the reporter, independently of the MBD and SET domain. Silencing is impaired by tudor domain deletion and inhibition of HDAC activity, as the tudor domain binds the corepressors mSIN3A/B, forming a larger complex with the histone deacetylases HDAC1/2 (Yang et al., 2003). In S2 cells, dSETDB1 targeting resulted in HKMT-dependent repression and deposition of H3K9m3, with concomitant enrichment for DNMT2 and DNAme (Gou et al., 2010). When targeting hTRIM28 to an artificial promoter in HEK293 cells, recruitment of RNA polymerase II (Pol-II) is reduced together with H3K4me3 enrichment, while H3K9me3, H3K27me3 and H4K20me3 become more abundant. The repression is dependent on the presence of hTRIM28, hSETDB1, HP1α and HP1β to less extent (Sripathy et al., 2006). SUMOylation of hTRIM28 by SUMO1 and an intact SIM is required for hSETDB1 recruitment, but not for HP1α deposition (Ivanov et al., 2007). However, targeting HP1α to pericentric heterochromatin results in SETDB1 recruitment and reestablishment of H3K9m3 in *Suv39h1/2* double-knockout mESC. Targeting of HP1α/β to a synthetic promoter amplified in Chinese hamster ovary cells DG44 also results in chromatin compaction and SETDB1 localization to the targeted HP1 foci, together with H3K9me3. No polycomb group (PcG) proteins co-localized to these foci (Verschure et al., 2005).

hSETDB1 N-terminal region interacts with the PHD-finger like domain of hDNMT3A/B, but not with hDNMT1, in transfected monkey kidney cell line (COS-7). Targeting either DNMT3A or SETDB1 to an artificial promoter when the other is co-expressed leads to stronger repression. H3K9me3 accumulation and low level of H3K9me2 is observed with decrease in H3K9ac, due to HDAC1 binding. The methylated promoter of the tumor suppressing gene *RASSF1A* in MDA-MB-231 cells shows accumulation of SETDB1, DNMT3A, HDAC1 and H3K9me3 (Li et al., 2006).

1.6.b. Endogenous genes

During cell division in HeLa cells, H3 methylation happens before its incorporation into the newly synthesized DNA (Sarraf and Stancheva, 2004). While the methyl-CpG binding protein

hMBD1 binds hSETDB1 throughout the cell cycle, they interact with the chromatin assembly factor hCAF-1 specifically during the synthesis phase (S phase), depending on actual occurrence of replication. Silencing maintenance of many hMBD1 targets, like *p53BP2*, depends on both DNAme and H3K9me3 deposition by hMBD1-SETDB1 complex during replication, but not on histone deacetylation. Absence of any of those components leads to H3K4me3 accumulation and transcription. The SUMOylation of any of the two hMBD1 sites by PIAS1/3 E3 SUMO-ligases disengages hSETDB1 from p53BP2 promoter, but not hMBD1, reducing the levels of H3K9me3 and disturbing repression (Lyst et al., 2006). Methylation of non-nucleosomal H3 by hSETDB1 was also observed in the context of its association with the hCAF-1-HP1α-TRIM28 complex, where it monomethylates core histones rather than mononucleosomes (Loyola et al., 2009).

Regarding early development, SETDB1 foci in mouse blastocysts co-localize with the promyelocytic leukemia nuclear bodies (PML-NB). This association was also observed in MEF, NIH/3T3 (Cho et al., 2011), mESC and HEK293T, with the SUMO1-modified version of SETDB1 (Yeap et al., 2009) and the shorter isoform (Cho et al., 2013). SETDB1 depletion inhibits MEF from reconstituting PML-NB after its dismantlement. Both PML and SETDB1 associate on the promoter of the PML target *Id2* to repress its transcription (Cho et al., 2011). The overall gene deregulation in *Setdb1* knockout differs from *Dnmt* TKO mESC. Most of SETDB1 binding sites are enriched for H3K9me3 in wild type mESC and lose this mark after SETDB1 knockout, but not in DNMT TKO mESC. However, only few such sites occurring at promoters become upregulated, with a minority being enriched for H3K9me3 in wild type mESC. Thus, in general, H3K9me3 deposition by SETDB1 is independent of DNAme (Karimi et al., 2011). Imprinted genes were shown to be regulated by SETDB1 in mESC (Yuan et al., 2009) in the context of ZFP57-TRIM28-SETDB1 complex. This ZFP is recruited by a specific methylated hexanucleotide that occurs in one or two copies in all murine and some human imprinted genes. Together with HP1γ, this complex binds preferentially the repressed allele and is required for maintaining H3K9me3 and DNAme and blocking H3K9ac deposition, while CTCF interacts with the other allele. The ZFP57-TRIM28 complex associates also with DNMT1, DNMT3A/B and UHRF1 (Quenneville et al., 2011). The maternally imprinted gene *hSNOG1* is instead bound by ZFP274 in hESC and is repressed by hSETDB1-dependent H3K9me3 and DNAme (Cruvinel et al., 2014). At imprinted and germline genes, SETDB1-dependent H3K9me3 is required for DNAme maintenance (Leung et al., 2014).

SETDB1 is also involved in gene regulation in the context of signal transduction pathways. In the nuclei of HeLa cells, hSETDB1 also interacts with the human serine/threonine kinase hAkt1, a central effector downstream of phosphoinositide 3-kinase (PI3K) required in many signal transduction pathways. Phosphorylation of the transcription factor hFKHRL1 by hAkt1 is enhanced by the interaction with hSETDB1 and retains hFKHRL1 in the cytoplasm, where it cannot promote expression of apoptotic genes, like FasL (Gao et al., 2007). Besides that, signaling pathways might induce HKMT-dependent repression, as depletion of the protein kinase CaMKII in MCF7 breast cancer cells led to impaired hSETDB1-dependent repression of a reporter

gene regulated by the estrogen receptor (Garcia-Bassets et al., 2007). In murine bone marrow-derived stroma ST2 cell line, Wnt-5a signaling through CaMKII-TAK1-TAB2 activates Nemo-like Kinase NLK, which phosphorylates SETDB1 at T976. This modified form associates with the chromodomain helicase DNA-binding protein CHD7 to repress the adipogenesis inducing nuclear receptor PPAR-γ through deposition of H3K9me2/3 at its binding site on gene promoters, consequently promoting osteoblastogenesis rather than adipogenesis (Takada et al., 2007). SETDB1 also complexes with the thyroid hormone receptor TRβ1 when it binds the corepressor Alien in murine carcinoma P19 cells. After stimulation with T3, this complex binds the negative thyroid hormone response element nTRE of the E2F1 and c-myc promoters to repress their activity with concomitant deposition of H3K9me3 (Hong et al., 2011). A role in silencing of quickly inducible genes during inflammation response was proposed for hSETDB1. The Interferon Regulatory Factor IRF5 induces pro-inflammatory cytokines, such as the tumor necrosis factor TNF. To prevent TNF expression for longer than needed, IRF5 recruits hTRIM28 and hSETDB1, for deposition of low H3K9me3 levels in HEK293 cells and in human GM-CSF differentiated macrophages. Even though these low levels are enough for silencing, they are easily removed upon a challenge, allowing quick reactivation of transcription (Eames et al., 2012).

Recruitment of hSETDB1 to 3' end of zinc finger genes (*Zfp*) in human immortalized myelogenous leukemia K562 cells was also observed. The KRAB zinc finger protein hZFP274 targets hTRIM28 to these specific genomic regions, which in turn targets hSETDB1 for deposition of H3K9me3 (Frietze et al., 2010). hATRX associates with this complex together with the atypical co-occupancy of H3K36me3 at such regions. Co-binding of ATRX was also seen in hESC and HeLa, but not in mESC and MEF (Valle-Garcia et al., 2016). Interestingly, a repetitive region previously known as MMSAT4 is present at the 3' end of almost all *KRAB-Zfp* genes and gets silenced by the TRIM28-SETDB1 complex (Kauzlaric et al., 2017).

An interesting interactor of SETDB1 is the methionine adenosyltransferase MAT2, which synthesizes SAM, a methyl donor for histone methylation. The cyclooxygenase gene *Cox-2*, important in inflammation, is bound by the transcription factor MafK, which recruits the subunits MAT2A/B. MAT2A, in turn, associates with SETDB1 to deposit H3K9me3 and repress transcription in MEF (Kera et al., 2013). Furthermore, a function in RNA-mediated silencing was described. In human breast cancer T47D cells, ncRNA are transcribed from promoters with androgen receptor and are targets of small interfering antigene RNA (agRNA) which recruit the hAGO2-hSETDB1 complex. hSETDB1-dependent H3K9me3 and hEZH2-dependent H3K27me3 are required for this silencing. The complex hSIN3-hHDAC2, but not hTRIM28 or hDNMT3A/B, is also recruited to those regions in an hAGO2- and hSETDB1-dependent manner (Cho et al., 2014).

In *Drosophila*, *dSetdb1* depletion upregulates genes mainly of the chromosome 4 (Brower-Toland et al., 2009). However, this finding opposes the general downregulation previously observed (Tzeng et al., 2007). In S2 cells, dSETDB1 is not expressed and the *Rb* gene gets

transcribed. Ectopic expression of dSETDB1 is enough to repress *Rb* transcription. dSETDB1 deposits H3K9me3 at a distal promoter region and triggers DNAme by dDNMT2, which spreads to a proximal promoter region with the help of HP1. H3K9me3 deposition at this proximal region depends on the MBD of dSETDB1 and the spread DNAme or silencing is not accomplished. This mechanism of *Rb* silencing is observed in the developing eye imaginal disc (Gou et al., 2010).

Overall, several genes were demonstrated to have their promoters regulated by SETDB1 in a methyltransferase dependent manner. The regulation of developmental genes, though, was merely based on genome-wide analyses and has not yet been well curated so far. Thus, how SETDB1 acts to suppress developmental-related genes, what exactly its roles are in keeping proper development and whether this is a direct role of SETDB1 and not some consequence of its other functions are still open questions.

1.7. Repression of transposable elements

A significant proportion of human and mouse genomes is comprised of transposable elements (TE), reaching 46 % and 37.5 %, respectively. These elements are divided in DNA transposons and retrotransposons (Wicker et al., 2007), with the last one relying on an RNA intermediate to amplify. Retrotransposons constitute 95 % of mouse TE and are classified in non-LTR and LTR (long-terminal direct repeats) retrotransposons. The non-LTR order forms the majority of retrotransposons (27 % of the mouse genome) and is constituted by autonomous LINE (long-interspersed) and non-autonomous SINE (short-interspersed nucleotide elements). The LTR order comprises around 9.9 % of both human and mouse genomes. While they are almost extinct in humans, many members are still active in mice. In mammals, LTR retrotransposons derived from vertebrate-specific endogenous retroviruses (ERV) superfamily. According to similarity to modern exogenous retroviruses (XRV), ERV are grouped in classes I, II and III, constituting 0.7 %, 3 % and 5.4 % of the mouse genome. Class I gathers the families MLV, GLN, VL30, MRRS, MRVY and MERVC; class II contains the superfamilies MMTV, IAP and MusD/ETn; and class III is composed by families MERVL and MaLR (Stocking and Kozak, 2008). Interestingly, ERV are differentially regulated during development and can even become expressed. While somatic cells rely on the DNAme machinery for silencing these elements, mouse early embryos and embryonic stem cells take advantage of histone modification mechanisms to keep them silent during the global DNA demethylation during the preimplantation stage (Gifford et al., 2013).

SETDB1 has an important role in TE silencing during reprogramming of DNAme in early stages of development. Transcriptional repression of ERV classes I (MLV, GLN) and II (IAP, MusD, ERVK10C) depends on SETDB1 binding and H3K9me3 deposition in mESC (Matsui et al., 2010; Karimi et al., 2011; Maksakova et al., 2011; Reichmann et al., 2012; Maksakova et al., 2013), but not on DNAme by DNMTs in general (Karimi et al., 2011; Reichmann et al., 2012). Ubiquitination of SETDB1 is essential for this regulation (Sun and Fang, 2016). LINE-1 show only mild upregulation after SETDB1 depletion. Loss of H3K9me3 at these ERV classes happens after

SETDB1 depletion (Matsui et al., 2010; Karimi et al., 2011), but not in *Dnmt* TKO (Karimi et al., 2011) or *Dnmt3a/b* DKO (Leung et al., 2014), and is accompanied by loss of H4K20me3 but not H3K9me2. However, depletion of SUV4-20H1/2 enzymes did not lead to reduction of H3K9me3. HP1α/β/γ is also partially lost at these ERV elements. TRIM28 binds them in mESC independently of SETDB1, even though binding of the latter depends on TRIM28. Knockdown of TRIM28 leads to similar phenotype. In MEF, ERV classes I and II are not bound by either SETDB1 or TRIM28 and only MLV show strong derepression after SETDB1 depletion. Deposition of SETDB1 is independent of DNMTs and their product and SETDB1 loss does not affect DNAme levels strongly in mESC (Matsui et al., 2010) or in blastocysts (Dodge et al., 2004). On the other hand, SETDB1 is involved in maintaining residual DNAme at specific H3K9me3-marked subfamilies of ERV class II (IAP, MusD and ERVK10C) and they preserve some DNAme in *Dnmt* DKO mESC. At these regions, SETDB1 opposes DNA demethylation activity, as they showed increased Tet1 binding and 5hmC levels in *Setdb1* KO mESC (Leung et al., 2014). Besides that, only IAPEz subfamily is further upregulated when DNMT1 was depleted together with SETDB1. Expression of genes near upregulated ERV tended to increase, along with chimeric transcripts (Karimi et al., 2011).

Loss of *Dnmt1* in mESC increases hemimethylated DNA, leading to prolonged UHRF1 binding that disrupts SETDB1 and KAP1 binding. H3K9me3 is partially lost and IAPEz transcription increases. This picture is reversed later after hemimethylated DNA is reduced. Loss of only *Uhrf1* resulted in the opposite picture, enhancing the repression status. *Setdb1*/*Dnmt1* DKO and *Setdb1*/*Uhrf1* DKO in mESC showed synergistic effect in IAPEz upregulation. In trophoblast stem cells, UHRF1 blocks SETDB1-dependent repression, as IAPEz expression increases after *Setdb1* knockout and gets more repressed in *Uhrf1* knockout (Sharif et al., 2016).

SETDB1 also deposits H3K9me3 at all H3.3-containing ERV. While HIRA deposits H3.3 at ERV enriched only for H3.3, DAXX and ATRX are responsible for deposition of H3.3 in ERV enriched for both H3.3 and H3K9me3. At IAPEz, H3.3 deposition decreases after *Trim28* deletion, but not after *Setdb1* deletion. Deletion of H3.3 reduces binding of both TRIM28 and DAXX to ERV sequences and decreases H3K9me3 levels at IAP, ERVK10C, ETn and MusD, leading to derepression of the first two subfamilies (Elsasser et al., 2015). However, it did not affect silencing of an IAPEz gag reporter, where expression of a H3.3 interaction-deficient DAXX could still rescue reporter repression in *Daxx* knockout mESC (Sadic et al., 2015).

Expression of the *Agouti* allele affecting coat color in mice is driven by an IAP retrotransposon. Heterozygous *Setdb1* or *Trim28* mutant mice are haploinsufficient, but survive and are fertile, though with lower probability of repressing paternally inherited *Agouti* allele. H3K9me3 is absent in mature sperm and global DNAme level is unchanged in mutant sperm. However, IAPEY showed loss of DNAme, which cannot be reversed by maternal contribution during preimplantation (Daxinger et al., 2016). In E13.5 primordial germ cells, some ERV elements (IAPEz and ERVK10C) are enriched for H3K9me3, DNAme and, differently from mESC,

H3K27me3. *Setdb1* deletion in those cells, driven by *Tnap* promoter, results in reduction of DNAme at LTR region of IAP which lost H3K9me3. The reactivation of a subset of ERV is sex-dependent. IAPEz and ETn were more upregulated in male, while ERVK10C was more upregulated in female cells. The derepression of some ERV resulted in upregulation of genes at their vicinity (Liu et al., 2014). Lack of SETDB1 in oocytes results in further deregulation of ERV classes II and III (Eymery et al., 2016) and LINE1 (Kim et al., 2016). At the X chromosome, only LTR order of X-linked repeats showed impaired silencing after *Setdb1* knockdown in differentiating mESC (Keniry et al., 2016).

In the hematopoietic system, MLV, MMTV and VL30 contain SETDB1-dependent H3K9me3 and are derepressed in *Setdb1*-deleted pro-B cells. Upregulation of some MLV elements, with mild or no DNAme loss, led to increase expression of their neighboring genes (Collins et al., 2015; Pasquarella et al., 2016). However, in *Setdb1* knockout GMP, LSK (Koide et al., 2016) and thymocytes (Takikita et al., 2016), IAP were also upregulated. ERV expression is directly dependent on the cell type, as a MLV element derepressed in *Setdb1* knockout pro-B cells stays silent in *Setdb1* knockout mESC, while the opposite happens to a nearby IAPEz sequence. Specific derepression of particular ERV is due to tissue-specific transcription factors, which explains why loss of *Setdb1*, H3K9me3 and DNAme not always lead to transcriptional activation. For instance, the derepressed MLV elements in mutant pro-B cells are bound by the B-lineage transcription factor PAX5, whose depletion abolishes derepression in mutant pro-B cells (Collins et al., 2015). Upregulation of nearby genes also happens after *Setdb1* deletion in mouse brain, due to derepression of IAP losing H3K9me3 and DNAme (Tan et al., 2012).

A MLV-based exogenous retrovirus gets quickly silenced in mESC with accumulation of H3K9me3, while de novo DNAme is only observed several days after infection. However, SETDB1 deletion leads to derepression of the reporter with concomitant DNAme loss, similarly to TRIM28 knockdown (Matsui et al., 2010). Similar SETDB1-dependent repression was observed in mESC, but not in MEF, for IAP LTR (Rowe et al., 2013) and IAPEz gag reporters (Sadic et al., 2015); for MSCV, MFG, IAP or MusD LTR in mESC (Maksakova et al., 2011); for MSCV LTR in pro-B cells (Collins et al., 2015); and when reporters with CMV or LTR promoters were introduced in HeLa cells (Poleshko et al., 2010). Interestingly, silencing of an IAPEz gag reporter becomes further impaired when *Setdb1* is depleted together with *Atrx* in mESC, indicating the occurrence of genetic interaction (Sadic et al., 2015). Whether an introduced ERV gets DNA methylated or not after transduction is dependent on the pattern of KRAB-ZFP expression. Both mESC and MEF, but not NIH/3T3 cells or 293T cells, express ZFP809 and can repress the PBS Pro derived from MLV. However, expression of this gene in 293T cells was enough for repression and deposition of DNAme. Injection of the IAP LTR reporter in fertilized mouse oocytes results in repression and DNAme. Even though transduced MEF cannot repress this reporter, it is repressed in MEF derived from transduced embryos. The early deposition of DNAme was kept

and is essential for the maintenance of the repression, which is impaired by treatment with the DNA methyltransferase inhibitor 5-aza (Rowe et al., 2013).

SETDB1 and the human silencing hub (HUSH) complex proteins are required for H3K9me3 deposition and repression of the spleen focus-forming virus promoter (SFFV) reporter in haploid human cell line KBM7 and in HeLa cells. The HUSH complex is formed in the nucleus by the transgene activation suppressor TASOR, the M-phase phosphoprotein MPP8 and periphilin. Reporter integrations silenced by the HUSH complex happened in the vicinity of H3K9me3-enriched regions, as a result of position-effect variegation (PEV), and are dependent on H3K9me3 binding by MPP8. H3K9me3 deposition at some endogenous loci and repression also depended on this complex in HeLa cells (Tchasovnikarova et al., 2015).

In NIH/3T3 fibroblasts, association of SETDB1 to the HP1α-CAF1 complex establishes non-nucleosomal H3K9me1 at HP1-associated pericentric heterochromatin during replication (mid S-phase). This complex is not associated with ATF7IP (Loyola et al., 2009). Besides that, a role for SETDB1 at major satellite repeats of pericentric heterochromatin was also described in concert with SUV39H1, G9a and GLP as a multimeric complex. This interaction is observed endogenously in HeLa, MEF and mESC. The stability and HKMT activity of the complex depends on the enzymatic activity and the chromodomains of SUV39H1. These four proteins work cooperatively to establish and maintain gene silencing at euchromatic G9a targets too (Fritsch et al., 2010).

In the wing imaginal disc, spread of DNAme due to dSETDB1-dependent H3K9me3 deposition is important for silencing of retrotransposons (Gou et al., 2010). In third instar larvae, however, loss of *dSetdb1* does not strongly impact pericentric heterochromatin or the centromere-proximal region of chromosome 4 (Riddle et al., 2012). In first instar larvae, upregulated transposons overlap in *dSetdb1*, HP1α and *dSu(var)3-9* mutants, but loss of *dSetdb1* has lower extent effects (Lundberg et al., 2013). In *C. elegans*, binding of SETDB1, HP1, CHD, ZFP and MBT homologs is well correlated with H3K9me3 and are associated with telomeres and repetitive elements, which can get upregulated in mutants of SETDB1 homolog (Zeller et al., 2016; McMurchy et al., 2017).

1.8. Consequences of SETDB1 overexpression

SETDB1 full-length and splicing variant are almost completely retained in the cytoplasm when overexpressed in NIH/3T3 fibroblasts, HEK293 cells, MEF and mESC. It reaches the nucleus only when the nuclear export protein CRM1 is blocked by leptomycin B (Cho et al., 2013). Inhibition of proteasome degradation by MG132 further increases nuclear localization (Tachibana et al., 2015). Mice with higher SETDB1 transgenic expression in the forebrain showed antidepressive phenotype. Increase in pericentric H3K9me3, at major satellite repeats, was observed. Besides that, at higher expression levels, SETDB1 targeted and repressed the NMDA receptor subunit *Grin2b*, similarly to what happens in human U87MG glioma cells. Interestingly, the TSS region of

Grin2b bound by TRIM28 forms a loop with a region 30 kb away inside the gene, which is bound by SETDB1 (Jiang et al., 2010).

In normal mature neurons, *Setdb1* expression is reduced while the MBD protein MECP2, which regulates repressive histone modifications, is upregulated. Overexpression of *Setdb1* in *Mecp2* neuronal knockout mice worsened the neurological impairment, instead of ameliorating it (Jiang et al., 2011). Furthermore, enhanced expression of SETDB1 in mice due to loss of its repressor CREB binding protein CBP increases H3K9me3 in striatal neurons. These mice show brain atrophy and striatal neuron dysfunction (Lee et al., 2008). In *Drosophila*, overexpression of d*Setdb1* is lethal and flies die at pupal stage. It leads to increase in global H3K9me1/2/3 levels, enhancing HP1 recruitment on chromosome 4 and on euchromatic arms (Seum et al., 2007).

Overexpression of *Setdb1* can also impair differentiation. When performed in differentiating myoblasts, it reduced the expression of differentiation markers, disrupting terminal differentiation (Beyer et al., 2016). Overexpression of wild type *Setdb1*, but not the ubiquitination site mutant, in HeLa cells increases H3K9me3 level at the target gene *SERPINE1*, decreasing expression. This suggests that ubiquitination enhances the activity at target loci in the cell (Ishimoto et al., 2016).

In summary, SETDB1 is involved in the repression of genes and repetitive sequences by different mechanisms and in numerous circumstances, especially during differentiation. Moreover, SETDB1 HKMT activity must be regulated properly according to each context. Several are the proteins which were shown to act in concert with SETDB1. However, only one of those interactors is regarded as a regulator of its methyltransferase activity and is known as ATF7IP.

1.9. The transcriptional cofactor ATF7IP

1.9.a. Structural features

ATF7IP was first identified as a murine ATFa-associated modulator (mAM) with 1306 aa and predicted molecular weight of 138 kDa, although it migrates as a 180 kDa protein (De Graeve et al., 2000). At the same time, it was found as a partner of Sp1, named p621 (Gunther et al., 2000). Later on, it was identified in HeLa cells as an interactor of transcriptional repression domain (TRD) of hMBD1, termed MCAF, for MBD1-containing chromatin associated factor (Fujita et al., 2003). The *Drosophila* homolog, *Windei*, has only 14.8 % identity with *Atf7ip*, however, the identity of the fibronectin type III repeat at the C-terminus is 36 %, with a similarity of 56 % (Koch et al., 2009).

A NLS close to the center of the protein is enough to target a cytoplasmic reporter protein to the nucleus. At the N-terminal side, a putative ATP binding site (ABS) is present, which was shown to be required for ATP hydrolysis activity of immunoprecipitated material. No canonical DNA binding sequence is observed (De Graeve et al., 2000). ATF7IP bears two other conserved domains, with identity of 76 % and 98 % between mouse and human homologs, which are important for interaction with other proteins. Domain 1 interacts with Sp1 (Fujita et al., 2003;

Ichimura et al., 2005; Liu et al., 2009) and SETDB1 (Ichimura et al., 2005). The C-terminal Domain 2 interacts with MBD1 and Sp1 (Fujita et al., 2003; Ichimura et al., 2005; Liu et al., 2009).

Although no covalent modification by SUMOylation is observed, a SIM between Domains 1 and 2 binds SUMO2/3 strongly and SUMO1 weakly. This binding is abrogated by D968A and L969A in humans. (Uchimura et al., 2006; Tanaka and Saitoh, 2010). Structural characterization of the SUMO3 binding to hATF7IP indicates a role for electrostatic interactions in the preferential association to SUMO3 (Sekiyama et al., 2008). Analysis of the phosphoproteome of hESC during differentiation identified five serine phosphorylation sites in hATF7IP (Rigbolt et al., 2011).

ATF7IP belongs to an evolutionarily conserved protein family and has a paralog in mammalians, called ATF7IP2. The gene *Atf7ip2* encodes a 681 aa protein with 27 % and 56 % similarity to the domains 1 and 2 of ATF7IP, respectively (Ichimura et al., 2005). In *Drosophila*, the C-terminal region of dATF7IP is enough for its nuclear localization. A coiled coil region (Domain 1) interacts with the dSETDB1 region N-terminally to its tudor domain (Koch et al., 2009).
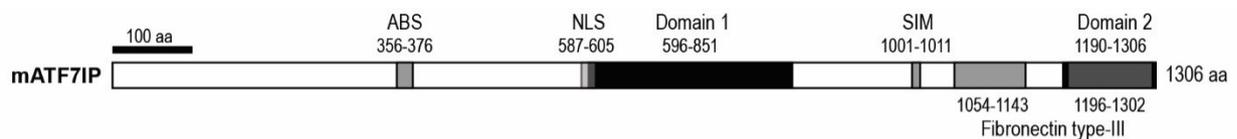
Figure 1.2 | Schematic representation of the domain structure of mouse ATF7IP.
ATF7IP contains a putative ATP binding site (ABS). Two conserved regions called Domains 1 and 2 are involved in protein interaction. A nuclear localization signal (NLS) overlaps with the beginning of Domain 1. A functional SUMO interaction motif (SIM) is present. Two fibronectin domains compose the C-terminus of the protein.

1.9.b. Expression and localization

Atf7ip is ubiquitously transcribed during embryogenesis, at E9.5, E11.5, E14.5 and E16.5. E11.5 embryos show stronger signals in the brain, spinal cord and somatic mesoderm. At E14.5 and E16.5, the nervous system kept strong signals, mainly in the forebrain, but numerous organs also showed high expression. Several adult mouse tissues express *Atf7ip*, for instance the epithelium of the stomach, epididymis, ductus deferens, tongue, intestine and uterus and in the spleen and some areas of the brain (De Graeve et al., 2000). During differentiation of mESC and hESC into embryoid bodies, *Atf7ip* expression decreases (Luzzani et al., 2011).

While *ATF7IP* is highly expressed in HeLa, HEK293T and K562 cells, *ATF7IP2* is not expressed only in HeLa cells. hATF7IP overexpressed in HeLa cells shows nuclear colocalization with hMDB1, hSETDB1 (Ichimura et al., 2005) and SUMO3 incapable of conjugation (Uchimura et al., 2006). hATF7IP forms foci in interphase nuclei of HeLa cells, which co-localized hMBD1 (Fujita et al., 2003). Mutation on the Domain 2 of *ATF7IP* or on the TRD of *MBD1* leads to dissociation of the complex and disruption of hSETDB1 localization, indicating that the latter is recruited to DNA methylated regions by hMBD1 via hATF7IP. The complex hMBD1-ATF7IP-SETDB1 also co-localizes with the three HP1 isoforms, which adopted a diffuse distribution when

hMBD1 was mutated. This co-localization is observed for the endogenous proteins as well in human cervix carcinoma C33A cells and overlap with heterochromatic foci, including with the H3K9me3 mark (Ichimura et al., 2005) and SUMO2/3. hATF7IP colocalization with the heterochromatic H3K9me3 mark was also observed in HeLa and MCF7 cells (Sasai et al., 2013). Interestingly, part of the endogenous hMBD1 is SUMOylated by SUMO1/2/3, what leads to higher stability in the association with hATF7IP. Depletion of either SUMO1 or SUMO2/3 leads to impaired hATF7IP, H3K9me3 and HP1β/γ co-localization to MBD1 foci (Uchimura et al., 2006).

dATF7IP is expressed in germ line, somatic follicle and nurse cells during fruit fly oogenesis. It partially colocalizes to HP1 foci at the oocyte nucleus and colocalizes exactly to dSETDB1 and POF at the fourth chromosome. dATF7IP, HP1 and POF overlap at many euchromatic bands of all polytene chromosomes in salivary glands. Deletion of dSETDB1 SET domain does not impair proper dATF7IP localization. However, deletion of dATF7IP affects dSETDB1 nuclear localization and protein stabilization in germ line cells (Koch et al., 2009). In HeLa cells, hATF7IP is almost exclusively nuclear and stabilizes nuclear hSETDB1, protecting against proteasomal degradation, while nuclear hSETDB1 also prevents hATF7IP proteasomal degradation (Timms et al., 2016).

1.9.c. ATF7IP-mediated transcriptional control

Ectopic expression of *Atf7ip* when the transcriptional activator complex ATFa-JunD is targeted to a reporter promoter leads to ABS-independent repression. However, artificial tethering of ATF7IP to a reporter promoter in mouse testis embryonal carcinoma F9 cells also leads to repression (De Graeve et al., 2000). Interaction with subunits of the general transcription factors TFIIE (α/β subunit) and TFIIH (ERCC2/3 and p34) and of the Poll-II (hRPB3/4/7/8) was observed ectopically in Sf9 (De Graeve et al., 2000) and for endogenous proteins in HeLa (Liu et al., 2009).

Whether hATF7IP exerts transcriptional repression or activation depends on the proteins complexing with it. Expression of *ATF7IP* in HeLa cells induces transcriptional activity of a reporter with promoters of *SNRPN* or *p16*, known to contain hSp1 binding motif and an associated CpG island. Similarly, co-expression of *ATF7IP* and *Sp1* in SL2 cells, which lack Sp1, MBD1 and DNAme, enhances transcription. However, when hMBD1 is also co-expressed, transcription is repressed instead (Ichimura et al., 2005). Both activation and repression happened in a Domain 2-dependent manner (Fujita et al., 2003). Point mutations on the TRD domain at residues important for hMBD1 repressive ability abrogates binding to hATF7IP. The promoter of *p16* is not methylated in the SBC-5 cell line and hATF7IP binds it via hSp1, independently of hMBD1. In the lung cancer cell line NCI-H1299, though, this promoter is methylated and can be bound by hMBD1 followed by hATF7IP, unless the point mutant hMBD1 is expressed instead. Interestingly, hMBD1 and hSp1 compete with each other for the binding to ATF7IP (Fujita et al., 2003).

hATF7IP is involved in autoregulation of the Epstein-Barr virus EBV transcription factor Rta, encoded by *BRFL1*, leading to activation of lytic cycle. Basically, Rta associates with Sp1 via

ATF7IP and is recruited to the Sp1 motif in the *BRFL1* promoter in P3HR1 lymphoma cells (Chang et al., 2005). Besides that, the Domain 1 of hATF7IP, Rta and the *BZLF1*-encoded transcription factor Zta form a complex at Zta response elements of EBV lytic genes, acting synergistically to increase their expression (Chang et al., 2010). hATF7IP regulates *BZLF1* promoter as well, by mediating the interaction between Rta and hATF2 after lytic induction. This complex also binds the promoter of *BMRF2*, a gene important for the viral infectivity (Lin et al., 2014).

The telomerase-associated genes *TERT* and *TERC*, transactivated by hSp1, show binding of both hATF7IP and hSp1 at their promoter regions. hATF7IP binding to the *TERT* promoter depends on hSp1, but the opposite is not true. Both proteins are required for proliferation, telomerase activity, proper expression of *TERT* and *TERC* and promoter occupancy by phospho-Pol-II and hERCC3. Some CpG sites between the Sp1-binding motifs in *TERT* promoter get methylated after depletion of either hATF7IP or hSp1, while hCTCF binding sites lost CpG methylation, even though it did not impact hCTCF binding. hMBD1, but not hSETDB1, is enriched in the *TERT* promoter and is independent of hATF7IP (Liu et al., 2009). hATF7IP is essential for stimulating growth rather than cellular senescence also by keeping expression of cell cycle, core histones and linker histone H1 genes and inhibiting Rb protein dephosphorylation. In human fetal lung IMR90 myofibroblasts, hATF7IP forms several foci colocalized to PML-NB, but not with H3K9me3. Depletion of this protein did not disassemble the PML-NB. However, proliferation was impaired and premature senescence triggered. Induction of senescence enhances has a similar effect, as hATF7IP accumulates to PML-NB in a SIM-dependent manner probably due to SUMO2/3 binding and this sequestration leads to cellular senescence (Sasai et al., 2013).

ATF7IP complexes with MBD1 and SETDB1 also to regulate the expression of X-linked genes for XCI in concert with DNAme. Knockdown of *Atf7ip*, *Mbd1* or *Setdb1* (but not other H3K9MT) in female MEF leads to weak activation of an X inactive-linked reporter, similarly to treatment with low levels of azacitidin, which removes DNAme. However, a combination of knockdown and azacitidin treatment or co-depletion with *Dnmt1* enhances reactivation. ATF7IP does not bind the X inactive during initiation of the process after induction of *Xist* expression in mESC. Enrichment was still not observed after retinoic acid differentiation (Minkovsky et al., 2014). Similar to what was shown for SETDB1 (Chapter 1.5.a. Embryonic stages), ATF7IP does not regulate *Xist* expression or the coating of the inactive chromosome by Xist RNA, even though concomitant depletion of *Atf7ip* and *Xist* also showed synergistic effects (Minkovsky et al., 2014). ATF7IP and SETDB1 functions in XCI may be involved with the chromodomain-containing transcriptional corepressor *Cdyl*, which binds H3K9me2 in the X inactive chromosome and interacts with ATF7IP1/2, SETDB1 and G9a during mESC differentiation (Escamilla-Del-Arenal et al., 2013).

The hATF7IP-MBD1 complex is also shown to promote tissue-specific antigen encoding genes. These genes are targets of the transcriptional regulator Aire, specific of specialized medullary thymic epithelial cells (mTEC). In these cells, Aire interacts with Domain 2 of hATF7IP

and with hMBD1, which act to recruit Aire to its targets, leading to transcriptional activation (Waterfield et al., 2014). An unrelated possible role of hATF7IP, which has been so far overlooked, is the involvement in repressing homologous gene targeting (Delacote et al., 2011).

1.10. Regulation of SETDB1 catalytic activity by ATF7IP

Co-fractionation of HeLa protein extracts monitoring HMTase activity and SETDB1 led to the identification of hATF7IP. When reconstituted from Sf9 cells, the mouse SETDB1-ATF7IP complex showed higher activity than SETDB1 alone. Besides, the activity of the complex was similar to equal amounts of HeLa cells native complex (Wang et al., 2003). This happens due to an increase in the turnover rate of the reaction. The same did not occur after reconstitution with the human homologs, where the complex rather showed lower activity (Basavapathruni et al., 2016). Unexpectedly, when the mouse proteins are combined after being individually purified from Sf9 cells, the methyltransferase activity is inhibited instead (Wang et al., 2003). Similar phenomenon was seen between HDAC and its modulator MTA2 (Zhang et al., 1999).

ATF7IP does not change substrate specificity, as mutations in other residues does not affect activity of either recombinant SETDB1 or the reconstituted complex (Wang et al., 2003). hATF7IP also did not change the activity of hSETDB1 over a large peptide panel of histones H3 and H4. Besides that, hATF7IP does not alter hSETDB1 affinity to substrate or SAM (Basavapathruni et al., 2016). The reconstituted complex has much higher efficiency in trimethylation in comparison to SETDB1 alone, while mono- and dimethylation are similar (Wang et al., 2003). In contrast, for the human reconstituted complex the enzyme turnover rate is rather lower for all substrates, indicating lower catalytic efficiency when compared to hSETDB1 alone. Interestingly, mono and dimethylation promptly accumulate over time, while H3K9me3 is much slower, independently of hATF7IP presence. Catalysis by hSETDB1 is distributive, meaning that the substrate is released right after catalysis (Basavapathruni et al., 2016).

Depletion of *SETDB1* in HEK293T and HeLa cells leads to mild reduction of global H3K9me2/3 levels. On the other hand, *ATF7IP* depletion decrease H3K9me3 and increases H3K9me2. Transcription of an *in vitro* assembled-chromatin, but not of naked DNA, is strongly repressed by hSETDB1 in a SAM-dependent manner, but is further repressed by hSETDB1-ATF7IP complex. Oligonucleosomes were marked with H3K9me2 by hSETDB1 alone, while in the presence of hATF7IP there was enrichment for H3K9me3 instead. The methylation is restricted to the promoter region and depends on the activator, which may open up chromatin (Wang et al., 2003).

In the case of the *SNRPN* promoter introduced in SL2 cells, where hATF7IP cooperates with hMBD1 to enhance silencing when ectopically expressed (Chapter 1.9.c ATF7IP-mediated transcriptional control), hSETDB1 further enhances repression. The same does not happen when hATF7IP2 is in the place of hATF7IP. In HeLa cells, *ATF7IP* depletion impaired MBD1 repression of this promoter and was only partially restored by *ATF7IP2* expression. Similarly, the methylated

promoter of *p53BP2* (Chapter 1.6.b. Endogenous genes) is bound by the hMBD1-SETDB1-ATF7IP complex (Ichimura et al., 2005).

The RNA-binding protein and transcription co-factor heterogeneous nuclear ribonucleoprotein hnRNPK interacts with TRIM28 in mESC. It is required for SUMOylation of TRIM28 by the SUMO E2 conjugating enzyme Ubc9 and concomitant recruitment of SETDB1 and ATF7IP. Knockdown of *Setdb1* or *Atf7ip* in mESC leads to derepression of proviral reporters MSCV-PBS and IAP LTR-PBS with loss of H3K9me3 (Thompson et al., 2015) and derepression of a MMLV reporter (Yang et al., 2015). Depletion of either of those proteins mildly derepresses ERV classes I and II (Thompson et al., 2015; Yang et al., 2015), with partial loss of H3K9me3. Depletion of *Hnrnpk* also impairs repression of these proviral reporters and ERV classes I and II, which are also bound by hnRNPK. While *Dnmt* TKO mESC show just mild upregulation of IAP, depletion of *Setdb1*, *Hnrnpk* or *Atf7ip* leads to further deregulation of ERV. Several SETDB1 target genes, including some germline genes bound by hnRNPK, get upregulated or even lose H3K9me3 after depletion of either *Setdb1* or *Hnrnpk* in mESC. Interestingly, depletion of *Atf7ip* increases SETDB1 occupancy at ERV (Thompson et al., 2015).

ERV deregulation shows strong correlation between *Setdb1*, *Atf7ip* or *Trim28* knockdown mESC. *Sumo2* knockdown reduces TRIM28 binding, including at the exogenous proviral reporter. Besides that, ERV regulated by the ZFP809-TRIM28-SETDB1 machinery overlap with the ones regulated by the histone chaperone SUMO2, but not by CHAF1A. However, CHAF1A interacts with SETDB1, KDM1a and HDAC2 to also regulate proviruses independently, by removing H3K4me3 and H3ac and providing a multi-layered silencing mechanism (Yang et al., 2015).

hATF7IP is also involved in silencing of a HUSH complex-repressed reporter in HeLa cells. The HUSH complex was described in a previous section (Chapter 1.7. Repression of transposable elements) and most of the loci which lose H3K9me3 after depletion of this complex also lose this mark both in *ATF7IP* and in *SETDB1* knockouts. Striking loss was observed at the body of zinc finger protein genes. Furthermore, transcriptome changes were also very similar in both knockouts, indicating the interdependency between those proteins (Timms et al., 2016).

In 3T3-L1 preadipocytes, SETDB1 binds the DNA methylated gene body of *Cebpa*, as mentioned before (Chapter 1.5.e. Musculoskeletal system). This region is also bound by MBD1 and ATF7IP. Depletion of either protein leads to *Cebpa* increased expression, reduced SETDB1 binding and loss of H3K9me3. Thus, MBD1 recruits ATF7IP and SETDB1 to methylated DNA targeted via its MBD (Matsumura et al., 2015).

In *Drosophila*, deletion of *dAtf7ip* leads to death at pupal stage. There were few escapers, which were very weak and died few days after becoming adults. The ovaries did not develop to bud off egg chambers from the germarium in mutant females. *dAtf7ip* is necessary for survival of germ line cells and for keeping H3K9me3 levels, similarly to *dSetdb1* loss (Koch et al., 2009).

Considering all that, the influence of ATF7IP on SETDB1 methyltransferase activity is directly supported only by in vitro experiments, even though the results were contrasting between the human and the mouse version of these proteins. In a cellular system, ATF7IP function as a cofactor is only supported by indirect observations, where loss of either protein leads to loci specific H3K9me3 reduction. However, studies addressing the actual mechanisms are lacking.

1.11. SETDB1 and ATF7IP in cancer and other diseases

The importance of *Setdb1* in neuronal development could already indicate possible roles in neuropathology. Indeed, higher levels of hSETDB1 expression and H3K9me3 accumulation is observed in neurons of Huntington's disease (HD) patients. Pharmacological treatment of HD mouse model to reduce Setdb1 expression ameliorates the situation (Ryu et al., 2006). SNP located in *ATF7IP* locus and in its vicinity exhibited association with age of onset of Huntington's disease (Valcarcel-Ocete et al., 2015) and major depressive disease (Song et al., 2013). Schizophrenia patients also showed higher *SETDB1* expression and H3K9me2 levels (Chase et al., 2013). Mutations in *SETDB1* locus (Cukier et al., 2012) or copy number variation due to deletion (Xu et al., 2016) might be involved in genetic causes of autism. Interestingly, nicotine treatment in neuronal cell lowered H3K9me2 levels and HKMT expression, including *Setdb1* (Chase and Sharma, 2013). In the case of HIV, hSETDB1 methylates the viral transactivator Tat and represses the HIV-1 promoter (Van Duyne et al., 2008).

The involvement of *SETDB1* in neoplasia was already predicted earlier due to the fact that this locus shows a great number of recurrent translocations (Harte et al., 1999). SETDB1 was found recurrently upregulated in non-small cell lung cancer cell lines (Watanabe et al., 2008; Lafuente-Sanchis et al., 2015) and amplified in patient tissues, where it was associated with advanced pathological stages (Inoue et al., 2015). Depletion of SETDB1 impaired proliferation and anchorage-independent growth of these cell lines, through inhibition of WNT signaling (Sun et al., 2015). Similar results were observed for other types of human lung cancer cell lines and patients (Rodriguez-Paredes et al., 2014). However, reduced SETDB1 expression was also observed as an enhancer of metastatic ability, due to derepression of actin polymerization (Wu et al., 2014). Some compounds were shown to decrease hSETDB1 levels in lung cancer cells (Lee and Kim, 2013; Noh et al., 2014), also without affecting other HKMT (Na et al., 2016).

In human hepatocellular carcinoma patient tissues, amplification of *SETDB1* and progressive higher levels of expression were observed, what correlated with lower survival (Wong et al., 2016). This also occurred in liver cancer cell lines, where SETDB1 dimethylates mutant p53 and protects it against degradation, regulating cancer cell growth (Fei et al., 2015a). *SETDB1* was also identified as a melanoma susceptibility locus (Macgregor et al., 2011), which is recurrently amplified in melanoma samples (Ceol et al., 2011) and is correlated with higher proliferation rate (Kostaki et al., 2014). A role of *SETDB1* is also observed in tissues and/or cell lines of human glioma (Spyropoulou et al., 2014), prostatic cancer (Sun et al., 2014), breast cancer (Zhang et al.,

2014; Regina et al., 2016) and colon carcinoma (Olcina et al., 2016), malignant pleural mesothelioma (Kang et al., 2016). Interestingly, hSETDB1 represses MLV in some carcinoma cells, with deposition of H3K9me3 at the LTR region (Wang and Goff, 2017). Due to its involvement in several diseases SETDB1 has been considered a promising therapeutic target, with some potential inhibitors already under study (Karanth et al., 2017).

Regarding ATF7IP, it was found strongly expressed in stomach, lung and breast cancerous tissues. Besides that, telomere-positive cancer cells showed high expression of ATF7IP and Sp1 (Liu et al., 2009). SNP in *ATF7IP* locus or on its vicinity showed association with a predictor for risk of prostate cancer (Jin et al., 2013), testicular germ cell tumor susceptibility (Turnbull et al., 2010; Kanetsky et al., 2011; Karlsson et al., 2013) and its aggressiveness (Lessel et al., 2012). Besides that, long-term culture of small cell ovarian carcinoma cells shows deletion of the region containing *ATF7IP* (Otte et al., 2012). Interestingly, in B-progenitor acute lymphoblastic leukemia, a translocation led to an in-frame fusion protein between *ATF7IP* and *PDGFRB* (Kobayashi et al., 2014) causing interleukin-independent proliferation (Ishibashi et al., 2016).

1.12. Concluding remarks and open questions

In summary, SETDB1 makes use of the methyltransferase ability from its SET domain to catalyze the H3K9me marks, especially the trimethylation, with the purpose of repressing retrotransposon activity and gene transcription (Kang, 2015). Even though ERV repression has been well characterized (Chapter 1.7 Repression of transposable elements), studies proving SETDB1 direct involvement in regulation of gene promoters are still lacking (Chapter 1.6 Silencing of euchromatic genes). For instance, this protein was linked to the regulation of developmental genes in mESC by inferences from genome-wide profiles, with no further curation of the data or at least not to a great extent. Regardless of that, SETDB1 acts in multiple contexts and for each of these specific tasks SETDB1 interacts with distinct set of proteins to become involved in different mechanisms of silencing. Some controversial studies suggested that ATF7IP is capable of functioning as a cofactor for SETDB1 catalytic activity, either to enhance the process in mouse (Wang et al., 2003) or to slow it down in humans (Basavapathruni et al., 2016). A drawback of those studies is that most of the observations come from in vitro systems and may not properly represent what happens in vivo. Either way, the fact that ATF7IP works as a cofactor suggests that the interaction of both proteins may happen in the majority of the contexts. Thus, several mechanisms of action for SETDB1 were revealed so far and happen in the most various contexts (Figure 1.3). However, the involvement of ATF7IP in most of these cases is not yet known. Considering all that, further studies characterizing the cooperation between SETDB1 and ATF7IP are required to elucidate the contexts in which they occur and the mechanisms by which they act.
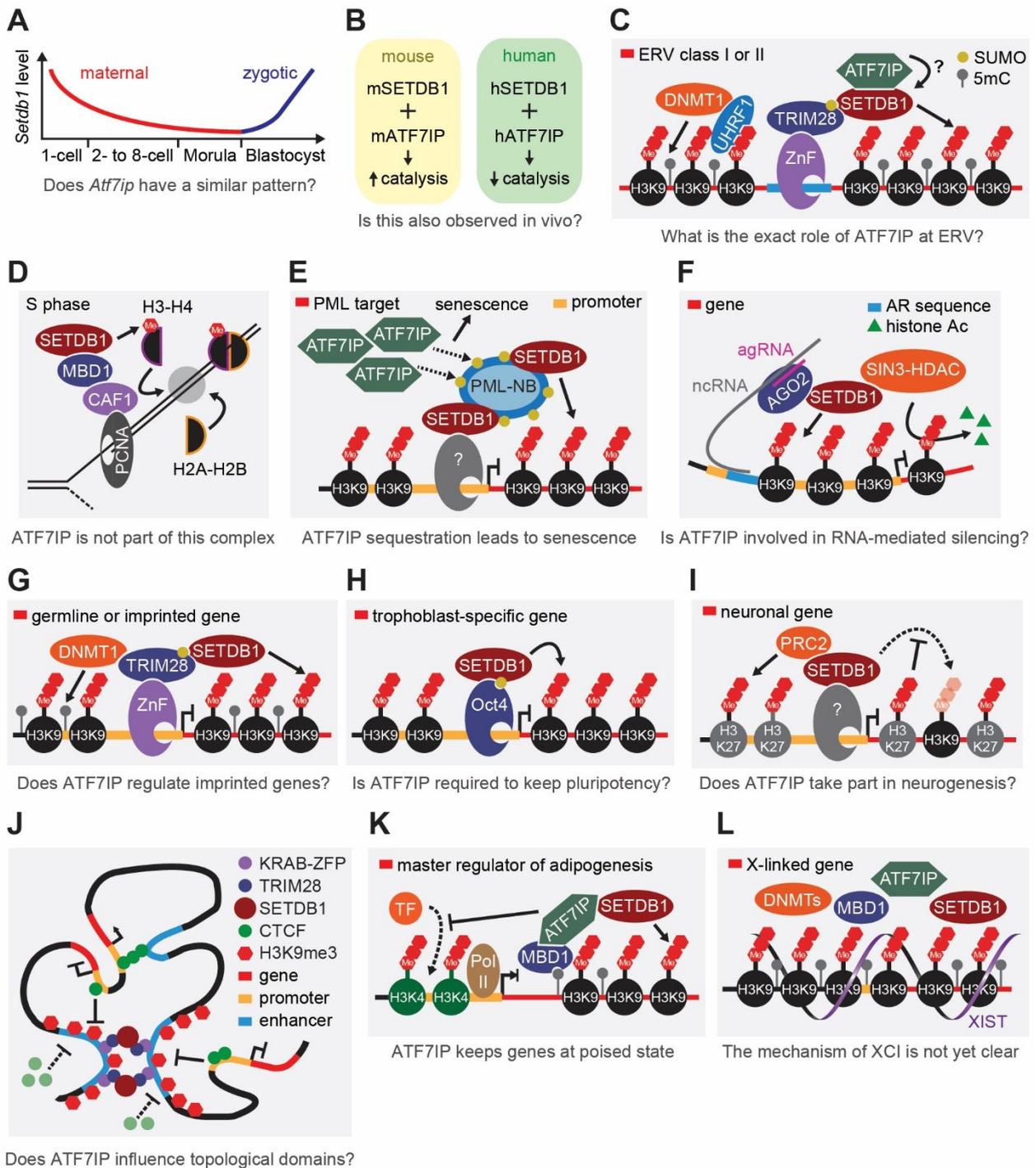
Figure 1.3 | Models for SETDB1 mechanism of action, its roles in different contexts and open questions.
(A) Pattern of *Setdb1* expression during early embryonic development. How does *Atf7ip* expression behave at these stages? (B) In the presence of mATF7IP, mSETDB1 shows higher catalytic activity towards H3K9me3 deposition in vitro, whereas, hSETDB1 activity is hindered in the presence of hATF7IP. Are these observations similar in a cellular system and in vivo? (C) In general, for TE silencing SUMOylated TRIM28 is targeted to ERV classes I and II by KRAB-ZFP and recruits SETDB1, which deposits H3K9me3 to block the binding of lineage specific transcription factors. ATF7IP is required for establishment of H3K9me3 and proper silencing of some of those ERV. (D) During cell division, CAF1-MBD1 complex recruits SETDB1 to the replication fork, where it methylates H3 before its incorporation onto DNA for the maintenance of repression. This process may not be linked to ATF7IP, as it is not part of the complex. (E) The maintenance of the PML-NB structure and the silencing of PML targets depend on SETDB1 and its methyltransferase activity. ATF7IP is not required as a structural component of PML-NB, but its sequestration to this complex due to SUMO binding induces senescence. (F) SETDB1 involvement in RNA-mediated silencing is also described. Essentially, agRNA-bound AGO2 binds ncRNA transcribed from promoters with the androgen receptor and recruits SETDB1 to tether SIN3A-HDAC1/2 complex to the neighbouring chromatin for

silencing. (G) KRAB-ZFP-dependent TRIM28 targeting is responsible for recruiting DNMTs to keep the levels of DNAme at imprinted and germline genes and for recruiting SETDB1-dependent H3K9me3. (H) Oct4 targets SUMOylated SETDB1 to trophoblast lineage-specific genes to restrict potential and keep pluripotent state. (I) SETDB1 recruits PRC2 for deposition of H3K27me3 at neuronal genes to hinder differentiation. The H3K27me3 mark inhibits SETDB1 activity. (J) SETDB1 is recruited by TRIM28 to KRAB-ZFP target sites to prevent excessive CTCF binding and regulate large topological chromatin domains which are neuron-specific. (K) Master regulators of adipogenesis have lineage-specific DNAme downstream of their promoter in progenitor cells and the active H3K4me3 mark upstream. MBD1 binds this mark and recruits ATF7IP, which, in turn, recruits SETDB1 to keep a poised bivalent state. H3K9me3 deposition prevents binding of an early adipogenesis transcription factor. (L) MBD1-SETDB1-ATF7IP complex act in concert with DNMT1 to perform XCI. However, the exact mechanism is not yet known. Regulation of several other developmental genes by SETDB1 has been suggested, however, it was not yet deeply investigated. Whether ATF7IP takes part in SETDB1 regulation of developmental genes is also unclear.

2. OBJECTIVES

The main goal of this work was to further characterize the function of ATF7IP as a cofactor of SETDB1 and how their interaction influences the methyltransferase activity of the latter in a cellular context representing the early developmental stage in mouse. The motivation to study the mechanisms of SETDB1-dependent repression were the indications of its role in regulating development. Besides that, its relationship with cancer and disease turns it into an attractive topic to gather knowledge which may increase opportunities for the development of future therapeutic strategies. The choice of the early development was due to the implications of this protein in maintenance of pluripotency and fate decisions. As mouse embryonic stem cells closely resemble this developmental stage, they were adopted as model system. To make this general goal feasible to address, the approach was to subset it in three more specific aims as follows:

- Characterize the roles of ATF7IP in embryonic stem cells and contrast it with what has been described for SETDB1. For that, the idea was to study the phenotype rising from the deletion of the gene coding for this protein by investigating the genome-wide transcriptional changes and the impact of its loss on differentiation. To compare it with the phenotype of SETDB1 depletion, publicly available datasets were reanalyzed.

- Determine the functional targets of the complex formed by SETDB1 and ATF7IP in embryonic stem cells. Functional targets were defined as genomic elements which are physically bound by this complex and have their epigenetic landscape primarily regulated by the activity of this complex to induce transcriptional repression. To this end, it was crucial to discover the genome-wide occupancy of both proteins as well as to figure out the consequences of ATF7IP loss to the genome-wide enrichment of different repressive and active histone 3 lysine 9 modifications.

- Identify the context where the SETDB1-ATF7IP complex exerts epigenetic and transcriptional regulation to get some hints in the mechanisms of action. In order to achieve that, the plan was to assess for the occurrence of genetic interaction between *Atf7ip* and other genes which may be acting in concert by using a silencing reporter assay. The genes analyzed would be based on the identification of putative protein partners of ATF7IP and other proteins involved in heterochromatin formation.

3. RESULTS

3.1. Generation of *Atf7ip* knockout mESC

The aim to understand the implications of ATF7IP on SETDB1 methyltranferase activity during early stages of embryonic development was first approached by reverse genetics. The *Atf7ip* locus was targeted for knockout by the CRISPR/Cas system, using a pair of sgRNA (Figure 3.1A). This system induces DNA double strand breaks that are later on repaired often by the non-homologous end-joining (NHEJ) repair machinery (Popp and Maquat, 2016) and leads to nucleic acid insertions or deletions at the damaged site. If the open reading frame changes, a premature stop codon rises from the frameshift mutated transcript to generate a truncated version of the protein. However, this type of mutation signals to the nonsense-mediated mRNA decay pathway to degrade these transcripts before they have the chance to be translated. In independent experiments, three *Atf7ip* KO mESC lines were generated, together with their respective control lines (Figure 3.1B). Isogenic population containing the same type of mutation were obtained by picking single clones and expanding them.

Mutants were primarily screened by PCR using a pair of primers flanking the target site (Figure 3.1C). A second step for the screening was to verify the absence of protein (Figure 3.1C). The identification of ATF7IP by western blot showed an interesting pattern of several bands with higher molecular weight than the predicted one. To make sure that those were not unspecific signals and to validate the commercial antibody (Ab) used, *Atf7ip* was depleted in mESC by two different short hairpin RNA (shRNA) separately. The shRNA expression vectors were packaged inside lentiviral particles used for delivery by transduction. In this way, cells integrated the shRNA-expressing construct and were permanently expressing it to keep a constant knockdown. Nuclear protein extracts probed with this Ab in western blot showed reduction of the signal of all the bands detected, proving its specificity (Figure 3.1D). Three control and three knockout clones showing complete loss of the ATF7IP were obtained (Figure 3.1E). The sequence of the locus in the mutants was verified by Sanger sequencing. For the first knockout cell line, several PCR products were cloned individually and sequenced to cover the sequence of both alleles (Figure 3.1F). Indeed, a nonsense mutation was produced, leading to loss of ATF7IP expression.

Figure 3.1 | Generation of *Atf7ip* knockout mESC.
(A) Region of the exon 2 of Atf7ip targeted by the sgRNA pair (purple). The PAM sequence recognized by the Cas9 is indicated (blue), along with the sites where double strand breaks are generated (blue arrows). (B) Experimental design for knocking out *Atf7ip* in mESC. Three control clones and three mutant ones were obtained. (C) UV-exposed agarose gel where the genotyping PCR products were separated to identify control and knockout clones. The first lane is the DNA ladder and is followed by a control sample and 14 knockout candidates. Below it is the digital image of the membranes used for blotting the nuclear extracts of the candidates after separation by SDS-PAGE. The membranes were probed with Ab against different ATF7IP and SUZ12 was used as loading control. Not all clones showed deletion of the target region even though the protein was absent. (D) Similar to the lower panel in C, but for the 6 selected clones alone. (E) Sanger sequencing of the cloned PCR products spanning the *Atf7ip* target site of the first mutant clone (ko25-1). Each clone represents one separate PCR product. The newly introduced premature termination codon is indicated in pink (F) Digital image of the membranes used for blotting the SDS-PAGE-separated

nuclear extracts of *Atf7ip*-depleted mESC after shRNA transduction. The membranes were probed with Ab against ATF7IP and exposed shortly and for a longer time. SUZ12 was used as loading control.

3.2. Loss of ATF7IP does not lead to changes in global H3K9me levels

Although loss of *Setdb1* leads to a severe phenotype, which is lethal for mESC (Chapter 1.5.a Embryonic stages), deletion of *Atf7ip* was not enough to lead to cell death in this system. Growth curves carried out by the laboratory technician Alexander Nuber, whose work supported this project, showed that these mutant cell lines behave similarly to control mESC in terms of growth (Nuber, Diploma thesis, 2017). Furthermore, cell cycle analysis via BrdU incorporation in newly synthesized DNA combined with staining of total DNA by propidium iodide, revealed no differences compared to control cells regarding progression through the distinct cell cycle phases.

In several cell types ATF7IP displays nuclear localization (Chapter 1.9.b Expression and localization), what is consistent with an interplay with SETDB1, a protein also present in the nucleus (Chapter 1.4.b Expression and localization patterns). Analysis of protein extracts from nuclear and cytoplasmic fractions indicated that the majority of ATF7IP peptides are retained in the nucleus of mESC (Figure 3.2), even though a portion still remains in the cytoplasm. Intriguingly, it seems that the post-translationally modified versions of this protein localize preferentially to the nuclear fraction. Whether one requires the other for the proper nuclear localization in mESC is still unclear. However, this is an unlikely situation as lack of nuclear SETDB1 would lead to cell death and contradicts the fact that the cells survive *Atf7ip* deletion.
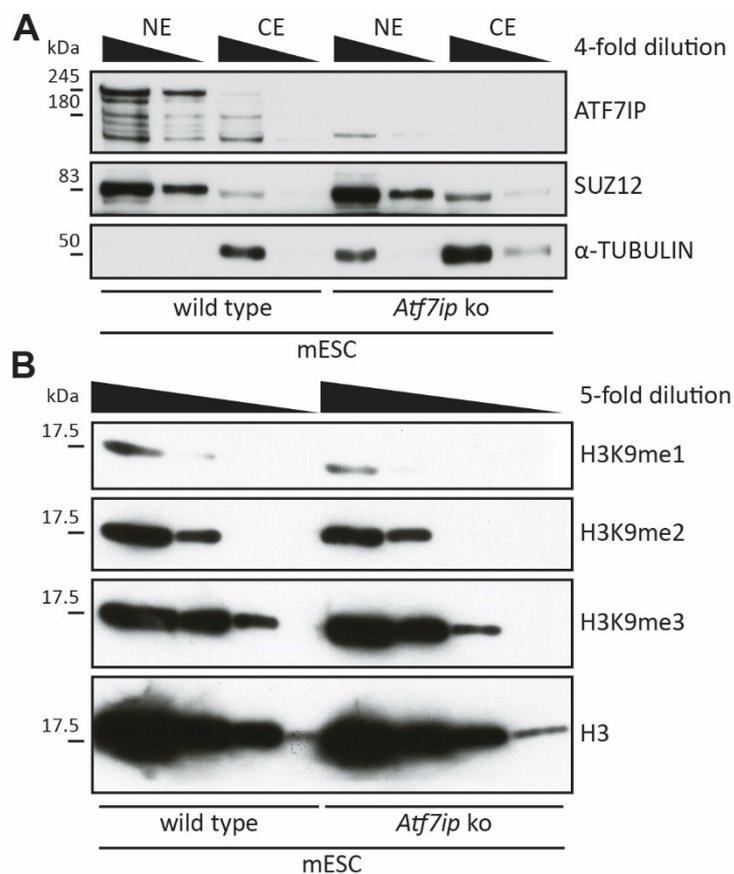
Figure 3.2 | Loss of ATF7IP does not lead to changes in global H3K9me levels.
(A) Digitalization of the membranes used for blotting the protein extracts of nuclear and cytoplasmic fractions of wild type and *Atf7ip* ko mESC after being resolved in SDS-PAGE. Samples were loaded in 4-fold dilutions as indicated on top. Membranes were probed with Ab against ATF7IP. Anti-SUZ12 and anti-α-TUBULIN Ab were used as loading control for nuclear and cytoplasmic extracts, respectively. (B) Digital image of the membrane used for blotting the acid extracts of nuclear histones from wild type and *Atf7ip* ko mESC after SDS-PAGE separation. Membranes were probed with Ab specifically recognizing H3K9 epitopes containing each of the modifications. H3 was used as loading control. Samples were loaded in 5-fold dilutions as indicated.

As SETDB1 methyltransferase activity was shown to be influenced by ATF7IP in vitro (Chapter 1.10 Regulation of SETDB1 catalytic activity by ATF7IP) the first step towards understanding the impact of the latter on the enzymatic activity of the former was to characterize the total levels of the different H3K9 marks. Histones were purified from wild type and *Atf7ip* knockout mESC by acid extraction. The levels of histone H3 and of the three methylation states of its lysine 9 were

quantified by Western blot using antibodies with high specificity for those marks and which did not exhibit cross reactivity to other marks (Hayashi-Takanaka et al., 2011). The total amount of histone H3 was unchanged between both samples. The same was true for the three H3K9me marks, indicating that if there is any alteration of SETDB1 activity due ATF7IP loss it does not largely affect global deposition of those modifications. This is consistent with the fact of SETDB1 not being regarded as responsible for the bulk H3K9me3 (Chapter 1.5.a Embryonic stages).

3.3. Transposable elements lose repression following ATF7IP deletion

In mESC the maintenance of repressed stated at ERV classes I and II is strictly dependent on SETDB1 enzymatic activity (Chapter 1.7 Repression of transposable elements). ATF7IP was found between the proteins which might assist SETDB1 in this task, as depletion by knockdown also led to upregulation of similar TE families (Chapter 1.10 Regulation of SETDB1 catalytic activity by ATF7IP). To test whether stable *Atf7ip* knockout cell lines lose proper control of transcriptional repression at those regions, TE transcript levels were quantified by reverse transcription quantitative PCR (RT-qPCR). Primers were designed based on consensus sequences for each ERV family, in order to amplify the majority of its elements. In consonance with previous findings, stable *Atf7ip* knockout cell lines showed impaired silencing of ERV class II, specially of MusD elements (Figure 3.3A). However, class I ERV did not become derepressed. Class III family MERVL did not show impairment as expected, considering that SETDB1 has no activity towards those elements.
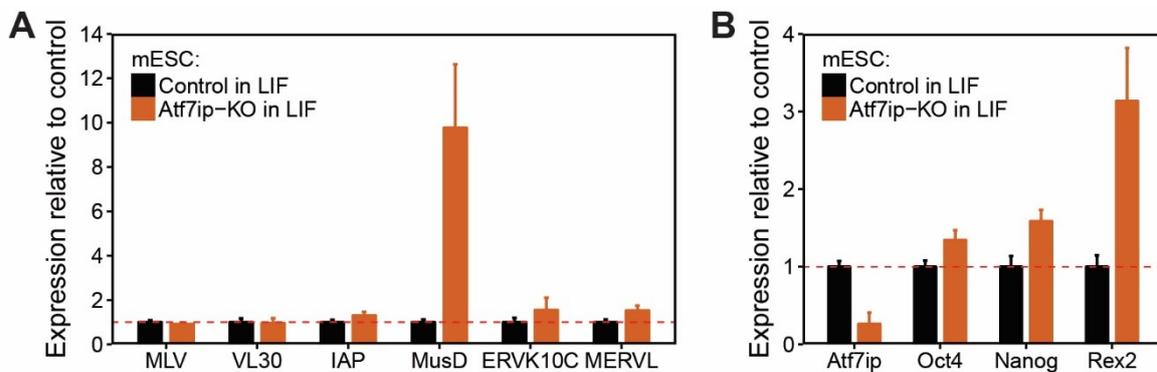
Figure 3.3 | Transposable elements lose repression following ATF7IP deletion.
(A) Expression levels of different TE families, members of ERV classes I (MLV and VL30), II (IAP, MusD and ERVK10C) and III (MERVL), in *Atf7ip* knockout mESC lines relative to control samples. For each replicate, expression was normalized by that of *Gapdh* and *Hprt*. The red dotted line indicates the level correspondent to no changes. (B) Similar to A, but for *Atf7ip* and a set of marker genes for embryonic stem cell pluripotent state.

To make sure mutant mESC did not lose the characteristic pluripotent state the expression levels of embryonic stem cell marker genes were assessed. The mutants showed comparable levels of three marker genes tested, except for one of them which exhibited rather higher transcriptional levels (Figure 3.3B). Besides that, the colonies display a round-shaped morphology, typical of this cell type (data not shown). Thus, the pluripotent state seems unaltered

in mESC lacking ATF7IP, differently from what is seen for SETB1 loss, where mESC that survive shortly the depletion tend to differentiate (Chapter 1.5.a Embryonic stages). Reduction of *Atf7ip* transcripts was also confirmed and is probably due to activity of the mRNA decay pathway.

3.4. Silencing kinetics of an exogenous IAPEz reporter is impaired by ATF7IP deletion

An internal fragment of the gag region from the IAPEz clade of TE containing around 160 bp is the minimal fragment required for recruitment of H3K9me3-dependent silencing through the TRIM28-SETDB1 pathway (Sadic et al., 2015). To understand the implications of ATF7IP on the silencing mechanism which acts upon this sequence, a reporter assay for exogenous TE silencing was carried out as previously described. Basically, this assay consists in delivering two reporter sequences by lentiviral transduction. The vector containing the enhanced green fluorescent protein (EGFP) controls for the virus titer (control), indicating the total amount of cells that are infected in the experiment. The second one is similar to the first, but the internal gag fragment is placed nearby the promoter (test) and is used for measuring the amount of cells able to use this fragment as a signal to repress transcription (Figure 3.4A). Each of these reporters are introduced in parallel in two batches of cells. Finally, the amount of EGFP-positive (EGFP+) cells in the test batch is measured by flow cytometry and is compared to the total EGFP+ cells captured in the control batch, giving the percentage of cells infected with the test reporter which remains EGFP+ (Figure 3.4F - second transduction step).

First of all, the results should be normalized by the total quantity of infective particles from each reporter. For that, the viral titer of both control and test viruses were measured (Figure 3.4B) in HeLa cells, where neither the control nor the test reporter sequences can be repressed. After that, two standardization procedures were carried out. In the first one, different viral titers were used to infect mESC to find out the minimum titer that leads to robust results (Figure 3.4C). In the second, a curve for the efficiency of infection of different starting amount of cells was obtained to identify the minimum cell number that needs to be seeded for getting replicable results (Figure 3.4D). To test the requirement of ATF7IP for the silencing of this gag sequence, knockout mESC lines were submitted to this reporter assay. While the control cells are able to rapidly silence the newly integrated gag reporter after lentiviral transduction, the mutant cells failed to do so and displayed a much slower kinetics of repression (Figure 3.4E).

To preclude the interference of possible variability between infection rates from control and knockout cell lines, the reporter assay was carried out on cells where *Atf7ip* was freshly knocked out. For that purpose, sgRNA targeting different portions of the Atf7ip locus were expressed separately in a mESC line constitutively expressing the Cas9 (Figure 3.4F). After that, those cells were used in the IAPEz-gag fragment reporter assay (Figure 3.4G). Several of the sgRNA led to impairment of silencing to levels similar to the one seen for the positive control *Daxx*, which was previously linked to this repression (Sadic et al., 2015). As the ATF7IP paralog ATF7IP2 was able to partially rescue MBD1-dependent repression (Chapter 1.10 Regulation of SETDB1 catalytic

activity by ATF7IP), its relevance for the gag fragment silencing was also addressed. However, no alteration of the repression was observed using sgRNA targeting different regions of this gene. Thus, ATF7IP, but not ATF7IP2, is involved in the SETDB1-dependent repression of exogenous TE coming from the IAPEz subfamily.
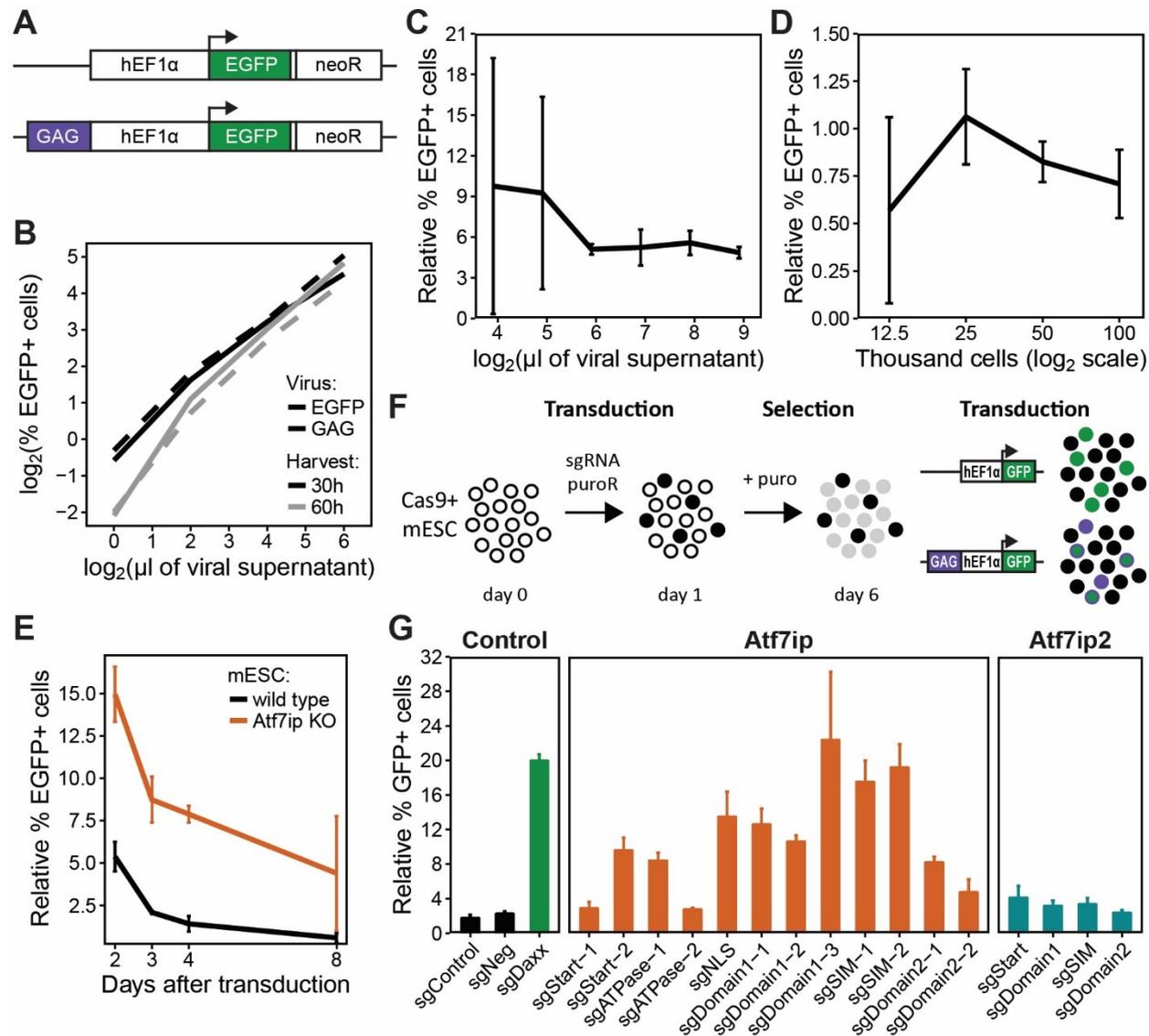
Figure 3.4 | Silencing kinetics of an exogenous IAPEz reporter is impaired by ATF7IP deletion.
(A) Diagram illustrating the control and test constructs for the IAPEz-gag fragment reporter assay, measured by EGFP expression driven by a promoter with strong activity in mESC. (B) Titration in HeLa cells of the viruses carrying either the control or the test construct. During viral production, two harvests were performed: one at 30 h and another at 60 h after co-transfection of HEK293T cells with packaging vectors and the reporter constructs. (C) Effect of the amount of viruses used for mESC transduction on the replicability of the experiment. (D) Effect of the amount of starting cell number on the robustness of the results. (E) Relative percentage of remaining EGFP+ cells after the gag fragment reporter assay in wild type and *Atf7ip* KO cells. (F) Experimental design for acutely knocking out *Atf7ip* in mESC. Cells constitutively expressing Cas9 were transduced with lentiviruses carrying a sgRNA expression construct. After antibiotics selection, and culture expansion, cells were divided in two batches to be transduced separately with the control and test reporter sequences. (G) Relative percentage of remaining EGFP+ cells after the gag fragment reporter assay in Cas9+ mESC transduced with sgRNA targeting different portions of either *Atf7ip* or *Atf7ip2*. An empty vector and a vector expressing a sgRNA targeting a non-related sequence were used as negative controls. A sgRNA targeting *Daxx* was used as positive control.

3.5. ATF7IP is not essential for *in vitro* differentiation in mESC

As the pluripotent state was kept in mESC lines not expressing ATF7IP, the ability to differentiate into the three germ layers was assessed. It could be the case that this protein is dispensable for the steady state of mESC, despite being required for the dynamic changes which take place during differentiation. To test that, mutant and control cells were triggered to form embryoid bodies (EB) by removing the leukemia inhibitor factor (LIF) from the medium and proceeding with the hanging drop method. With that, cells are stimulated to differentiate towards cell lineages from any of the three germ layers. Transcript levels were measured by RT-qPCR at the initial time point before differentiation was triggered and at days 5 and 10 after differentiation. Similarly to what was observed in control mESC, *Atf7ip* ko cells displayed gradual decrease of the pluripotency marker genes over time of EB formation (Figure 3.5A). Even though the process was slower than in the control samples, expression in mutants reached levels comparable to control at day 10. Regarding mesoderm-related genes, their expression in mutants also followed the same trend seen in the reference samples (Figure 3.5B). Of note, the gene *Nodal* is already expressed in the pluripotent state, because of its involvement in early embryonic development. Finally, the late differentiation genes implicated in either endoderm or ectoderm development showed no changes between knockout and control cells (Figure 3.5C). Hence, the dynamic changes of stage-specific gene expression are independent of the expression of *Atf7ip*.

Not only transcription of genes but also of TE is targeted for differential regulation during development (Chapter 1.7 Repression of transposable elements). Their activity is directly involved in controlling tissue-specific regulatory networks, as they may act as transcriptional enhancers (Xie et al., 2013; Garcia-Perez et al., 2016). To verify whether loss of ATF7IP would impair the regulation of specific ERV families, the changes in their transcription was followed along the EB differentiation. In normal situation, members of all three ERV classes remain not altered during the onset of differentiation (Figure 3.5D). This picture changes at later stages, when their expression gets enhanced, with some of them showing stronger activation like the IAP elements. On the other hand, mutant mESC exhibit the opposite behavior, as ERV which are initially upregulated during mESC stage tend to get repressed, while the transcription of the others is reactivated to a lower extent or does not show any alteration. This indicates that TE repression becomes independent of ATF7IP at later stages of differentiation. Besides that, the alteration of the epigenetic landscape at those regions might signal for their repression in differentiated cells by the respective mechanism governing at that stage.

Interestingly, *Atf7ip* expression is gradually stimulated upon differentiation (Figure 3.5E), indicating some importance during differentiation, even though mutant cells are capable of forming late stages EB. Indeed, presence of ATF7IP is correlated with the maintenance of proper levels of expression at least at the imprinted gene *H19*, as *Atf7ip* ko mESC showed stronger activation of transcription at this locus during differentiation into EB (Figure 3.5F).
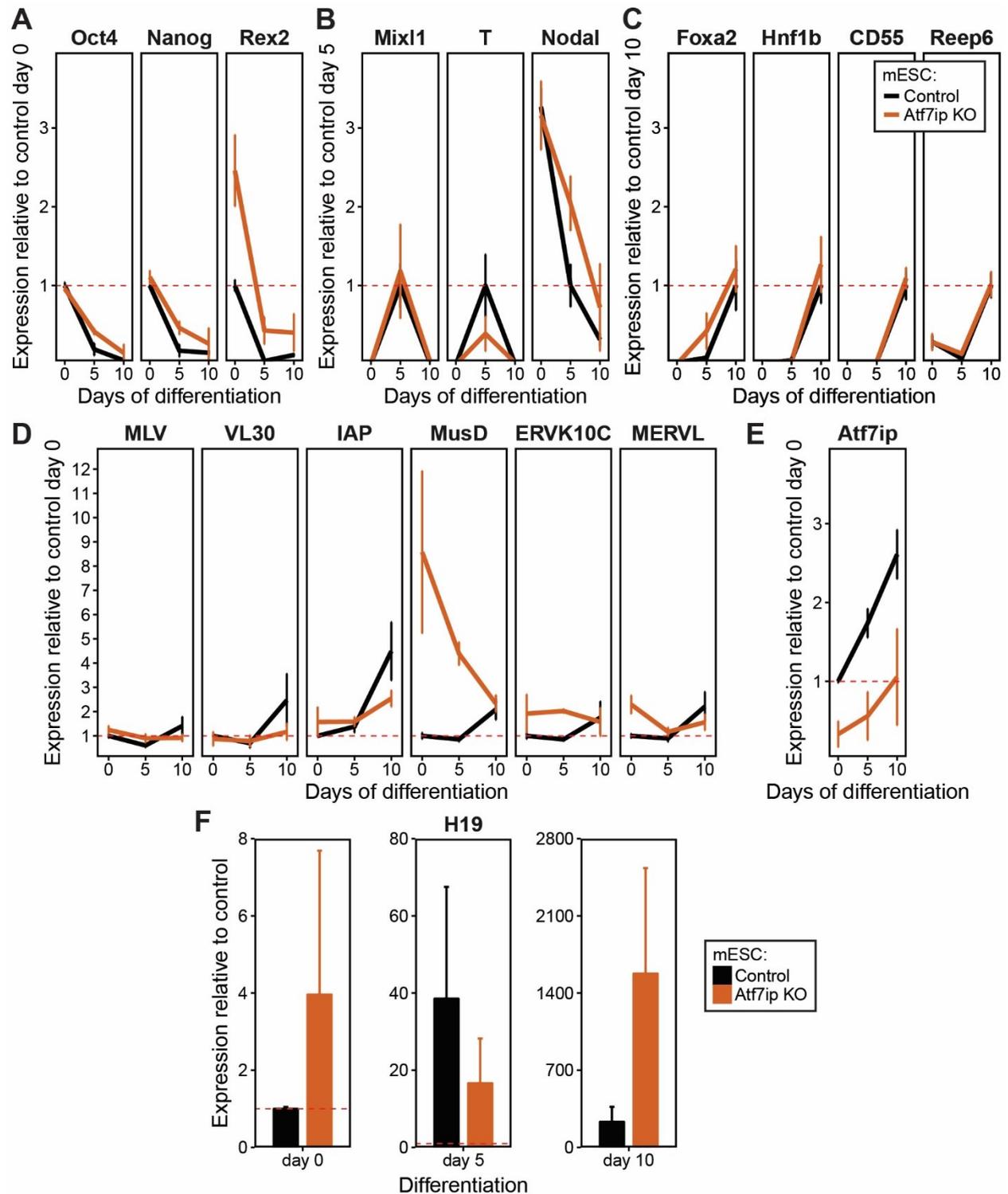
Figure 3.5 | ATF7IP is not essential for *in vitro* differentiation in mESC.
(A) Expression levels of different marker genes for pluripotency during the time course of differentiation into EB, in *Atf7ip* knockout mESC lines and control samples relative to the latter at day 0, before differentiation was triggered. For each replicate, expression was normalized by that of *Gapdh* and *Hprt*. The red dotted line indicates the level correspondent to no changes. (B) Same as in A, but the expression of genes marker for the mesoderm state were analyzed and normalized to control samples at day 5, when cells reach this state. (C) Same as in A, but for endoderm marker genes and an ectoderm marker (Reep6) normalized to control samples at day 10, when cells reach later stages of differentiation. (D) Same as in A, but for distinct ERV from classes I (MLV and VL30), II (IAP, MusD and ERVK10C) and III (MERVL). (E) Same as in A, but levels of *Atf7ip* expression were analyzed instead. (F) Same as in A, but for the imprinted gene *H19*. Note that the scales are different between each of the days of differentiation, as the levels of expression become much higher than the previous time point.

3.6. Developmental genes are misregulated in *Atf7ip*-depleted mESC

The derepression of some ERV families following *Atf7ip* depletion in mESC as measured by RT-qPCR (Chapter 3.3 Transposable elements lose repression following ATF7IP deletion) indicated the possibility of global changes in gene expression. To unravel the genome-wide changes in transcription, rRNA-depleted total RNA isolated from 3 biological replicates of control and *Atf7ip* KO mESC were quantified by RNA high-throughput sequencing (RNA-Seq). From the 228 significantly misregulated genes around half showed increased expression in mutants (Figure 3.6A). Comparison to published RNA-Seq datasets for *Setdb1* KO mESC and its respective control cell line (Karimi et al., 2011) indicated a more severe phenotype when SETDB1 is lost (Figure 3.6B), consistent with the fact that these cells do no survive long (Chapter 1.5.a Embryonic stages). However, several of the genes showing expression changes in *Atf7ip* KO cells overlapped with those misregulated after SETDB1 loss (Figure 3.6C). The up- and downregulated gene sets in both *Atf7ip*- and *Setdb1*-depleted cells were enriched for biological process gene ontology (GO) terms related to development and differentiation, especially within the downregulated ones (Figure 3.6D). Hence, development-related genes may be regulated by a joint effort of both proteins. Unexpectedly, these results indicate that the presence of the complex SETDB1-ATF7IP might be involved in hindering further repression of those genes, as they undergo downregulation when either protein is depleted in mESC.
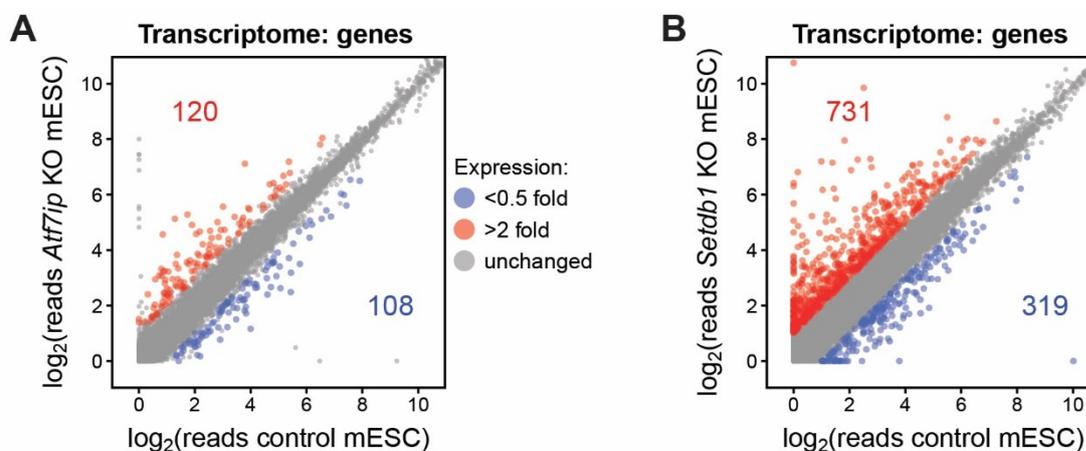
Figure 3.6 | Developmental genes are misregulated in *Atf7ip*-depleted mESC.
(A) Correlation of the expression levels (given by RPKM) of all protein-coding UCSC genes between *Atf7ip* KO mESC and control cells. Average of three biological replicates is plotted. Up- and down-regulated genes with more than two-fold changes are highlighted. (B) As in A, but for *Setdb1* KO mESC versus its control cell line. The raw data was previously published (Karimi et al., 2011) and only one replicate was available. (C) Scaled Venn diagram representing the overlap of upregulated (in the left) and downregulated genes (in the right) across *Atf7ip* and *Setdb1* knockout mESC. (D) GO term enrichment analysis for biological processes of up- and downregulated genes (upper and lower panels respectively) in either *Atf7ip* or *Setdb1* mutant cells. Shown is the probability based on the p-value. The number on top of each bar indicates the amount of genes fitting the specific GO term for each mutant.

C

Setdb1 KO / Atf7ip KO

680 | 51 | 69 — Up

Setdb1 KO / Atf7ip KO

300 | 19 | 89 — Down

D

GO term enrichment

mESC:
- Setdb1 KO
- Atf7ip KO

genes up:
- meiosis I — 9 / 17
- meiotic cell cycle — 12 / 31
- synapsis — 6 / 12
- male gamete generation — 13 / 37
- piRNA metabolic process — 0 / 8

genes down:
- tissue development — 30 / 49
- cell surface receptor signaling pathway — 34 / 57
- anatomical structure morphogenesis — 35 / 72
- anatomical structure development — 52 / 117
- single−multicellular organism process — 54 / 124
- cellular response to growth factor stimulus — 15 / 22
- multicellular organism development — 48 / 106
- single−organism developmental process — 52 / 122
- developmental process — 52 / 122
- animal organ morphogenesis — 19 / 32
- embryo development — 20 / 35
- system development — 43 / 92
- cellular developmental process — 41 / 87
- regulation of multicellular organismal development — 25 / 55
- anatomical structure formation in morphogenesis — 18 / 35
- positive regulation of developmental process — 20 / 40
- regulation of multicellular organismal process — 30 / 73

$\log_{10}$(p−value)

3.7. TE families derepressed overlap in mESC lacking ATF7IP or SETDB1

The global transcription profiles of *Atf7ip* KO mESC were generated using paired-end reads, which results in a better mappability of the reads to the reference genome as it increases the number of uniquely assigned reads (Corley et al., 2017). With these data in hand, the differential expression of all annotated repeats defined in the University of California Santa Cruz (UCSC) genome browser were assessed. Several TE families exhibited expression in *Atf7ip*-deleted cells higher than 2-fold over the control samples (Figure 3.7A). Similarly to what was seen for transcription at coding genes (Chapter 3.6 Developmental genes are misregulated in *Atf7ip*-depleted mESC), the misregulation of repetitive regions in *Setdb1* mutant cells was stronger, with five times as many families being derepressed (Figure 3.7B). Nonetheless, the majority of the upregulated families following *Atf7ip* deletion overlaps with the misregulated ones in *Setdb1*-deleted mESC (Figure 3.7C). Consistent with the RT-qPCR data (Chapter 3.3 Transposable elements lose repression following ATF7IP deletion), the classes of TE identified as derepressed in *Atf7ip* or *Setdb1* mutants or in both at the same time are mainly ERV from classes I and II (Figure 3.7D).
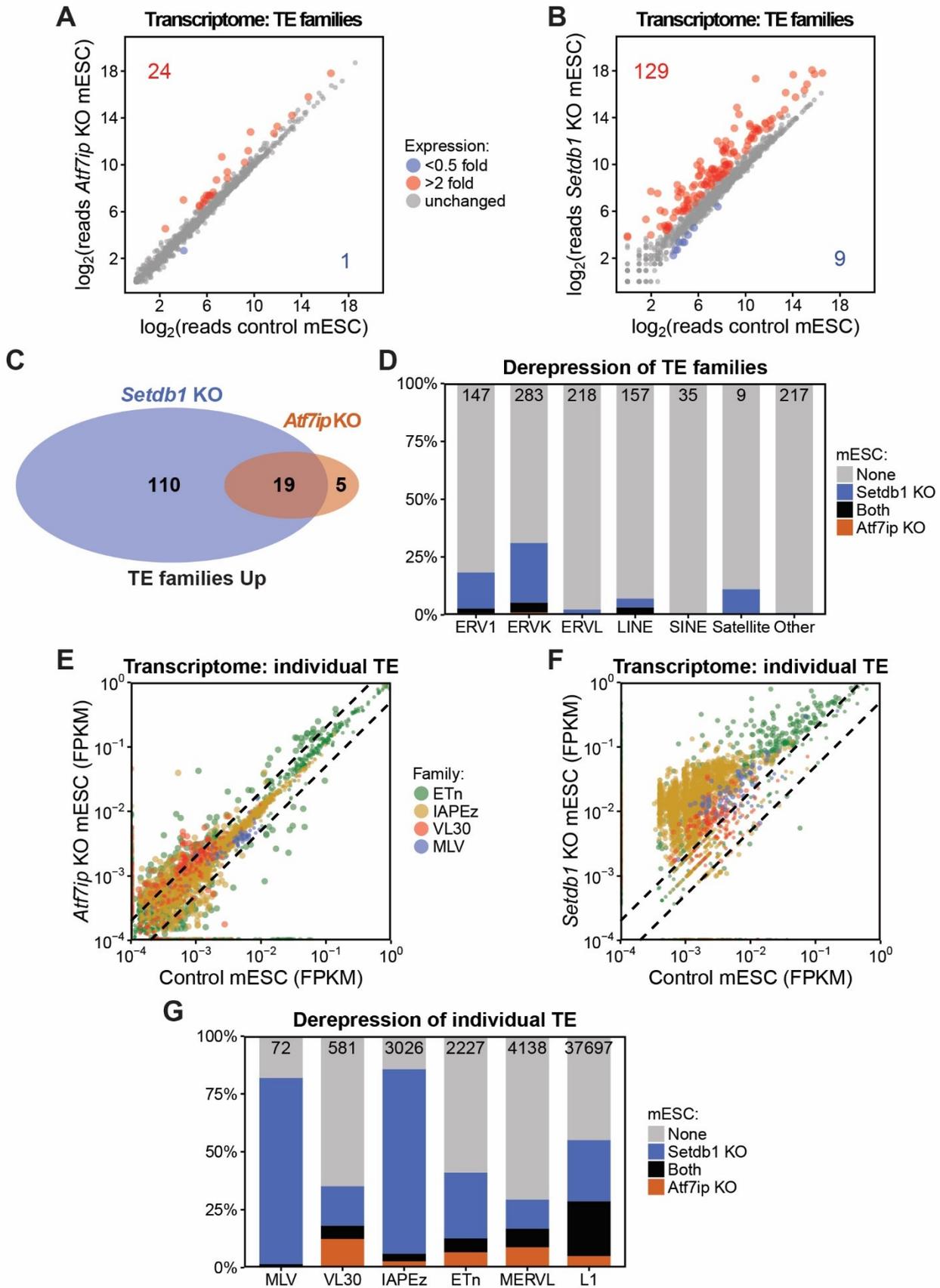
Figure 3.7 | TE families derepressed overlap in mESC lacking ATF7IP or SETDB1.
(A) Correlation of the expression levels (given by RPKM) of all families of TE as defined by the UCSC database between *Atf7ip* KO mESC and control cells. Average of three biological replicates is plotted. Up- and down-regulated families with expression higher or lower than two-fold over the control are highlighted. (B) As in A, but for *Setdb1* KO mESC versus its control cell line. The raw data was previously published (Karimi et al., 2011) and only one replicate

was available. (C) Scaled Venn diagram representing the overlap of derepressed TE families across *Atf7ip* and *Setdb1* knockout mESC. (D) Percentage of families of different TE classes which become upregulated in either *Atf7ip* or *Setdb1* mutants. The total number of families composing each of the classes depicted are shown on the top of their respective bar. (E) Correlation of the expression levels (given by RPKM) of all individual elements belonging to the indicated TE subfamilies as annotated in the UCSC database between *Atf7ip* KO and control mESC. Average expression of the three biological replicates is plotted. Dashed lines are drawn at the threshold of two-fold (upper) or half-fold (lower) expression. (F) As in E, but for *Setdb1* KO mESC versus its control cell line using the sole replicates available in the published raw data (Karimi et al., 2011). (G) Percentage of individual elements which are upregulated in either *Atf7ip* or *Setdb1* KO cells. The total amount of loci forming each of the subfamilies depicted are shown on the top of their respective bar.

The reads were then mapped to individual elements composing each of the subfamilies MLV, VL30, ETn, IAPEz, MERVL and LINE-L1 (Figure 3.7E). While in SETDB1-depleted cells most IAPEz and ETn elements were derepressed (Figure 3.7F), in ATF7IP-depleted cells they were only in part upregulated (Figure 3.7E). On the other hand, individual MLV which also undergo strong reactivation in *Setdb1* mutants, did not show any increase in expression in *Atf7ip* mutants. This can be easily noticed when a distribution of derepressed TE are plotted according to the subfamilies they belong (Figure 3.7G). Interestingly, when looking at individual repeat loci, some LINE-L1 and MERVL elements also showed upregulation in both mutant types. Whether this is a direct effect of The SETDB1-ATF7IP complex requires the investigation of their binding sites, which is going to be discussed later, even though SETDB1 activity on those sites was not appreciated in previous works (Chapter 1.7 Repression of transposable elements).

3.8. mESC partially adapt to the loss of ATF7IP

The mild changes in the transcriptome of *Atf7ip* KO mESC in comparison to control cells can also be observed by principal component analysis (PCA). The first component, based on differences which cover 50 % of the variance between all the samples, distinguishes well the mutant cell lines from the control group (Figure 3.8A). On the other hand, the second component, which represents 30 % of the variance, is not consistent in grouping the mutant genotype away from the control. A reason for this might be that the several passages required for isolation and expansion of clones originating from single colonies gives enough time for the cells to adapt to the knockout condition.

To test the hypothesis of adaptation to the knockout condition in *Atf7ip* mutants, transcription levels were measured by RT-qPCR in wild type mESC few days after deletion of the gene. This constitutes an acute knockout situation, where one can capture the very early responses. Expression of different subfamilies of ERV classes I, II and III were analyzed at days 0, 2 and 4 after knockout induction (Figure 3.8B). Knockout of *Dnmt1* was performed as a positive control, as it was shown previously that its deletion in mESC leads to upregulation of TE early after induction of knockout, while they become repressed again at later time points (Sharif et al., 2016). Consistently, similar derepression was observed in the present experiment.
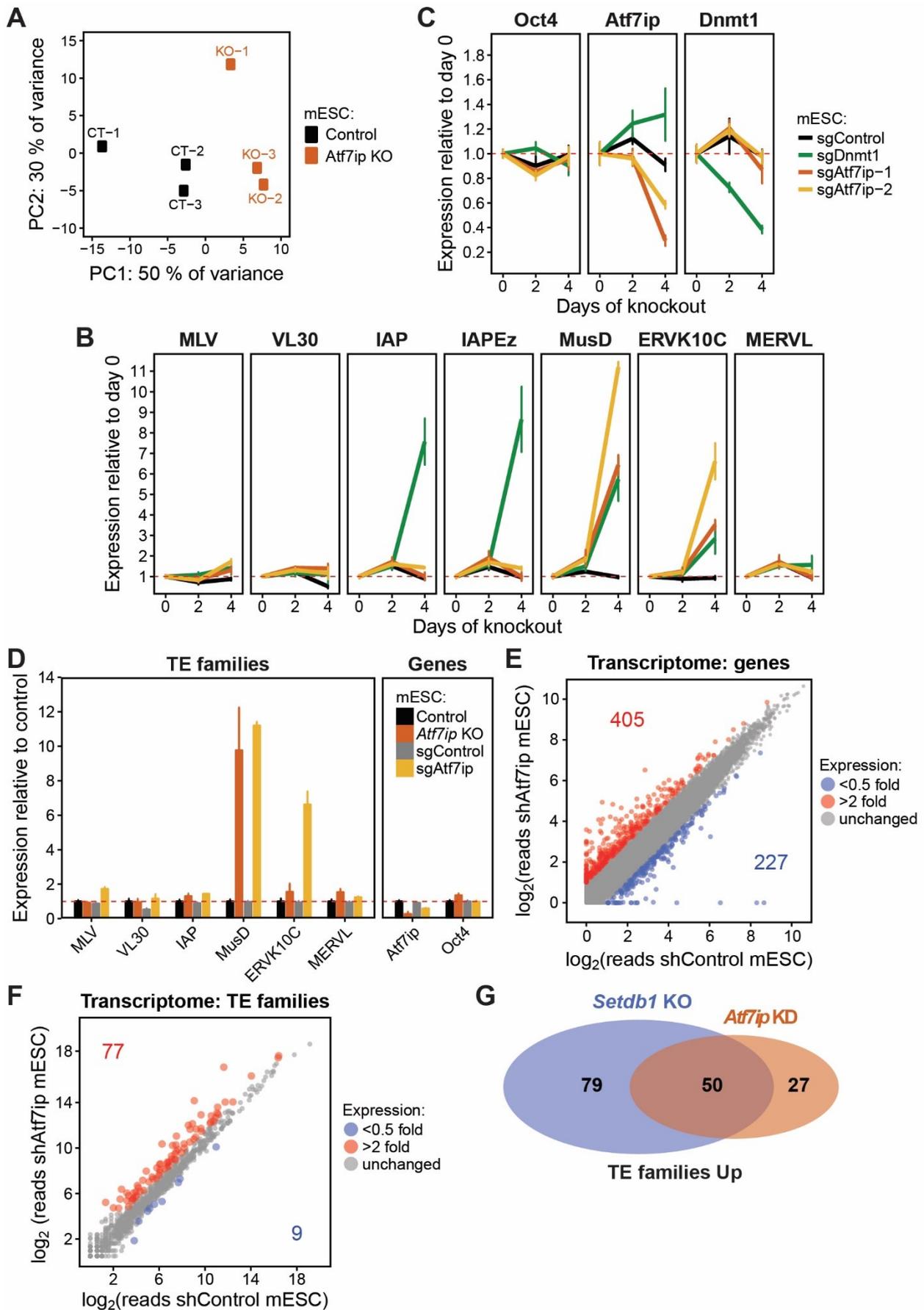
Figure 3.8 | mESC partially adapt to the loss of ATF7IP.
(A) PCA projection of control and *Atf7ip* knockout stable cell lines. Each dot represents the dataset originating from one sample. (B) Expression levels of different TE subfamilies during the time course of acute knockout induction via

CRISPR/Cas9, relative to non-transfected samples at day 0. *Dnmt1* was used as a positive control for reactivation of TE. For each replicate, expression was normalized by that of *Gapdh* and *Hprt*. The red dotted line marks the level where no changes are observed. (C) Same as in B, but for the expression of the genes targeted by sgRNA in the experiment in B. Expression of *Oct4* was verified as a means to assess whether the cells still kept the pluripotent state. (D) Re-plotting of the data shown in B for the sgControl and the sgAtf7ip with higher efficiency and for the data presented in (Figure 3.3), to compare side by side the differences between TE derepression in stable knockout cell lines and acutely deleted pool of cells. (E) Correlation of the expression levels (given by RPKM) of all protein-coding UCSC genes between mESC expressing shRNA to knockdown *Atf7ip* and cells expressing a control shRNA. Up- and down-regulated genes with more than two-fold changes are highlighted. The raw data was previously published (Yang et al., 2015) and only one replicate was available. (F) Same as in E, but for TE families. (G) Scaled Venn diagram depicting the overlap of derepressed TE families across *Atf7ip* knockdown and *Setdb1* knockout mESC.

Acute deletion of *Atf7ip* also led to upregulation of some TE subfamilies, namely MusD and ERVK10C, even though IAP elements did not show increased expression as seen for *Dnmt1* KO. Expression of the genes targeted for deletion were also verified and displayed reduction over time in the respective knockouts (Figure 3.8C). Levels of *Oct4* transcripts remained similar during the course of the experiment, indicating that cells were not affected by loss of pluripotency. Of note, deletion of *Dnmt1* led to gradual increase in *Atf7ip* expression levels over time (Figure 3.8D).

The impact in the whole transcriptome of an acute depletion of ATF7IP was investigated using readily available data previously published, where knockdown was achieved by shRNA expression (Yang et al., 2015). As expected, many more genes were dysregulalted (Figure 3.8E) when compared to the knockout cell lines (Figure 3.6A). Similar effect was observed for the number of derepressed TE families (Figure 3.8F), which was also higher than the amount of derepression in stable knockout cell lines (Figure 3.7A). When contrasted with the families upregulated in SETDB1 knockout mESC, a larger overlap was observed (Figure 3.8G). Thus, cells are able to cope with *Atf7ip* deletion in the long term and manage to partially contain the reactivation of several TE and genes. To achieve this adaptation, they probably take advantage of other silencing mechanisms which act in parallel, though independently of the pathway involving ATF7IP.

3.9. Loss of DNA methylation further enhances *Atf7ip* knockout phenotype

The DNAme pathway is known to act as a second layer of silencing over some TE families in mESC, specially of IAPEz, which display enhanced derepression when deletion of *Setdb1* is combined with that of *Dnmt1* (Chapter 1.7 Repression of transposable elements). Treatment of mESC with the two inhibitors (2i) of MEK and GSK3β maintains cells in a naive ground state by reducing global DNAme (Sim et al., 2017). Addition of vitamin C (vitC) to the growth medium promotes Tet activity of DNA demethylation enhancing loss of DNAme and inducing a blastocyst-like state with effects beyond those obtained with 2i (Blaschke et al., 2013). TE repression is ensured by SETDB1-dependent H3K9me3 and PcG-mediated H3K27me3 deposition after the intense loss of DNAme (Walter et al., 2016). To investigate the roles of SETDB1-ATF7IP complex while precluding the influence of silencing by DNAme, control and *Atf7ip* mutant mESC were

transferred to 2i medium complemented with vitC (2i+vitC) for induction of global DNAme reduction. Transcript levels were quantified by RT-qPCR at early and late time points.

After 14 days of culture in 2i+vitC *Oct4* expression showed some decrease in expression, while *Nanog* became upregulated (Figure 3.9A). Interestingly, expression changes for *Oct4* and *Rex2* over the course of the experiment were similar between control and mutant samples. However, *Nanog* transcriptional levels at the last time point, which were higher in control cells, exhibited an opposite behavior in cells lacking ATF7IP. As expected, some of the TE groups underwent activation during early time points, though they got repressed at the latest point, with the exception of MusD, which could not be reach down to the initial levels (Figure 3.9B). Besides MLV, all the TE families tested suffered enhanced derepression in the mutant when compared to the kinetics in the control samples. Contrasting to the reduction of expression at the latest time point in the control cells to levels similar to the initial ones, all TE reached the end of the experiment with transcriptional levels higher than the initial ones in the mutants. MusD elements were again the exception and showed opposite effects, with expression at the latest time point lower than the one measured before transferring the cells to 2i+vitC.
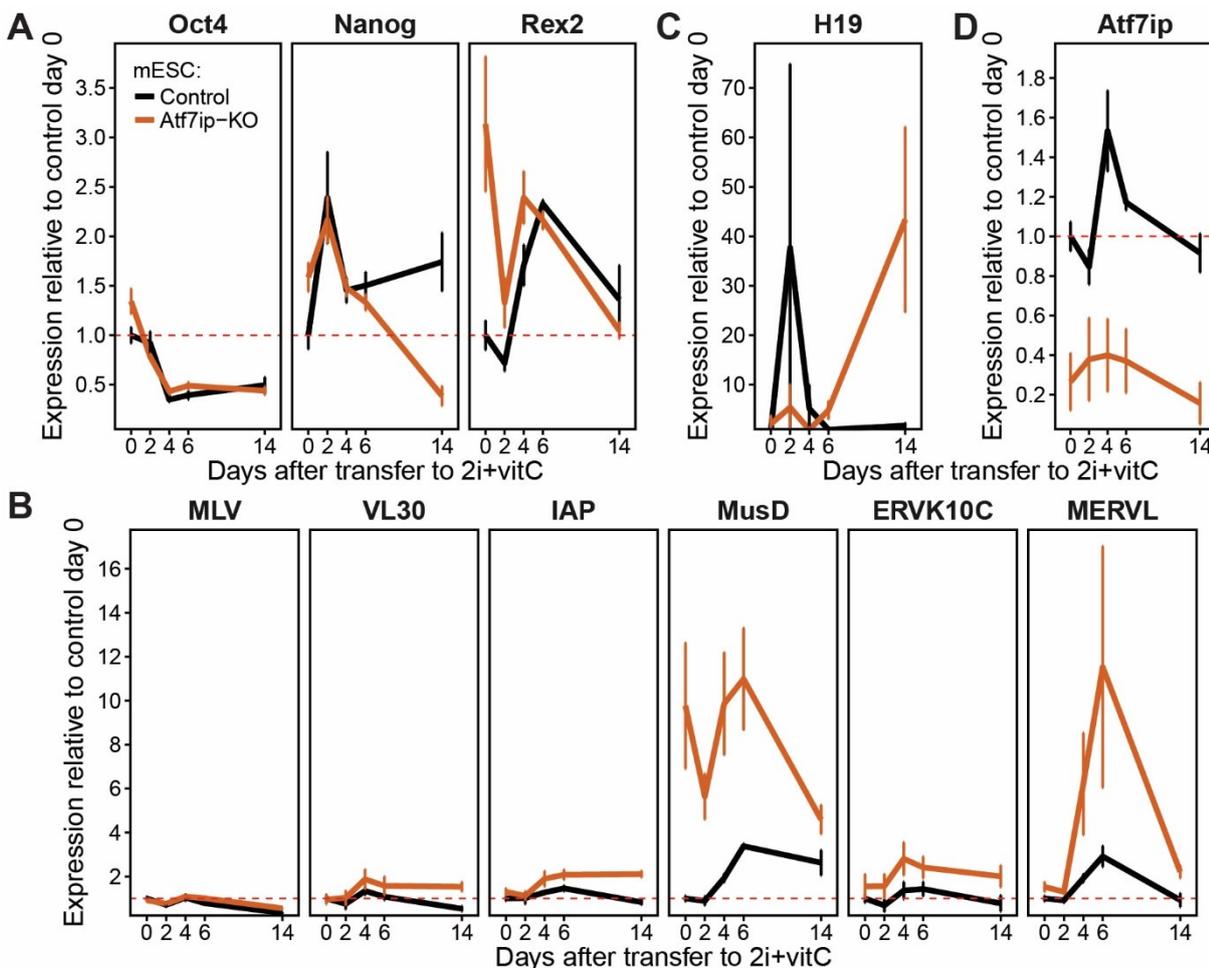
Figure 3.9 | Loss of DNA methylation further enhances *Atf7ip* knockout phenotype.
(A) Expression levels relative to control day 0 of different pluripotency marker genes during the time course after transferring *Atf7ip* knockout mESC lines and control samples to medium containing 2i complemented with vitamin C.

For each replicate, expression was normalized by the geometric mean of *Gapdh* and *Hprt*. The red dotted line indicates the level correspondent to no changes. (B) Same as in A, but for distinct ERV from classes I (MLV and VL30), II (IAP, MusD and ERVK10C) and III (MERVL). (C) Same as in A, but levels of expression of the imprinted gene *H19* were analyzed instead. (D) Same as in A, but for *Atf7ip*.

Regarding the imprinted gene *H19*, the long exposure to 2i+vitC medium led to strong and stable derepression of this gene, but only in the *Atf7ip* mutant mESC (Figure 3.9C). Of note, *Atf7ip* expression increases during the beginning of DNAme loss, probably as a positive feedback of a synergic pathway with the aim to counter the effects of DNA demethylation and prevent transcriptional derepression (Figure 3.9D). This result is consistent with the one observed right after deletion of *Dntm1* in mESC (Chapter 3.8 mESC partially adapt to the loss of ATF7IP), where *Atf7ip* exhibited similar upregulation.

3.10. mESC are more prone to differentiate when lacking both ATF7IP and DNA methylation

As the DNAme mechanisms somehow lessen the effects of *Atf7ip* deletion in mESC over time, the impact in differentiation of ATF7IP deletion coupled to global loss of DNAme was investigated. For that, cells grown for several passages in 2i+vitC were used in EB formation assay. Transcription levels were measured by RT-qPCR at days 0, 5 and 10 after induction of differentiation. When compared to differentiation of primed cells grown in medium containing only LIF (Chapter 3.5 ATF7IP is not essential for *in vitro* differentiation in mESC), control cells with depleted DNAme took longer to partially lose pluripotency as measured by the expression of a few marker genes, even though *Oct4* expression did not reduce (Figure 3.10A). However, genes involved in mesoderm formation were already expressed at levels similar to the ones seen at day 5, when they were supposed to have the peak of transcription (Figure 3.10B). Besides that, these genes did not undergo differential regulation of expression over time during differentiation. Similar trend was observed for the endoderm gene *CD55* and the ectoderm marker *Reep6*, which were already expressed in stem cell stage (Figure 3.10C). On the other hand, the other two endoderm-related genes tested became upregulated only in the last time point analyzed, similar to what happens to the primed cells in LIF.

In the case of the *Atf7ip* KO cells, depletion of DNAme seemed to facilitate the differentiation into EB. In the first 5 days, transcription of pluripotency genes was already strongly downregulated (Figure 3.10A). Marker genes for endoderm, but not for ectoderm, were induced very early and at high levels (Figure 3.10C), even when compared to what is seen for control and mutant cells with normal DNAme levels in LIF medium (Figure 3.5C). Also unexpected was the reduction of their expression at later stage of EB differentiation, which was concomitant with the increase in transcription of some of the mesoderm genes later than normal (Figure 3.10B).

In control cells, which presented reactivation of TE at day 10 of EB formation when coming from medium with LIF only (Figure 3.5D), depletion of DNAme led to an anticipation of this reactivation to day 5 (Figure 3.10D). In contrast to what happens with *Atf7ip* KO cells in the primed

state, where TE expression is either reduced or unchanged, mutant cells in 2i+vitC-induced naïve state lacking proper DNAme lost repressive control over these regions and most of the ERV families tested became upregulated already in the first 5 days of differentiation. The only exception were the MusD elements, which rather showed some decrease in transcription.
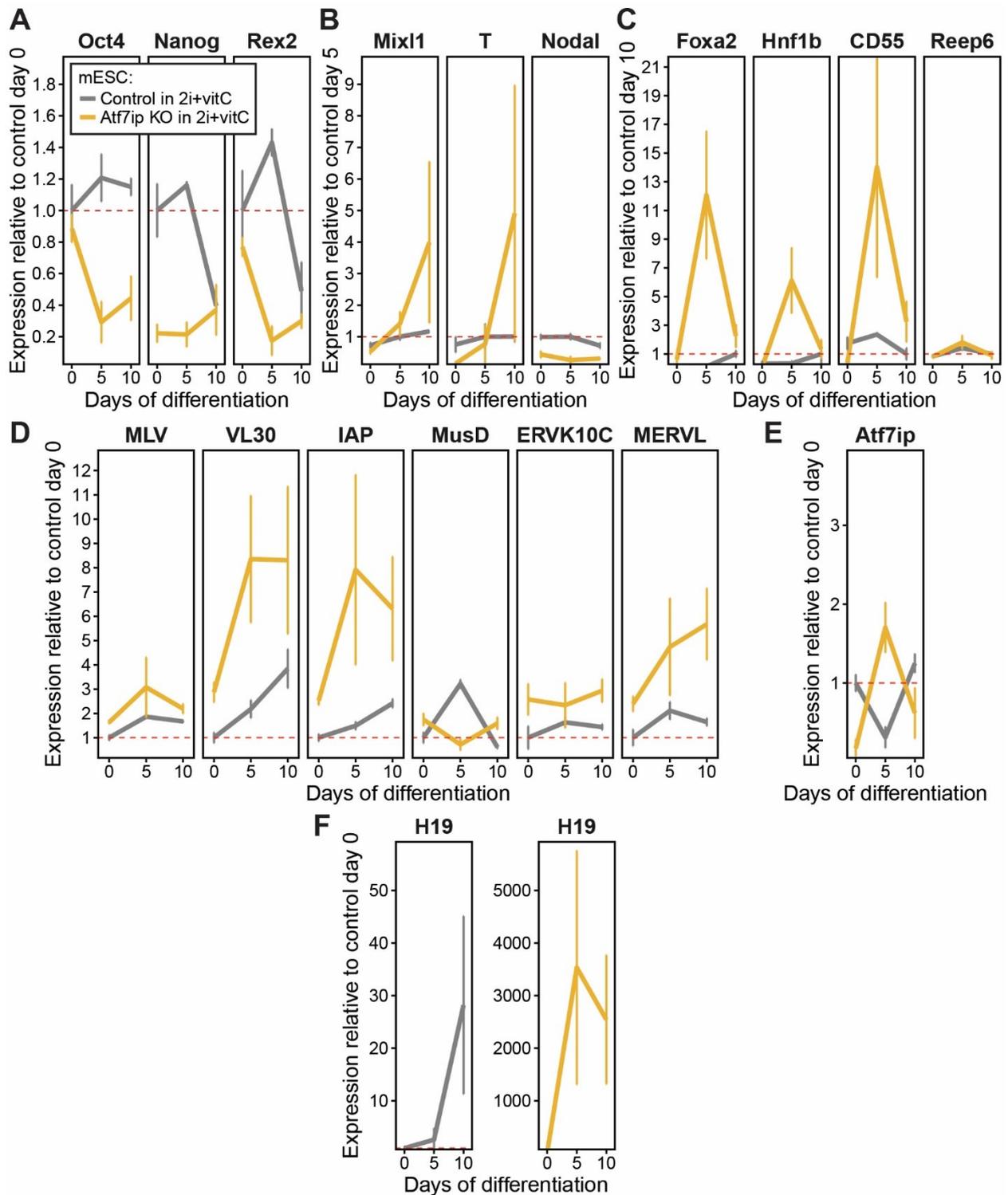
Figure 3.10 | mESC are more prone to differentiate when lacking both ATF7IP and DNA methylation.
(A) Expression levels of different genes involved in pluripotency during the time course of differentiation into EB in *Atf7ip* knockout mESC lines and control samples grown for several passages in medium containing 2i and vitamin C to reduce global DNAme levels. The values shown are relative to the control cells at day 0, before triggering

differentiation. In each replicate, expression was normalized to *Gapdh* and *Hprt*. The red dotted line indicates the level correspondent to no changes. (B) Same as in A, but the transcription of genes marker for the mesoderm state were assessed and normalized to control samples at day 5, when cells reach this state. (C) Same as in A, but for endoderm-related genes and an ectoderm marker (Reep6) normalized to control samples at day 10, when cells reach later stages of differentiation. (D) Same as in A, but for distinct ERV from classes I (MLV and VL30), II (IAP, MusD and ERVK10C) and III (MERVL). (E) Same as in A, but transcriptional levels of *Atf7ip* were analyzed. (F) Same as in A, but for the imprinted gene *H19*. Note the different scales between both plots, as the upregulation is much stronger in mutant samples.

Regarding *Atf7ip* expression, instead of gradually increasing over time it showed strong downregulation at day 5, while an opposite trend was observed in the knockout mutants (Figure 3.10E). The imprinted gene *H19* became upregulated, although not as strong as in the primed control cells containing DNAme (Figure 3.10F). Even though primed *Atf7ip* KO mESC showed gradual increase in expression of this gene achieving very high levels at day 10, removal of DNAme prior to induction of differentiation in EB enhanced this effect and led to an extremely strong upregulation already at day 5 Altogether, these results indicate that there is a crosstalk between DNAme and SETDB1-ATF7IP complex in regulating genes and TE expression changes during differentiation. Furthermore, ATF7IP seems to be required for proper transcriptional regulation in the context of development when DNAme is depleted in mESC, suggesting an important role for the SETDB1-ATF7IP complex during very early stages of embryonic development (pre-implantation embryo) marked by global erasure of DNAme (Atlasi and Stunnenberg, 2017).

3.11. ATF7IP is involved in SETDB1-dependent TE repression in primary MEF

The fact that SETDB1 knockout embryos do not survive and die very early during the pre-implantation period indicates its fundamental roles in embryonic development (Chapter 1.5.a Embryonic stages). As an important interactor and cofactor of SETDB1 catalytic activity, ATF7IP might also play a role during these early developmental stages as it is indicated by the impairment in transcription regulation in the absence of ATF7IP and DNAme (Chapter 3.10 mESC are more prone to differentiate when lacking both ATF7IP and DNA methylation). Some of the results obtained by the laboratory technician Alexander Nuber support the idea that ATF7IP functions with SETDB1 during early embryonic development. In summary, he showed that in crosses between heterozygous *Atf7ip* knockout adults, no homozygous mutants puppets were born. Furthermore, homozygous mutants could only be observed until blastocyst stage and they already did not reach Mendelian rate (Nuber, Diploma thesis, 2017) indicating a phenotype as early as in *Setdb1* KO embryos. Interestingly, these mutants failed to properly form the ICM.

All those observations suggest a role for the complex SETDB1-ATF7IP in the pluripotent stem cells forming the embryo. To investigate the possibility of this complex acting at the layer of cells giving support for the formation of the embryo proper, primary mouse embryonic fibroblasts (pMEF) isolated from homozygous embryos conditional knockout (cKO) for either *Atf7ip*

(*Atf7ip* p/p) or *Setdb1* (*Setdb1* p/p) were experimented. In these cells lines, the flox allele (p) can be efficiently deleted by transducing the cells with viruses carrying a cassette for Cre recombinase expression. Besides that, cells expressing this protein can be selected by an antibiotic resistance which is expressed in tandem with the Cre coding sequence due to the presence of an internal ribosome entry site (IRES). Deletion was then induced in these cKO cell by lentiviral transduction and expression was quantified at days 0, 6 and 12 by RT-qPCR.

The responses regarding transcriptional changes after induction of deletion were very similar between both genotypes (Figure 3.11A). Basically, TE from different ERV classes became gradually upregulated over the course of the experiment, with the exception of the ERV class I MLV and the class III MERVL. Derepression of IAP was only very mild compared to others. Upregulation of imprinted genes were mainly seen following *Atf7ip* deletion in the first 6 days after induction, however, in cells of both genotypes these genes downregulated in the last time point (Figure 3.11B).
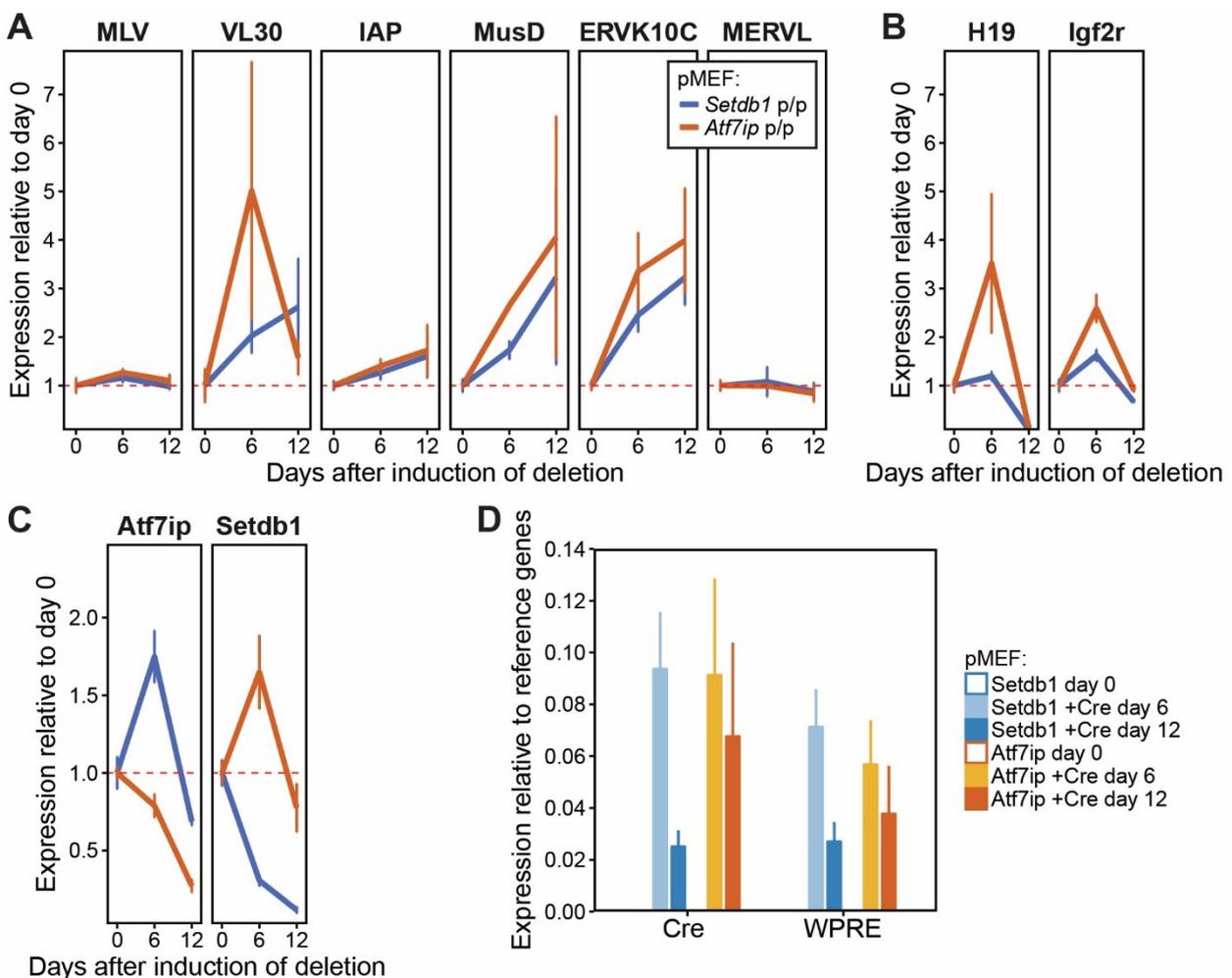
Figure 3.11 | ATF7IP is involved in SETDB1-dependent TE repression in primary MEF.
(A) Expression levels of distinct ERV from classes I (MLV and VL30), II (IAP, MusD and ERVK10C) and III (MERVL) during the time course after induction of *Atf7ip* or *Setdb1* deletion in the respective cKO pMEF relative to day 0, right before deletion. The flox alleles were deleted by expression of lentivirally-delivered Cre recombinase. Transduced cells were selected for the antibiotic resistance. Expression was normalized to that of *Gapdh* and *Hprt*. The red dotted

line indicates the level correspondent to no changes. (B) Same as in A, but for two imprinted genes. (C) Same as in A, but for the flox genes targeted for deletion in each cKO line. (D) Same as in A, but for the Cre recombinase and the puroR sequence relative to reference genes *Gapdh* and *Hprt*.

When transcription of the genes targeted for deletion was analyzed in both cKO pMEF during induction of deletion, a positive feedback between expression of both genes was observed in the first 6 days (Figure 3.11C). While transcripts of one gene were reduced due to deletion, the other became upregulated. Such a response was expected, as similar results were observed in HeLa cells (Timms et al., 2016) and in mESC (Thompson et al., 2015). Of note, the turnover of *Atf7ip* transcripts seem longer than for *Setdb1*, as 6 days after deletion they still accounted for around 80 % of the initial expression. Expression of Cre recombinase and the puromycin resistence (puroR) were confirmed (Figure 3.11D).

3.12. Generation of FLAG-ATF7IP knockin mESC

The interaction between SETDB1 and ATF7IP in mESC is already well defined (Chapter 1.10 Regulation of SETDB1 catalytic activity by ATF7IP). To confirm that this interaction is not impaired by expression of a tagged version of SETDB1 mESC, co-immunoprecipitation (Co-IP) was carried out using an Ab against the FLAG epitope and nuclear extracts from a FLAG-SETDB1 knockin (KI) mESC line previously generated (Fan, PhD Thesis, 2015). As expected, SETDB1-ATF7IP association could be captured in this cell line (Figure 3.12A). As SETDB1 activity is targeted towards histone H3, the functions of this complex SETDB1-ATF7IP on transcriptional regulation must happen in concert with its association to chromatin.

A way to define the genome-wide occupancy of chromatin interactors with high specificity is to take advantage of the FLAG tag, for which Ab with strong affinity and great specificity are available. As FLAG-SETDB1 KI cells were readily available, the first step to define the genomic binding of both proteins was to obtain a FLAG-ATF7IP KI mESC line as well. For that purpose, the CRISPR/Cas system was used to target the C-terminus of the *Atf7ip* (Figure 3.12B). During homology-directed repair (HDR) by the intrinsic cellular system, the FLAG-tag could be inserted in the target place. Two isogenic clones were isolated and expanded by Zeyang Wang, a PhD student from the same laboratory who helped in this specific task. Correct HDR was confirmed by PCR (Figure 3.12D) and Sanger sequencing of the products (Figure 3.12E). The translation of the tagged protein was confirmed by using nuclear protein extracts for Western blot (WB), probing it with a FLAG-Ab (Figure 3.12F).

Figure 3.12 | Generation of FLAG-ATF7IP knockin mESC.
(A) Digital image of the membrane used for blotting the FLAG Co-IP material of FLAG-SETDB1 KI mESC nuclear extracts after separation by SDS-PAGE. Membrane was probed with Ab against different epitopes and LAMIN B1 was used as a control for non-specific binding. (B) Sequence of the C-terminal of *Atf7ip* locus indicating the stop codon (pink). The sgRNA sequence is also shown (purple) together with the PAM sequence (blue). The PAM is recognized by the Cas9, which generate double strand breaks at the region indicated with blue arrows. (C) Diagram of the HDR template design and of the genotyping scheme. (D) UV-exposed agarose gel to distinguish between wild type and knockin clones. As the PCR product are quite similar in size, a restriction enzyme site was added to the HDR template. (E) Sanger sequencing of the PCR products of the *Atf7ip* locus from two independent clones isolated after CRISPR knockin. The FLAG-tag (green) was correctly recombined into the locus before the endogenous stop codon (pink). The PAM sequence (blue) was mutated (red) to avoid cleavage of the HDR template by the Cas9. (F) Digital image of the membrane used for blotting nuclear extracts of the two FLAG-ATF7IP KI mESC clones after separation by SDS-PAGE. Membrane was probed with Ab against different epitopes and SUZ12 was used as loading control. FLAG-SETDB1 KI cells nuclear extract was used as positive control for the FLAG-Ab.

3.13. SETDB1 and ATF7IP co-occupy gene promoters and transposable elements

With FLAG-tagged mESC lines at SETDB1 and ATF7IP loci in hand, the genomic binding of these two proteins was identified by chromatin immunoprecipitation (ChIP). At first, quantitative PCR was used to analyze the purified material (ChIP-qPCR) from Flag-*Atf7ip* KI mESC (Figure 3.13A) and Flag-*Setdb1* KI mESC (Figure 3.13B). As there was clear enrichment over the input material when compared to negative regions, chromatin immunoprecipitation followed by high throughput sequencing (ChIP-Seq) was carried out. Reads mapping to multiple locations were discarded and peaks close to the background noise from input sample were filtered out. Looking at all peaks ranked by the strength of their signal, three distinct clusters according to presence of either protein can be readily visualized (Figure 3.13C). The majority of the peaks belong to the cluster containing SETDB1 only. However, around one third of SETDB1 binding sites are co-bound by ATF7IP (Figure 3.13D). Interestingly, the strength of SETDB1 occupancy increases with that of ATF7IP, especially at sites with the highest occupancy (Figure 3.13E).

The three clusters were then investigated for the enrichment of biological process GO terms (Figure 3.13F). For this analysis, only peaks 2 kb away from an annotated gene TSS were included. Co-bound regions showed an enrichment for terms related to meiosis, cell cycle DNAme and neuronal-related processes. ATF7IP sole peaks were enriched for rRNA processing and transcription initiation. SETDB1 sole peaks, in turn, were associated with several terms related to development, especially of neurons. Using GREAT to assign peaks to genes contained in regulatory domains which overlap with the peak region (McLean et al., 2010), the majority of the sites co-bound by SETDB1 and ATF7IP were associated with two genes (Figure 3.13G). This indicates that the complex formed by these proteins may be involved in controlling regulatory domains involving multiple genes.

When the sole and common peaks were assessed for the genomic annotation of the regions containing the binding sites, almost half of the co-bound regions occurred at ERV (Figure 3.13H - upper panel). SETDB1 sole peaks were very abundant at intra- and intergenic regions, while ATF7IP is bound to many promoters without the former. Of note, around half of the SETDB1-bound promoters were also occupied by ATF7IP (Figure 3.13H - lower panel). Indeed, the closest SETDB1 peaks to TSS co-occurred with ATF7IP (Figure 3.13I). However, the majority of the peaks around the TSS were represented by ATF7IP sole peaks. Interestingly, common binding sites at ERV regions tended to occur closer to TSS, especially for class II elements (Figure 3.13J). For this analysis reads with multiple alignment were kept to avoid bias. Thus, the SETDB1-ATF7IP complex is frequently observed at ERV elements and close by gene promoters.

A ChIP: FLAG Ab

B ChIP: Flag-*Setdb1* mESC

D

SETDB1   ATF7IP

5355   1666   4384
(47%)   (15%)   (38%)

C

SETDB1   ATF7IP   INPUT

E Genome coverage

Peaks:
Atf7ip
Setdb1
Both

G Peak association with genes

25   379   1261   1

F GO Term: Biological process

Atf7ip only

ncRNA processing 42
rRNA processing 27
meiotic nuclear division 26
homologous chromosome segregation 13
meiosis I 19
histone mRNA metabolic process 6
DNA strand elongation 6
DNA methylation or demethylation 12
regulation of protein ubiquitination 22
DNA recombination 24
negative regulation of protein ubiquitination 11
oogenesis 12
DNA-templated transcription, initiation 12
positive regulation of stem cell proliferation 9

Setdb1 only

neuron projection morphogenesis 25
neural precursor cell proliferation 11
forebrain development 17
mesonephric tubule morphogenesis 6
cell fate commitment 13
muscle cell migration 6
mesenchyme development 12
ureteric bud development 7
nerve development 6
central nervous system neuron differentiation 11
kidney epithelium development 8
mesoderm development 7

Both

meiotic nuclear division 14
homologous chromosome segregation 7
spermatogenesis 18
meiosis I 8
learning 9
cell cycle process 27
DNA methylation or demethylation 7
gene silencing by RNA 7
neuron-neuron synaptic transmission 7
regulation of cAMP metabolic process 6
rhythmic process 10
regulation of neurotransmitter levels 7
multicellular organismal signaling 6
embryonic limb morphogenesis 6
insulin secretion 7

$-\log_{10}$(p-value)

H

ATF7IP only 4384
Both 1666
SETDB1 only 5355

Annotation:
Promoter
Intragenic
Intergenic
ERV
SINE/LINE
Others

Promoter 1117
Intragenic 2660
Intergenic 2304
ERV 3005
SINE/LINE 1621
Others 698

Factor:
ATF7IP only
Both
SETDB1 only

I Peak distribution: TSS

Overlap:
Atf7ip
Setdb1
Both

Distance to TSS

J Common peak distribution: ERV

Class:
ERV1
ERVK
ERVL

Distance to TSS

Figure 3.13 | SETDB1 and ATF7IP co-occupy gene promoters and transposable elements.
(A) Enrichment as percentage of input at positive (IAP and *Nnat*) and negative (*Ptch1* and *Tia1*) control regions obtained by ChIP-qPCR using sheared chromatin of nuclear extracts from Flag-*Atf7ip* KI mESC. Wild type mESC were used as mock sample. Error bars represent standard deviation between technical replicates. (B) Enrichment as percentage of input at positive (IAP and *H19*) and negative (*Ptch1* and *Tia1*) control regions obtained by ChIP-qPCR using sheared chromatin of nuclear extracts from Flag-*Setdb1* KI mESC. Beads not coated with Ab were used as mock sample. Error bars represent standard deviation between technical replicates. (C) Heatmap of binding sites obtained by ChIP-Seq using KI mESC lines clustered according to presence of SETDB1 or ATF7IP. The peaks were ordered decreasingly according to the strength of the signal. (D) Scaled Venn diagram representing the overlap between SETDB1 and ATF7IP binding sites. (E) Correlation between the genome coverage of ATF7IP and of SETDB1 of all binding sites normalized to input. (F) GO term enrichment analysis for biological processes of sole and common SETDB1/ATF7IP peaks. Shown is the probability based on the p-value. The number on top of each bar indicates the amount of genes fitting the specific GO term. (G) Number of common ATF7IP/SETDB1 peaks associated with putative regulation of 0, 1, 2, or 3 genes according to GREAT analysis. (H) Percentage of genomic annotations associated with sole and common SETDB1/ATF7IP peaks (upper panel). The lower panel shows a replot of the data from the upper panel to emphasize the percentage of sole or common peaks occurring at each genomic annotation. (I) Distribution of peak counts 2 kb around the TSS for sole and common peaks. (J) Same as in I, but for peaks occurring at regions containing annotation for different ERV classes in the range of 100 kb around the TSS. Reads with multiple alignment were included.

3.14. Promoters and ERV bound by SETDB1-ATF7IP respond differently to *Atf7ip* deletion

To associate the binding of SETDB1-ATF7IP complex to the functional activity of SETDB1 to deposit H3K9me3 mark, the enrichment for different H3K9 modifications were investigated in control mESC and compared to *Atf7ip* KO cell lines. SETDB1 binding sites enriched for H3K9me3 most often co-occurred with ATF7IP (Figure 3.14A). When the co-bound regions were analyzed for fluctuations in the enrichment for the distinct H3K9 modifications after *Atf7ip* deletion almost no peaks showed changes in H3K9me3 deposition (Figure 3.14B). However, while H3K9me2 tended to become increased ERV-associated peaks, promoter associated peaks displayed an opposite trend. H3K9ac decreased at ERV instead and increased at a few promoter regions. Interestingly, concomitant binding of SETDB1 and ATF7IP nearby TSS was associated with presence of H3K9ac and absence of the repressive H3K9me2/3 marks (Figure 3.14C). On the other hand, common peaks far away from the TSS were usually devoid of H3K9ac and were marked by H3K9me3. In general, no strong differences in behavior were observed between sole SETDB1 peaks and the common ones, except for H3K9me2, which became slightly more enriched in peaks far from an annotated TSS and H3K9ac, which was already low in control samples and became even lower (Figure 3.14D). Intriguingly, investigation of publicly available datasets (Karimi et al., 2011) showed that SETDB1 sole peaks did not undergo any changes in H3K9me3 levels when *Setdb1* was deleted in mESC, whereas the deposition of this mark was strongly impaired in mutant cells at common peaks.
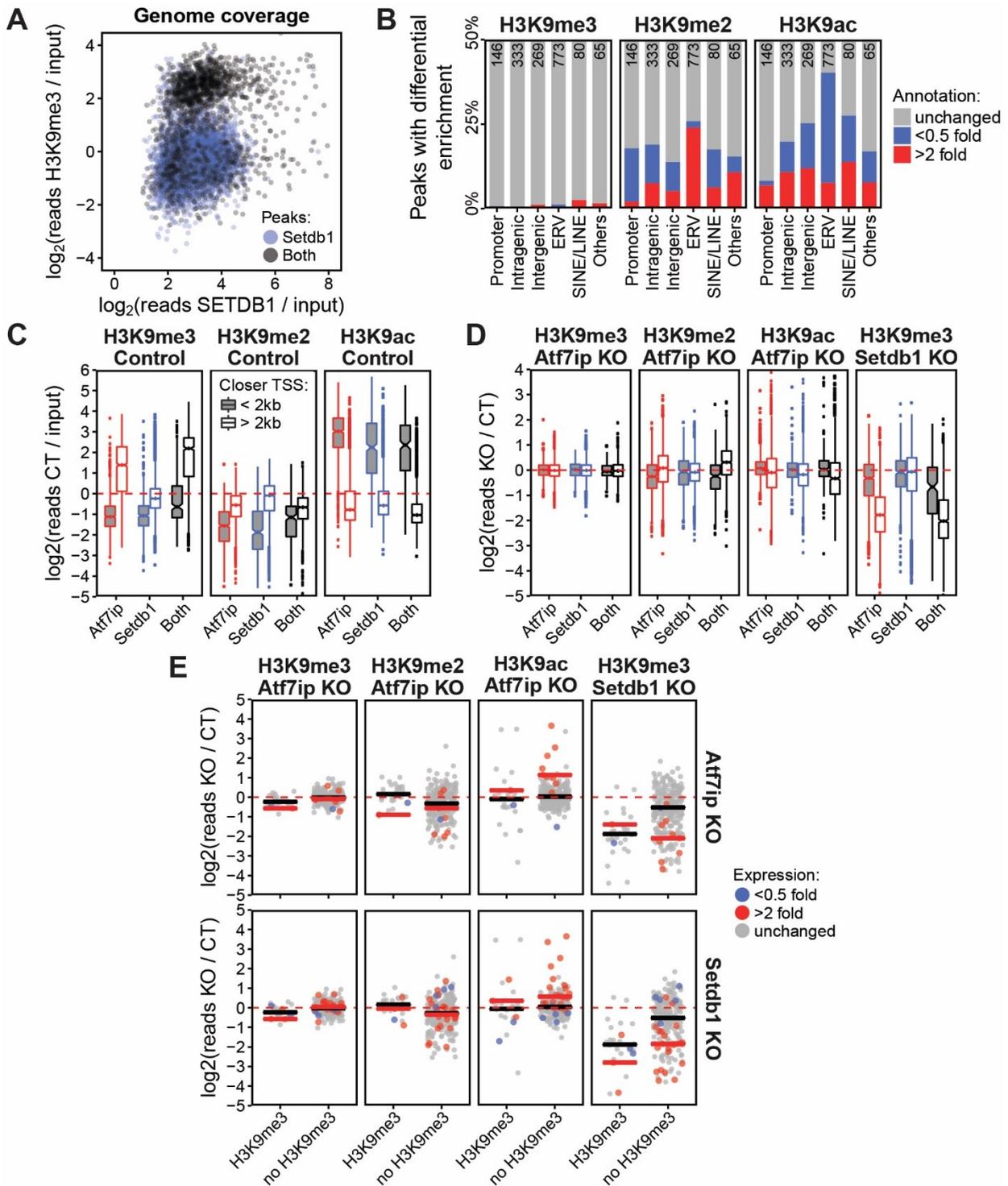
Figure 3.14 | Promoters and ERV bound by SETDB1-ATF7IP respond differently to *Atf7ip* deletion.
(A) Correlation between the genome coverage of SETDB1 and H3K9me3 at all SETDB1 binding sites normalized to input. The color indicates whether the peak is a sole peak or co-occurs with ATF7IP. (B) Percentage of common SETDB1/ATF7IP binding sites which show differential deposition in H3K9 modifications according to the genomic annotations associated with them. (C) Coverage normalized to input material for the distinct H3K9 modifications in control mESC at sole and common binding sites according to the distance from the closest annotated TSS. (D) Same as in C, but for the fold change in Atf7ip KO cells over control mESC. (E) Fold enrichment of H3K9 modifications in Atf7ip KO cells over control mESC for all common binding sites in the vicinity of annotated TSS according to the presence or absence of H3K9me3. The colors depict the expression changes in Atf7ip KO (upper panel) or Setdb1 KO mutants (lower panel). A black line is drawn at the median of all peaks, while the red line represents the median of only the peaks nearby an upregulated gene. The red dotted line corresponds to no changes. (F-J) Cumulative plot for the coverage of SETDB1, ATF7IP, input material, and H3K9 modifications in control and Atf7ip-deleted cells, for
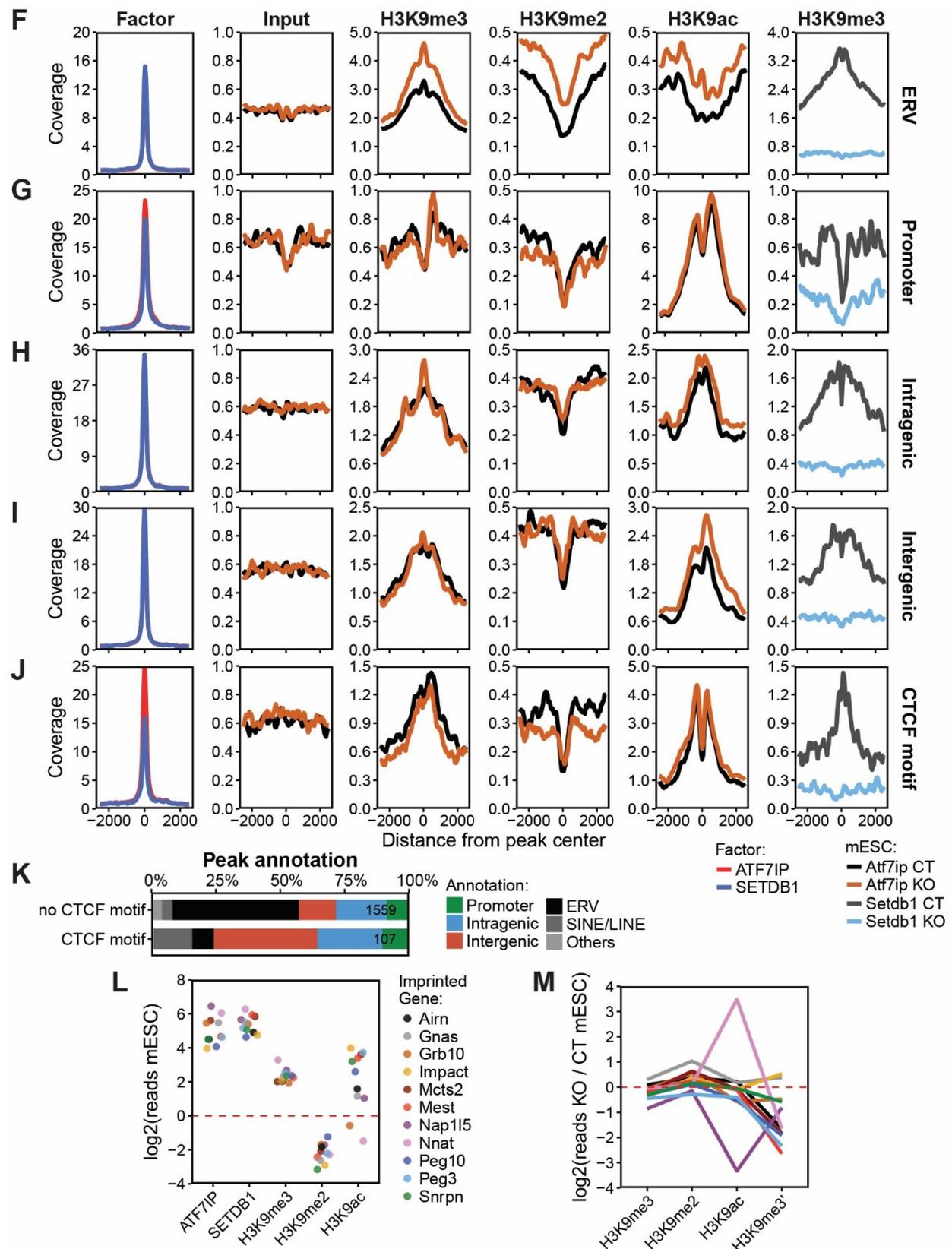
common peaks at different genomic annotations and at peaks containing CTCF motifs. (K) Percentage of genomic annotations associated with common SETDB1/ATF7IP peaks according to the presence or not of CTCF motifs. (L) Coverage normalized to input material for the SETDB1, ATF7IP and H3K9 modifications in control mESC for imprinted genes showing SETDB1 and ATF7IP binding at their promoter region. (M) Fold enrichment of H3K9 modifications in Atf7ip KO or Setdb1 KO cells over control mESC for the genes in L. Setdb1 KO mESC data is based on previously published datasets (Karimi et al., 2011).

The common peaks in the vicinity of annotated TSS did not show important changes in H3K9 modifications after *Atf7ip* deletion, regardless of the levels of H3K9me3 where they localized (Figure 3.14E - upper panel). However, the few upregulated genes in the mutant mESC bearing common peaks at their TSS tended to lose H3K9me2 and gain H3K9ac. Interestingly, none of those genes were marked by H3K9me3 in control cells. Similar picture is seen for the genes upregulated in *Setdb1* KO mESC and co-bound by SETDB1 and ATF7IP (Figure 3.14E - lower panel). When the coverage of the H3K9 modifications at common binding sites were analyzed in both control and *Atf7ip* KO mESC, distinct trends were observed at the different genomic annotations (Figure 3.14F-I). At ERV-associated sites, high levels of H3K9me3 were observed in control cells, together with low enrichment for H3K9ac, while *Atf7ip* deletion led to increase in H3K9me2/3 and in the acetylation mark. Similar picture was observed at intra- and intergenic binding sites, although to a lower extent. Co-bound promoters, on the other hand, suffered slight decrease in H3K9me2/3 after *Atf7ip* deletion, keeping high levels of H3K9ac. Intriguingly, the first nucleosomes downstream of the TSS showed increase in H3K9me3 instead. Besides that, *Setdb1* deletion led to severe reduction of H3K9me3 at common biding sites, independently of the genomic context.

Considering the recently described role of SETDB1 in impeding CTCF biding to its motif and controlling regional genomic insulation (Chapter 1.5.c Neuronal development), changes in enrichment for H3K9 modifications at common binding sites containing CTCF motifs were assessed (Figure 3.14J). These sites behaved in a similar fashion to what is seen for co-bound promoters, even though the majority of them occur at intra- and intergenic regions (Figure 3.14K). Regarding the role of SETDB1-ATF7IP complex at imprinted genes (Chapter 1.6.b Endogenous genes), several of them displayed strong occupancy of both factors in mESC (Figure 3.14L). Besides lacking H3K9me2, these regions were in general enriched for H3K9me3 and H3K9ac in control cells. In the absence of ATF7IP imprinted genes co-bound by SETDB1-ATF7IP tended to lose H3K9me3, while increasing the deposition of H3K9me2 (Figure 3.14M). Deletion of *Setdb1* also led to impaired deposition of H3K9me3 at their promoters. In summary, common binding sites for SETDB1 and ATF7IP at ERV regions show opposing response to *Atf7ip* deletion compared to promoters co-bound by these proteins, whereas both sites lose H3K9me3 following *Setdb1* deletion.

3.15. SETDB1-ATF7IP complex binds and silences several ERV from classes I and II

The binding of both SETDB1 and ATF7IP to regions containing ERV annotations (3.14 Promoters and ERV bound by SETDB1-ATF7IP respond differently to *Atf7ip* deletion) and their previously described role in repressing some TE families (Chapter 1.10 Regulation of SETDB1 catalytic activity by ATF7IP) strongly suggested the cooperation of these proteins in silencing repetitive regions. To investigate this matter, all the reads (including the ones with multiple alignments) from occupancy in Flag KI mESC and from ChIP-Seq and RNA-Seq in

control and *Atf7ip* KO mESC were mapped to UCSC annotated repeat clades. In general, the clades strongly bound by SETDB1 also displayed strong signal for ATF7IP over the input material (Figure 3.15A), suggesting the collaboration between these proteins. Interestingly, these regions tended to be enriched for H3K9me3 in control cells, indicating the importance of this mark for their repression (Figure 3.15B). Subdividing all the TE into different classes, it becomes clear that SETDB1 and ATF7IP forms a complex specifically at ERV, but not at ERV class III (Figure 3.15C).

To analyze the impact of ATF7IP loss at all TE families individually according to their respective class the occupancy data for factors and histone modifications was integrated with the transcriptome in control and *Atf7ip* KO cells. Basically, the data was ordered decreasingly according to ATF7IP binding and local regression was performed for all datasets at three ERV classes and at LINE as control region. The strength of the signal for SETDB1 and ATF7IP was very consistent throughout the different families and practically remained at background levels at ERV class III and LINE, as expected (Figure 3.15D). Interestingly, the presence of both factors at ERV classes I and II correlated with transcriptional derepression after *Atf7ip* deletion, even though no changes in H3K9me3 deposition could be observed. For ERV class I, ATF7IP loss was followed by H3K9me2 increase together with H3K9ac. Unexpectedly, for ERV class II H3K9ac levels became lower in the mutant.
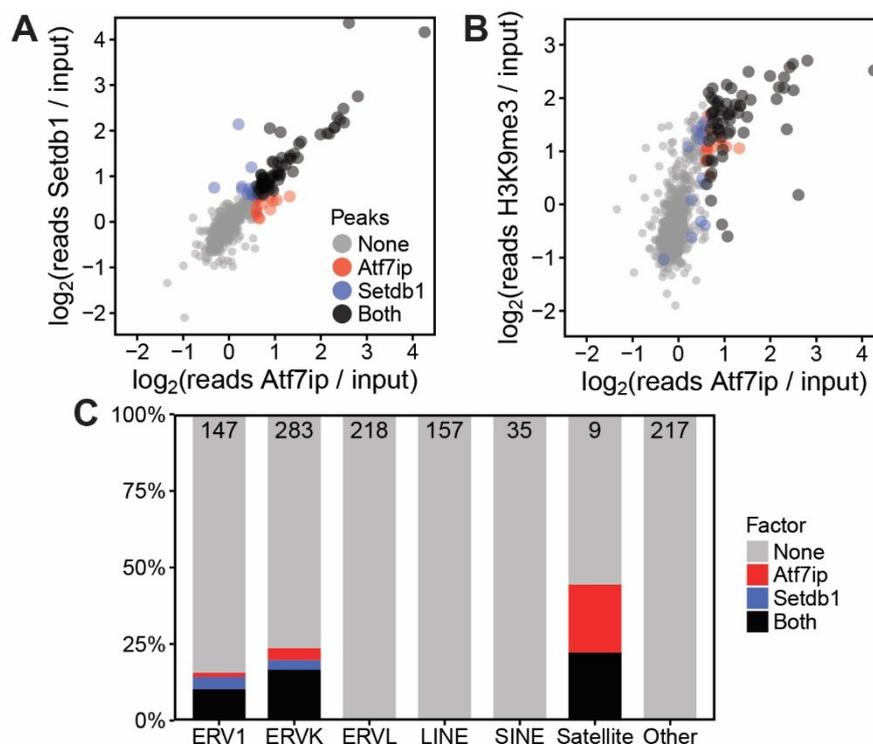
Figure 3.15 | SETDB1-ATF7IP complex binds and silences several ERV from classes I and II.
(A) Correlation between the genome coverage of SETDB1 and ATF7IP at all UCSC annotated TE families normalized to input signal. The color indicates whether the peak is a sole peak or co-occurs with ATF7IP. (B) Same as in A, but for ATF7IP and H3K9me3. (C) Percentage of families belonging to different TE classes which are bound by either ATF7IP or SETDB1 or both together. The total number of families composing each of the classes depicted is shown on the top of their respective bar.
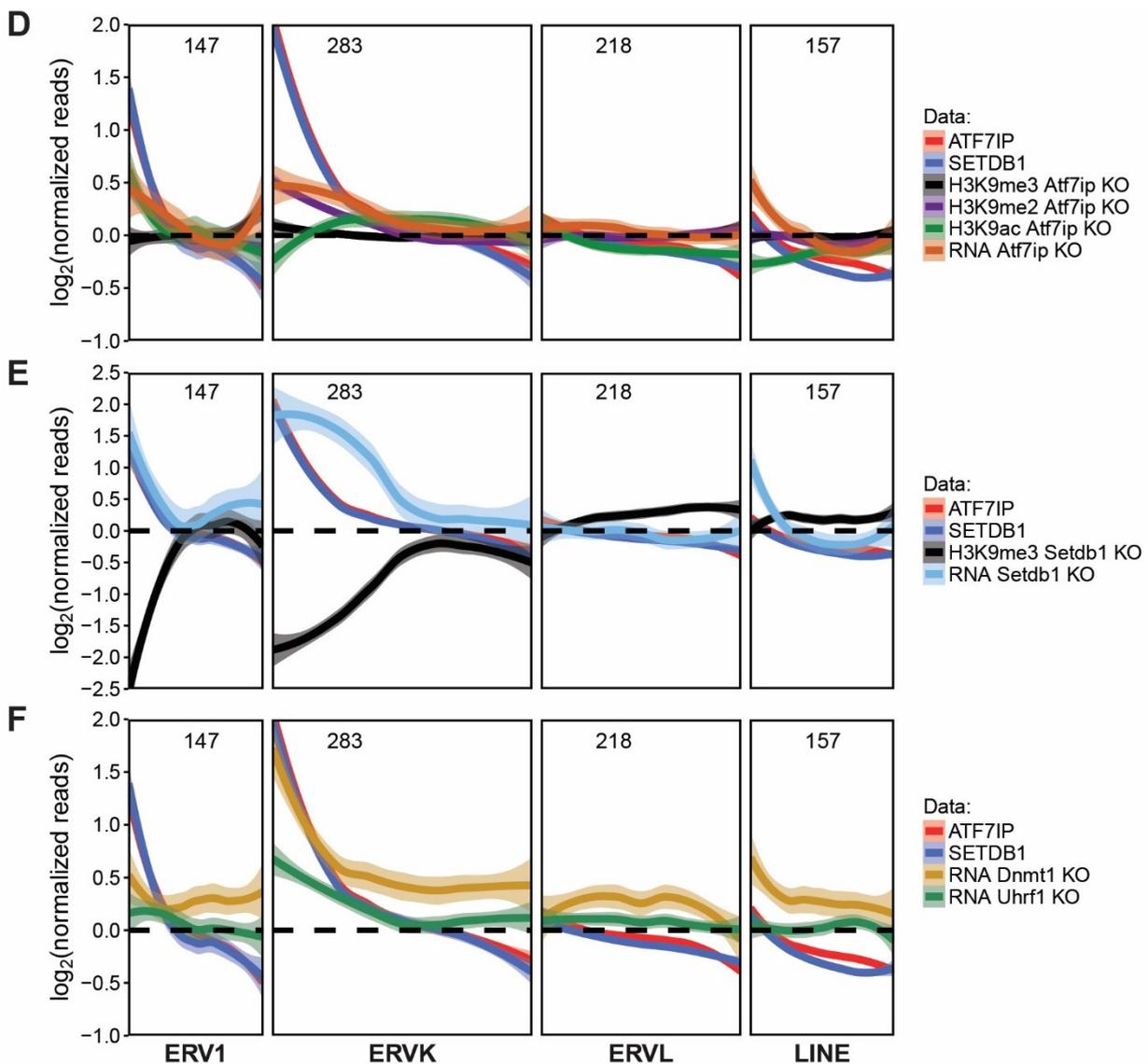
(D) Local regression of the coverage of each dataset upon all the TE families according to their origin. The data was normalized to input signal and, for the KO datasets, also to the control samples. Note that the data was ordered according to the strength of ATF7IP binding. Each data point represents one family and the panels are scaled according to the number of families. The area around the lines indicates the respective standard error. The number on top of each panel represents the number of families of the respective TE class. (E) Same as in D, but for *Setdb1* KO mESC using publicly available datasets (Karimi et al., 2011). (F) Same as in D, but for *Dnmt1* KO and *Uhrf1* KO mESC using published datasets (Sharif et al., 2016).

Assessing published datasets for the impact at these regions of *Setdb1* deletion in mESC (Karimi et al., 2011), H3K9me3 levels were strong lost and followed by transcriptional activation (Figure 3.15E). Publicly available datasets were also used to investigate the overlap of SETDB1-ATF7IP complex activity with DNAme machinery (Sharif et al., 2016). The co-binding of SETDB1 and ATF7IP was also correlated with stronger upregulation of TE when DNAme machinery is impaired (Figure 3.15F). Thus SETDB1-dependent H3K9me3 deposition can be independent of ATF7IP in a long term after its deletion, but is not enough for keeping their TE targets completely repressed. Besides that, DNMT1 plays an important role as a co-repressor of the TE regulated by the SETDB1-ATF7IP complex.

3.16. DNAme synergizes with SETDB1-ATF7IP repression at IAPEz but not at MusD

To assess the functions of SETDB1-ATF7IP complex at TE in a more specific manner, the analyses were narrowed down to individual elements by counting the number of tags mapping to all UCSC annotated copies from IAPEz, ETn and MERVL. Almost all MLV and IAPEz copies and several ETn elements were co-bound by both factors (Figure 3.16A). After ordering all copies according to the levels of ATF7IP signal and performing local regression, some trends could be observed for each clade (Figure 3.16B). Even though IAPEz and ETn belong to the same ERV class, they had some discrepancies in the response for ATF7IP loss. For instance, while the H3K9ac levels dropped considerably at the former after *Atf7ip* deletion, they remained the same in the latter. Besides that, H3K9me3 levels showed higher increase in IAPEz. On the other hand, the enrichment for H3K9me2 increased in the co-occupied copies of both subfamilies. As expected, no changes were observed for MERVL copies. Investigating published datasets, loss of SETDB1 (Karimi et al., 2011), DNMT1 or UHRF1 (Sharif et al., 2016) led to derepression of IAPEz copies (Figure 3.16C). Interestingly, the elements containing intermediate levels of SETDB1 and ATF7IP were the ones exhibiting the strongest derepression. *Dnmt1* deletion, however, did not cause upregulation of ETn elements, indicating that DNAme does not play a role in silencing these regions. Thus, DNAme may be related to the strong loss of H3K9ac at IAPEz in *Atf7ip* KO mESC.

To corroborate these findings, the fluctuation in H3K9 modifications due to *Atf7ip* deletion in mESC was evaluated by ChIP-qPCR in three biological replicates for control and mutant (Figure 3.16D-G). These observations were in fact replicable. To assess the influence of DNAme on the repression of these TE and imprinted genes, control and *Atf7ip* KO mESC were transferred to 2i+vitC medium (Chapter 3.9 Loss of DNA methylation further enhances *Atf7ip* knockout phenotype). Strikingly, the higher H3K9me3 levels observed in mutant cells grown in LIF medium were reduced again, reaching levels similar to the ones in control samples (Figure 3.16H-I). This suggests that presence of ATF7IP may refrain SETDB1 from excessive H3K9me3 deposition at regions where DNAme synergizes with ATF7IP for silencing (Figure 3.9B). When the cumulative coverage for SETDB1, ATF7IP and H3K9 modifications are analyzed along the consensus sequence for IAPEz and MusD, the gag region containing higher signals for those factors show stronger increase in H3K9me3 deposition at IAPEz after *Atf7ip* deletion when compared to MusD (Figure 3.16J-K). Besides that, deletion of *Dnmt1* only affects IAPEz, whereas *Setdb1* loss affects both as seen before (Figure 3.16C). Binding of neither SETDB1 nor ATF7IP is seen at MERVL and any alterations following *Setdb1* or *Atf7ip* deletion may be due to indirect effects (Figure 3.16L).
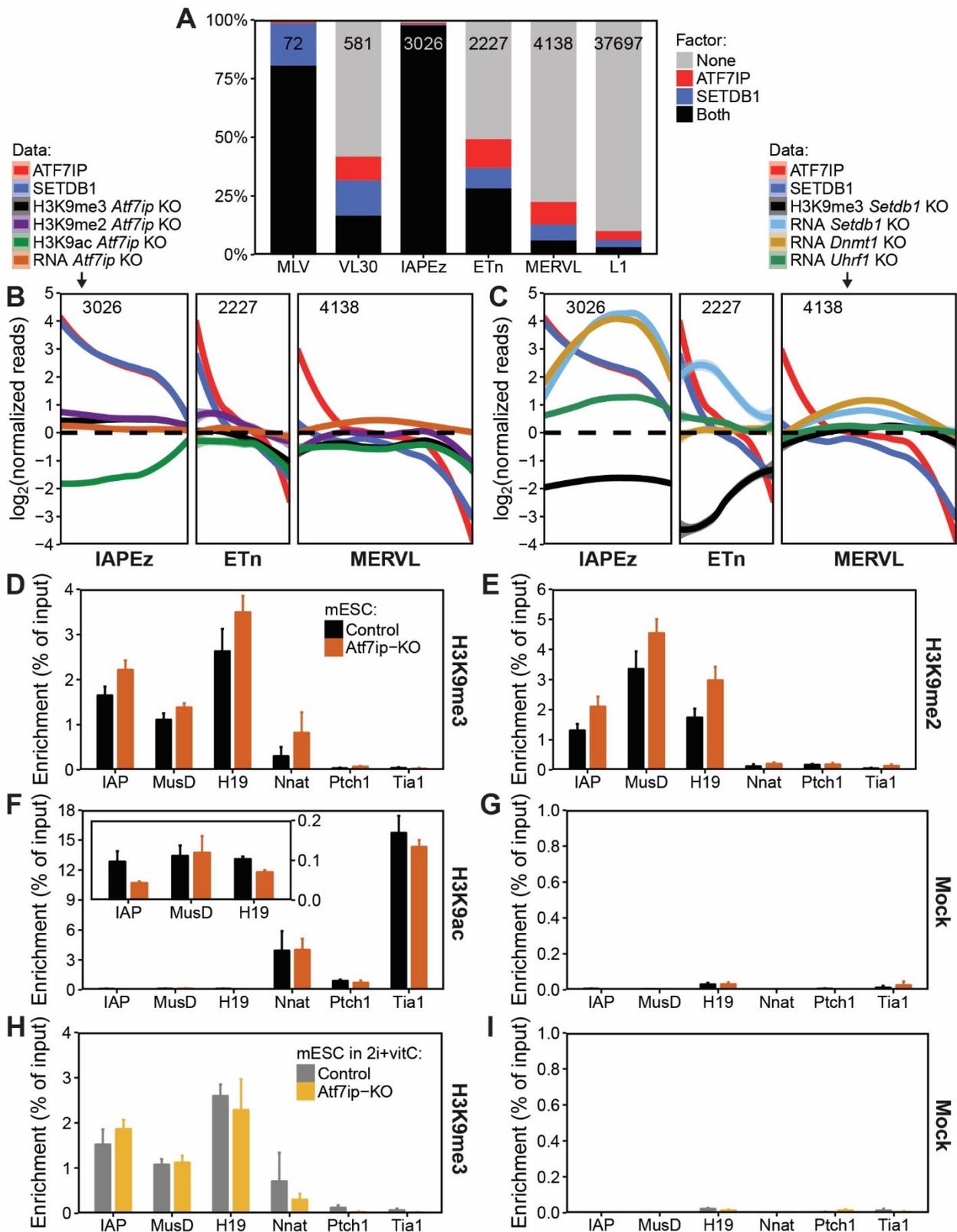
Figure 3.16 | DNAme synergizes with SETDB1-ATF7IP repression at IAPEz but not at MusD.
(A) Percentage of copies belonging to different TE families which are bound by either ATF7IP or SETDB1 or both together. The total number of copies composing each of the families depicted is shown on the top of their respective bar. (B) Local regression of the coverage of each dataset upon all the TE copies of each family. The data was normalized to input signal and, for the KO datasets, also to the control samples. Note that the data was ordered according to the strength of ATF7IP binding. Each data point represents one copy of the TE and the panels are scaled according to the number of individual elements. The area around the lines indicates the respective standard error. The number on top of each panel shows the number of copies of the respective TE family. (C) Same as in B, but for *Setdb1* KO (Karimi et al., 2011), *Dnmt1* KO and *Uhrf1* KO mESC (Sharif et al., 2016) using publicly available
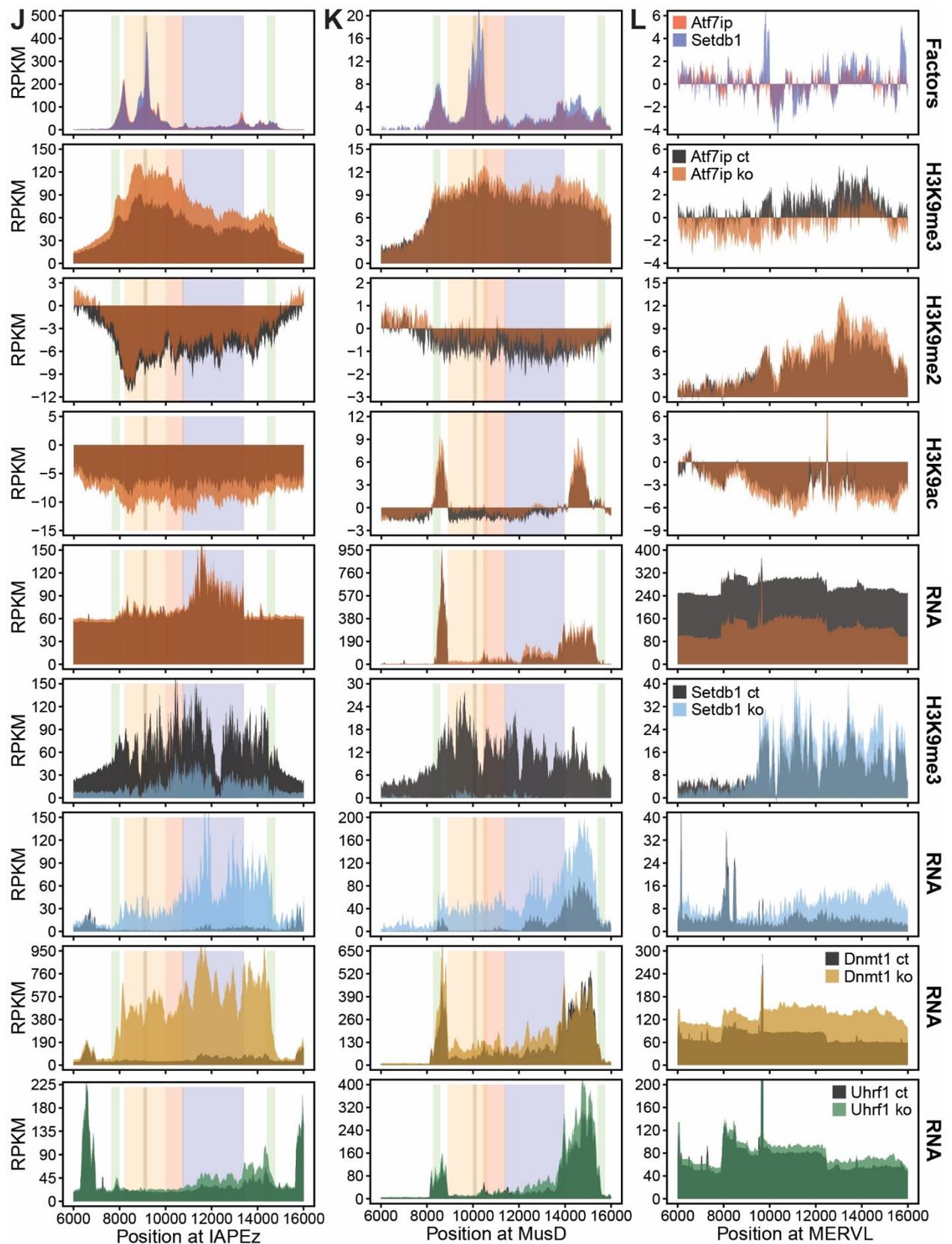
datasets. (D-G) Enrichment as percentage of input for H3K9 modifications at TE and imprinted genes in control and *Atf7ip* KO mESC. For the mock sample, sheared chromatin was incubated with beads not covered with any specific Ab. (H-I) Same as in D, but for control and *Atf7ip* KO cells grown in 2i+vitC medium. (J) Cumulative coverage for factors, H3K9 modifications and transcriptome along the consensus sequence for all IAPEz elements. The colors indicate the following regions at the ERV: LTR, gag, pro and pol. The darker region inside the gag represents the fragment used for the exogenous TE silencing reporter assay. (K) Same as in J, but for MusD elements. (L) Same as in J, but for MERVL elements. Information about subregions is not available for this elements.

3.17. Generation of a platform for multifunctional editing of *Atf7ip* locus in mESC

Targeted genome editing is a very powerful tool for studying the functions of a gene. A platform for quick and versatile engineering of a locus can be created by insertion of the multifunctional integrase (MIN) tag containing the phage attachment site (*attP*) for site-specific recombination with the bacterial attachment site (*attB*) as previously described (Mulholland et al., 2015). The serine integrase coded by the bacteriophage Bxb1 catalyzes the recombination generating attachment site junctions (*attL* and *attR*) as a byproduct of the reaction of integration (Singh et al., 2013). Thus, a modular system is created, where vectors carrying a desired DNA sequence at the right position can be integrated at the target locus by transient co-transfection with the Bxb1 integrase (Figure 3.17A).

Figure 3.17 | Generation of a platform for multifunctional editing of *Atf7ip* locus in mESC.
(A) Diagram of the site-specific recombination mediated by the Bxb1 integrase. Bxb1 recognizes and binds as a dimer to the *attP* and *attB* sequences to mediate recombination, generating *attL* and *attR* scars containing one piece of each of the recombination sites. (B) Sequence of the N-terminal region of *Atf7ip* locus indicating the start codon (orange). The sgRNA sequence is also shown (purple) together with the PAM sequence (blue). The PAM is recognized by the Cas9, which generate double strand breaks at the region indicated with blue arrows. (C) Diagram of the HDR template design and of the genotyping scheme. (D) UV-exposed agarose gel to distinguish between wild type and knockin clones. M indicates the lane containing the ladder. (E) Sanger sequencing of the PCR products of the *Atf7ip* locus from four independent clones isolated after CRISPR knockin. The FLAG-tag (green) was correctly recombined into the locus after the endogenous start codon (orange). The PAM sequence (blue) and the sgRNA target site (purple) were mutated (red) to avoid cleavage of the HDR template by the Cas9.

For modifying the *Aff7ip* locus, the Cas9 was targeted to a region nearby the start codon, with the aim to insert the MIN-tag at the 5'-end of the coding sequence (Figure 3.17B). In this way, any sequence integrated in the locus later on by recombination would be placed at the beginning of the transcript (Figure 3.17C). To favor the HDR the Cas9 was tagged with a peptide which signals for nuclear import during mitosis, when DNA damage are more prone to be fixed via HDR. Single clones were isolated to obtain isogenic cell lines and insertion of the tag was screened by PCR with primers surrounding the tagged region (Figure 3.17D). Correct sequence of the locus was verified by sequencing PCR products covering the whole targeted region (Figure 3.17E). Hence, a platform for fast integration of sequences of interest in the *Atf7ip* locus was generated and allows the switching of the expression of the endogenous protein by that of mutated sequences keeping the physiological conditions and the endogenous expression levels.

3.18. Domains 1 and 2 are essential for ATF7IP to exert its repressive functions

The *Atf7ip* locus encodes for a protein containing several interesting domains and motifs as introduced previously (Chapter 1.9.a Structural features). To study the contribution of each of its features for the regulation of SETDB1 activity several mutant forms of ATF7IP were cloned in a vector containing the *attB* sequence. In this way, these mutants could be integrated in the *Atf7ip* locus at the beginning of the coding sequence in the mESC line containing the *attP* sequence right after the start codon (Figure 3.18A). Of note, the vectors used for integration contain a late polyadenylation (polyA) signal following the mutant coding sequence for efficient transcription termination (Schek et al., 1992), making sure transcription does not go further into the locus. This means that the endogenous gene is knocked out as a consequence of the recombination.

Integration of the mutants of interest was done by co-transfecting the MIN-*Atf7ip* KI mESC line with a plasmid coding for the Bxb1 integrase and another plasmid containing the *attB* followed by a cDNA of interest and an antibiotic resistance. As not all the cells show integration in both alleles, the strategy adopted was to first integrate an empty vector, for knockout only (Figure 3.18B). Afterwards, isogenic clones containing integration in only one allele were screened by PCR (Figure 3.18C). This heterozygous clone was then used to express the panel of *Atf7ip* mutants separately by site-specific recombination (Figure 3.18B). Cells containing one knockout allele and a second one expressing the mutant of interest were analyzed in rescue experiments.

The different mutants chosen to be analyzed included: empty vector (to generate knockout cells), full-length cDNA (to account for the ability to rescue using the complete coding sequence), deletion of Domain 1 keeping the NLS, deletion of the whole Domain 1, deletion of Domain 2, mutation at the SIM, deletion of the ABS, mutation at the NLS and a truncation formed only by the NLS-Domain 1 (Figure 3.18D). Cloning of all these sequences were performed with the help of Büşra Turgu, a master student from the Faculty of Biology - LMU, whom I supervised during a two-months internship.
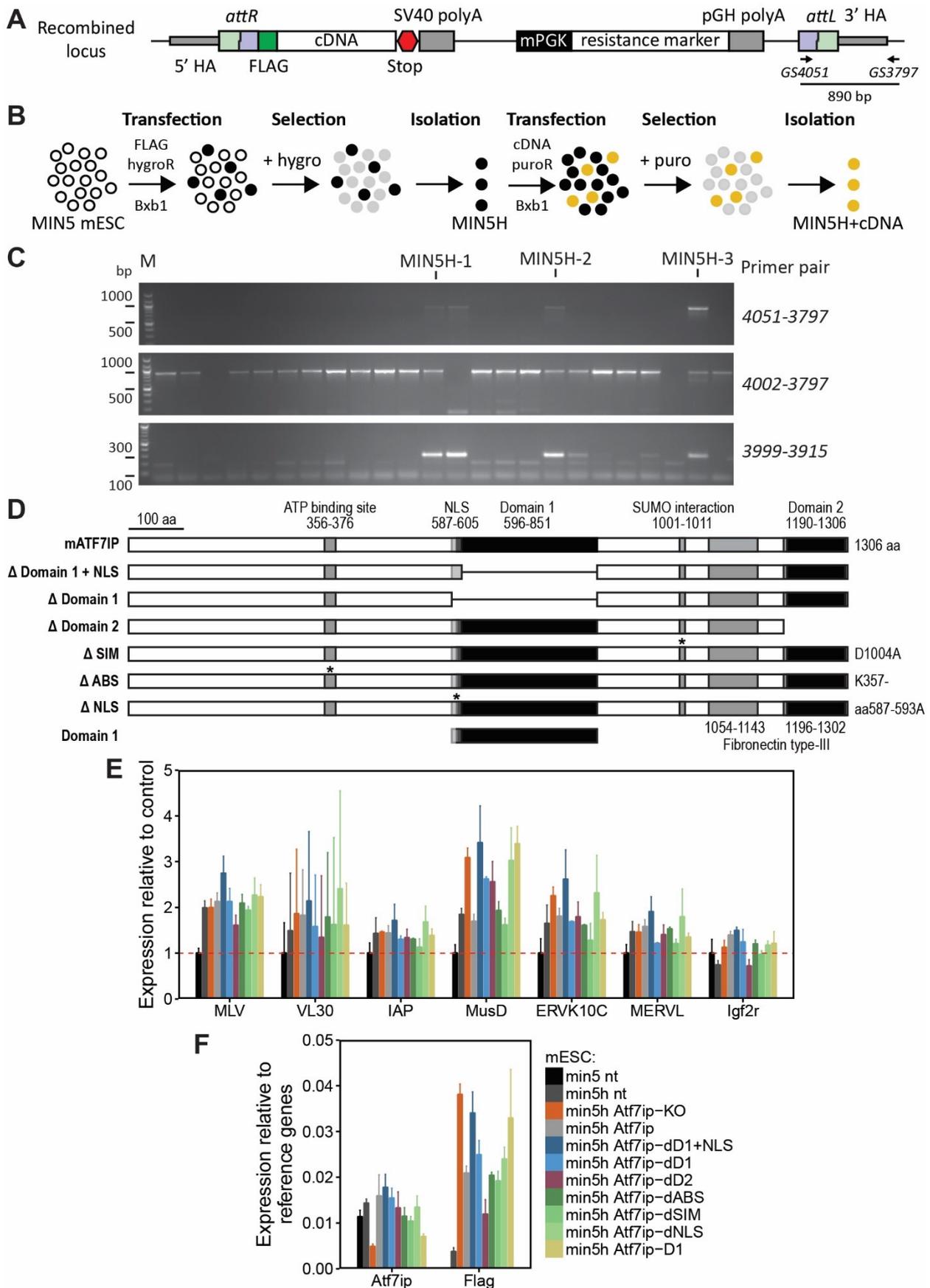
Figure 3.18 | Domains 1 and 2 are essential for ATF7IP to exert its repressive functions.
(A) Graphical representation of the MIN-tagged Atf7ip locus after Bxb1-mediated recombination either with an empty vector to knockout the allele or a vector containing a mutant sequence to replace the expression of the endogenous one. A cassette for antibiotic resistance is part of the vector and also gets integrated. (B) Experimental design for

generation of mESC lines where one *Atf7ip* allele is knocked out and the other is replaced by different mutant sequences via Bxb1-mediated recombination. A first round of co-transfection of Bxb1 expressing vector and an empty vector containing only the Flag sequence after the *attB* together with the hygroR cassette. Isogenic cell lines were isolated and expanded after long term selection and used in a second round of co-transfection, this time with a vector containing different *Atf7ip* mutants following the *attB* and carrying the puroR cassette. Again, isogenic clones were isolated and expanded. (C) UV-exposed agarose gel to distinguish between wild type and knockin cell lines. (D) Graphical representation of the full-length ATF7IP and the panel of mutants used in rescue experiments of Min-*Atf7ip* KI mESC via Bxb1-mediated recombination. The different domains and motifs are depicted. The asterisks indicate the position of the point mutations. The specific aa targeted are specified on the right side of the sequence. (E) Expression levels of different TE families, members of ERV classes I (MLV and VL30), II (IAP, MusD and ERVK10C) and III (MERVL), and of the imprinted gene *Igf2r* in Min-*Atf7ip* KI mESC after Bxb1-mediated recombination. The expression levels are relative to the homozygous Min-*Atf7ip* KI samples. For each replicate, expression was normalized by the geometric mean of *Gapdh* and *Hprt* expression. The red dotted line indicates the level correspondent to no changes. (F) Same as in E, but for *Atf7ip* and the *Flag* sequence relative to reference genes *Gapdh* and *Hprt*.

To finally investigate the importance of the selected regions for proper function of ATF7IP, cells were harvested 6 days after the second round of recombination leading to expression of the mutants on the place of the endogenous locus. Expression of endogenous retroviruses were assessed by RT-qPCR. The removal of one allele already led to derepression of some TE to some extent, especially of MusD elements, and silencing was further impaired by knocking out the second allele (Figure 3.18E,F). Expression of full-length *Atf7ip*, but not of Domain 1 truncation, was enough to keep MusD transcription repressed to the same level as observed in the heterozygous cell line. However, either constructs were able to keep repression of the other TE tested. Interestingly, *Atf7ip* mutants lacking either Domain 1 or Domain 2 could not prevent derepression. Similarly, mutation at the NLS also impaired silencing. However, the ABS and the SIM did not seem required by ATF7IP to maintain proper transcriptional silencing.

3.19. ATF7IP interacts with proteins related to translation, transcription and cell cycle

The ability of SETDB1-ATF7IP complex to exert functions in multiple contexts may be due to other auxiliary proteins which could work along with this complex. The interactors of SETDB1 have been investigated by high throughput technique previously (Thompson et al., 2015), however, the interactome of ATF7IP remains unclear. To address which proteins could be associated with the complex via an interaction with ATF7IP, a high throughput approach was adopted. Taking advantage of the Flag-*Atf7ip* KI mESC line generated, co-IP was carried out against the FLAG epitope. Considering the importance of SUMOylation for some of the SETDB1 interactions to take place, activity of SUMO-specific proteases was blocked by addition of the cysteine protease inhibitor N-ethylmaleimide (NEM). At first, the immunoprecipitates were evaluated by WB (Figure 3.19A) and silver staining (Figure 3.19B) to assess the efficiency of capturing the target protein and the level of non-specific binding of the background. These experiments were performed with the help of Ren Xie, a master student from the Faculty of Biochemistry - LMU, whom I also supervised during a two-months internship. After that, the proteins in the immunoprecipitated material were digested into smaller peptides and sequenced

by mass spectrometry (co-IP-MS), which was performed with the help of PD. Andreas Schmidt through the ZFP Mass Spectrometry Facility of the Biomedical Center - LMU.

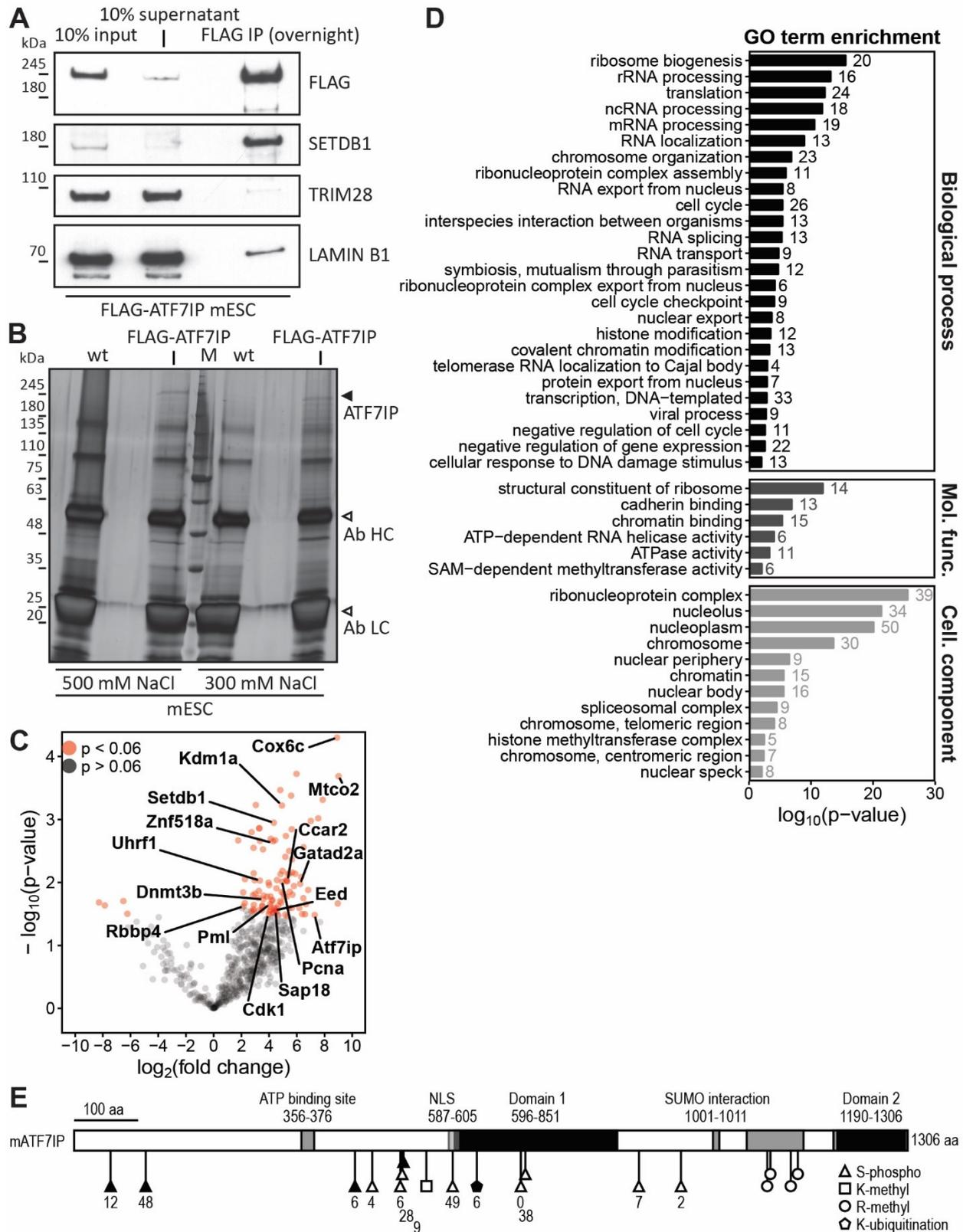**Figure 3.19 | ATF7IP interacts with proteins related to translation, transcription and cell cycle.**
(A) Digital image of the membranes used for blotting the FLAG Co-IP material of Flag-*Atf7ip* KI mESC nuclear extracts after separation by SDS-PAGE. Membranes was probed with Ab against different epitopes and LAMIN B1 was used as a control for non-specific binding. (B) Digital image of the silver stained SDS-PAGE gel used for

separating the FLAG Co-IP material of Flag-*Atf7ip* KI mESC nuclear extracts. The closed arrowhead indicates the ATF7IP band and the two open arrowheads indicates the heavy-chain (HC) and light-chain (LC) of the Ab used. (C) Volcano plot displaying all identified peptides according to the enrichment over the control and the levels of significance. Significant proteins identified ($p < 0.05$) are highlighted in red. A few proteins of interest are annotated in the plot. (D) GO term enrichment analysis for biological processes (upper), molecular function (middle) and cellular component (lower panel) of significantly identified proteins in the co-IP-MS of FLAG Ab immunoprecipitates from Flag-*Atf7ip* KI mESC nuclear extracts. Shown is the probability based on the p-value. The number on top of each bar indicates the amount of genes fitting the specific GO term. (E) Scaled graphical representation of the *Atf7ip* locus and the position of PTM identified in the co-IP-MS results (open symbols) or only in the PhosphoSitePlus database (closed symbols). The number nearby the modification indicates the number of datasets where this specific modification was observed and which are available at PhosphoSitePlus database.

The identified peptides were then aligned to the mouse UNIPROT database. After filtering the significantly enriched proteins, a list of around 100 candidate ATF7IP partners were obtained (Figure 3.19C). The specific enrichment of SETDB1 was strikingly detected. For the peptide counts and the protein coverage of the top 80 candidates, consult the Appendix (Table 9.1). A GO term enrichment analyses of all the significantly identified partners identified a series of interesting terms, for instance, the ones related to transcription, translation, cell cycle control and chromatin modification (Figure 3.19D). Of note, several PTM were identified including phosphoserine, methyl-arginine and methyl-lysine (Figure 3.19E). An investigation of other possible modifications of ATF7IP was done by inspecting the comprehensive database PhosphoSitePlus (Hornbeck et al., 2015) available online (http://www.phosphosite.org). Altogether, these data indicate that ATF7IP is heavily modified, what is in agreement with the several bands identified in WB (Chapter 3.3 Transposable elements lose repression following ATF7IP deletion), and some of these PTM may be involved in promoting the functions of this protein.

3.20. ATF7IP acts in the context of SETDB1 complex and synergizes with other proteins

A way to investigate the functional relationship between pathways is by assessing the occurrence of genetic interaction, a phenomenon which describes the synergistic effects of phenotypes arising from the concurrent deletion of two genes (Mani et al., 2008). In this way, elements contributing to compensatory effects via independent pathways can be identified. On the opposite fashion, genes not showing additive phenotypes could be components of the same silencing machinery, in case they are known to act on the context in question. It is clear that silencing mechanisms can present overlapping functions, especially those acting upon TE (Chapter 1.7 Repression of transposable elements). In order to identify the specific context in which ATF7IP acts for the regulation of ERV repression and what are other parallel pathways able to compensate for ATF7IP loss, genetic interaction with several other genes was tested by taking advantage of the reporter assay for exogenous retrotransposon silencing (Chapter 3.4 Silencing kinetics of an exogenous IAPEz reporter is impaired by ATF7IP deletion).

The list of genes to be tested was assembled based on factors which are known to be involved in TE repression. Besides that, some of the proteins and ribonucleoproteins (RNP) identified as

ATF7IP interactors were added to the list (Chapter 3.19 ATF7IP interacts with proteins related to translation, transcription and cell cycle). Firstly, to verify whether any of those genes gathered on this list are involved in silencing of the IAPEz-gag reporter, each of the genes were knocked out individually in wild type mESC. Six days after deletion by CRISPR/Cas system, the knockout cell pools were analyzed in the reporter assay (Figure 3.20A). As expected, deletion of the factors known to silence TE led to impaired silencing of the reporter. Besides them, only *Uhrf1*-deleted cells exhibited some disruption of silencing. Unfortunately, the possibility of low efficiency knockout induction by each sgRNA cannot be discarded, as most of them were not assessed for efficiency prior to this screening. When probed for genetic interaction with *Atf7ip*, most of those genes did not show synergistic effects. The most striking genes leading to a cumulative effect were *Atrx*, *Morc3* and *Dnmt1* (Figure 3.20B). The role of *Atrx* in regulating this reporter sequence was previously shown (Sadic et al., 2015), whereas the involvement of Morc3 was recently identified by Sophia Groh during her Master thesis (Groh, Master Thesis, 2015) and is currently under investigation as the focus of her PhD thesis. Regarding *Dnmt1*, IAPEz regulation by the encoded protein was already described (Sharif et al., 2016). Of note, the strong effect of *Setdb1* deletion for the silencing of the reporter in wild type mESC was not observed when it was deleted in *Atf7ip* KO cells. This is an indication that both proteins act in the context of the SETDB1-ATF7IP complex and that initiation of exogenous TE repression by SETDB1 cannot take place in the absence of its partner protein.

To validate some of the previous results, the experiments were repeated with more replicates. Interestingly, both *Atrx* and *Daxx* showed further impairment in silencing of the IAPEz-gag fragment reporter (Figure 3.20C). The effects of acute dKO were also tested to prevent any possibility of long term adaptation of the cells. For this, cells were co-transfected with a plasmid expressing only a sgRNA against *Atf7ip* and another vector expressing a sgRNA against a second target gene, the Cas9 and antibiotics resistance, using a higher molar ratio of the first one. In this way, most of the transfected cells would incorporate the former, however, only cells transfected also with the latter would survive selection and express Cas9. Firstly, cells were immuno-stained for ATF7IP 6 days after KO induction to measure the efficiency of obtaining a dKO cells. The amount of cells displaying loss of ATF7IP signal were similar between single KO and dKO experiments (Figure 3.20D). Finally, these cells were probed for ability to silence the TE reporter and the synergistic effects were again observed in the dKO cells (Figure 3.20E).

Figure 3.20 | ATF7IP acts in the context of SETDB1 complex and synergizes with other proteins.
(A) Relative percentage of remaining EGFP+ cells during a time course after the gag fragment reporter assay in wild type mESC following acute deletion of a gene of interest. Genes were grouped in different categories, according to their functions. Three controls were used: non-transfected cells and cells transfected with either an empty vector or with a sgRNA targeting a non-related intergenic region. (B) Same as in A, but for Atf7ip KO mESC after acute co-deletion of a second gene of interest. (C) Relative percentage of remaining EGFP+ cells during a time course after the gag fragment reporter assay in wild type mESC after acute deletion of a gene of interest in either wild type mESC or in Atf7ip KO cells. (D) Percentage of cells not stained for ATF7IP after acute deletion of a gene of interest or dKO of Atf7ip and a second gene in wild type mESC. (E) Relative percentage of remaining EGFP+ cells during a time
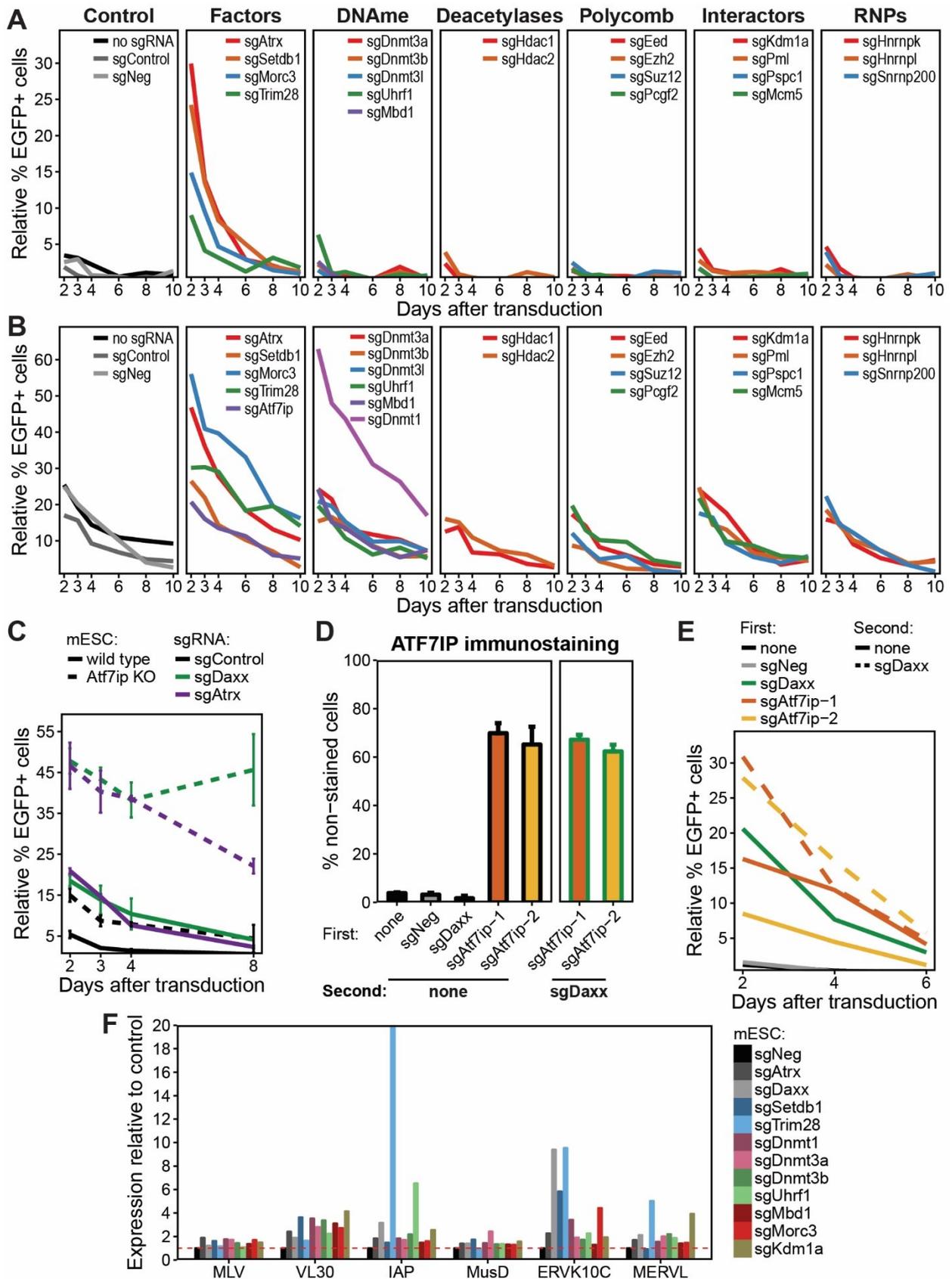
course after the gag fragment reporter assay using the cells in D. (F) Expression levels of different TE families, members of ERV classes I (MLV and VL30), II (IAP, MusD and ERVK10C) and III (MERVL) in *Atf7ip* KO mESC after co-deletion of a second gene to test for genetic interaction. The expression levels are relative to the sgRNA targeting a negative control region. Expression was normalized by the geometric mean of *Gapdh* and *Hprt* expression. The red dotted line indicates the level correspondent to no changes.

When genetic interaction was assessed for its involvement in the regulation of endogenous TE, some interesting trends were observed. For instance, IAP elements were strongly impaired in *Atf7ip* KO mESC after depletion of *Trim28* or *Uhrf1*. MusD showed slight upregulation after *Dnmt3a* co-deletion, whereas ERVK10C were mainly affected by co-deletion of *Daxx*, *Setdb1*, *Trim28*, *Dnmt1* and *Morc3* (Figure 3.20F). In summary, the SETDB1-ATF7IP complex have a role in silencing newly integrated exogenous TE and is assisted by other pathways which may act to ensure silencing maintenance. This suggests a similar role for endogenous TE when they actively transpose. Besides that, regulation of endogenous TE, especially IAP and ERVK10C elements, is strongly dependent on other proteins, which compensate for *Atf7ip* loss in mESC.

4. DISCUSSION

4.1. Establishment of tools for the investigation of SETDB1 regulation by ATF7IP

Several functions of the protein SETDB1 were so far described. The characterization of the molecular mechanism utilized by such proteins with a large spectrum of functions is no easy task. The requirement for specialized methods to address the most specific biological questions has increased exponentially the number of tools available. In this scenario, the CRISPR/Cas9 system emerged as a promising means for genome editing allowing for innumerable applications (Shalem et al., 2015). Gene knockout, knockin, transcriptional repression and activation are some of them. It became an essential technique for this work and by using it three very important tools could be generated, which permitted the investigation of SETDB1 regulation by its cofactor ATF7IP. Firstly, *Atf7ip* KO mESC lines could be generated and characterized (Figure 3.1). Using these cells, the functions of ATF7IP could be clarified by reverse genetics, where the phenotype was characterized applying various techniques, for instance, transcriptomics. Besides that, these cell lines could be potentially used for verifying whether SETDB1 depends on ATF7IP to bind non-methylated gene promoters. SETDB1 recruitment to endogenous retroviruses was shown to be dependent on TRIM28 and does not rely on ATF7IP (Thompson et al., 2015), whereas its targeting to DNA-methylated gene loci requires ATF7IP binding to MBD1 (Ichimura et al., 2005; Matsumura et al., 2015). However, the mechanism of recruitment to non-methylated promoters has not yet been investigated.

Another very important tool which was developed in this work was the Flag-*Atf7ip* KI mESC lines, which were obtained by homologous recombination triggered after Cas9 double-strand break (Figure 3.12). The usage of the Flag-tag in cases where a specific Ab against the protein of interest is not available or does not work in a specific technique has become very common and brings along the advantage of the availability of highly specific Ab (Zhang et al., 2008). The presence of this tag can be beneficial for a wide range of techniques involving the immune-identification of proteins. The use of Flag-tagged proteins in ChIP experiments, for example, has been shown to give results comparable to when Ab against the endogenous protein is used (Mazzoni et al., 2011). Considering that, the generation of Flag-*Atf7ip* KI mESC lines helped the determination of the genome-wide occupancy of ATF7IP in these cells. Besides, potential proteins partners of ATF7IP could be identified by performing mass spectrometry of the material obtained in co-IP experiments using FLAG Ab. In addition, localization assays by immunohistochemistry is another example of techniques that can be performed with high specificity by using this cell line.

Lastly, the generation of Min-*Atf7ip* KI mESC lines opened a wide range of possibilities to analyze this locus (Figure 3.17). This short tag can be used for site-specific recombination via Bxb1-mediated integration of a provided template. As a result of the integration the endogenous sequence of the targeted locus gets knocked out, whereas the integrated sequence gets

transcribed using the endogenous promoter. Thus, different constructs can be expressed at physiological conditions and endogenous levels (Mulholland et al., 2015). In this work, several ATF7IP mutants and one truncation could be evaluated, together with the full-length sequence and an empty one. With that, distinct domains or specific aa residues of this protein were assessed for their requirement for transcriptional repression in the context of SETDB1. Furthermore, this system can be used for further characterization of the mechanism of SETDB1-ATF7IP complex activity by testing the implications of each of the ATF7IP domains in the methyltransferase activity.

4.2. SETDB1 and ATF7IP can either function in the same context or act independently

Even though ATF7IP forms a complex with SETDB1 to act as a cofactor and regulate its methyltransferase activity (Wang et al., 2003; Basavapathruni et al., 2016), not all functions of each of these proteins seem to overlap. The first indication for this is the lethality observed after deletion of *Setdb1* in mESC (Yuan et al., 2009; Lohmann et al., 2010), which is not recapitulated by *Atf7ip*-depleted mESC (Figure 3.1). Furthermore, transcriptome changes were not as severe in the latter in comparison to the former, with few genes being commonly misregulated (Figure 3.6). Besides that, investigation of the genome-wide occupancy of both proteins reveals an overall discrepancy between their binding sites, with nearly one fourth of SETDB1 targets being shared with its partner ATF7IP (Figure 3.13C-D). Whereas SETDB1 tends to bind intra- and intergenic regions by itself, ATF7IP occupies very often promoter regions independently of the former (Figure 3.13H). Unexpectedly, several GO terms related to development and differentiation were enriched at SETDB1 sole targets. Evidences for the existence of independent roles for these proteins are also available in the literature. For instance, ATF7IP has also been implicated in transcriptional activation of promoters containing Sp1 motifs in concert with Sp1 and when SETDB1 is absent (Ichimura et al., 2005). In a similar fashion, transactivation of the promoter of telomerase-associated genes by both Sp1 and ATF7IP was demonstrated (Liu et al., 2009). In contrast to that, establishment of non-nucleosomal H3K9me1 at pericentric heterochromatin by SETDB1 in the context of HP1α-CAF1 complex during DNA replication was shown not to be linked to ATF7IP (Loyola et al., 2009).

Despite all the independent functions exhibited by either protein, several evidences for the functional overlap between both SETDB1 and ATF7IP were gathered in this work. First of all, the majority of ERV classes I and II bound by SETDB1 was also co-occupied by ATF7IP (Figure 3.15C) and this result was reproduced when individual TE were analyzed (Figure 3.16A). This cooperation in TE silencing was also described before (Thompson et al., 2015; Yang et al., 2015). The observation that silencing of an exogenous retrovirus reporter is not further impaired by co-deletion of *Setdb1* in *Atf7ip* KO mESC reveals another context where these proteins act together (Figure 3.20A-B) and supports previously published evidences for their cooperation in repressing exogenous TE in the context of the HUSH complex (Timms et al., 2016). Although ERV

represented the clearest target of the SETDB1-ATF7IP complex in this work, many genes also displayed co-binding of these factors at their promoter region (Figure 3.14B) and imprinted genes were a good example (Figure 3.14L). In the past, ATF7IP has been implicated in transcriptional modulation of promoters containing Sp1 motifs, where co-binding of MBD1 and SETDB1 resulted in transcriptional repression (Ichimura et al., 2005). Recently, the repression of an adipocyte master regulator was also reported to be dependent on the presence of both ATF7IP and SETDB1 (Matsumura et al., 2015). The strong interaction between these proteins detected by co-IP-MS also supports the importance of the formation of this complex in mESC (Figure 3.19), which was previously described (Thompson et al., 2015). Curiously, the same way SETDB1 is essential during early embryonic development with knockout leading to peri-implantation lethality (Dodge et al., 2004; Keniry et al., 2016), *Atf7ip* knockout mice also exhibit very strong phenotype and do not survive through blastocyst stage (Nuber, Diploma thesis, 2017). This suggests an essential role of the SETDB1-ATF7IP complex as early as in the mouse blastocyst stage.

4.3. Adaptation to *Atf7ip* deletion allows proper cell differentiation in presence of DNAme

One way to study the differentiation capacity of mESC is to derive embryoid bodies from an agglomerate of cells by removing LIF from the medium. This way, pluripotency is lost and the cells can give rise to progenitors from the three germ layers. *Setdb1* null mutant mESC, though, showed intrinsic signs of pluripotency loss together with impaired potential to differentiate and form EB (Bilodeau et al., 2009; Yuan et al., 2009; Lohmann et al., 2010). Unexpectedly, *Atf7ip* null cells had slightly higher levels of pluripotency marker genes (Figure 3.3) and were capable of differentiating into EB (Figure 3.5). The fact that these mutants survive and even show potential to form differentiated cell types can be explained by a process of adaptation that the cells undergo after gene deletion. More specifically, the cells try to compensate the loss of function of a gene by changing expression levels of other genes that may act via parallel pathways (El-Brolosy and Stainier, 2017), what might explain why *Atf7ip* KO cell lines exhibit such a low variation from control cells and are not clearly grouped in PCA (Figure 3.8A). This is supported by the observation that many more genes were found to be dysregulated when the phenotype of ATF7IP loss was analyzed few days after depletion, as observed in knockdown cells in comparison to stable knockout cells kept in long term culture (compare Figure 3.7A-B and Figure 3.8E-F). In addition, acute *Atf7ip* deletion led to ERV derepression to a higher extent than the observed in KO cell lines (Figure 3.8D).

Global DNA demethylation stimulated by medium switch, where LIF medium is complemented with 2i+vitC, completely impaired the dynamics of gene expression during EB formation. Most probably this happened due to the fact that the cells took longer to repress the pluripotency markers and that the mesoderm-related genes were already expressed in mESC stage. Such behavior was reported before, where long exposure to 2i molecules led to irreversible epigenetic and genomic changes which negatively impacted male mESC developmental potential (Choi et

al., 2017). These changes were represented mainly by karyotypic aberrations and, more importantly, by loss of proper DNAme at imprinting control regions.

Unexpectedly, loss of ATF7IP along with DNAme depletion was somehow enough to allow for the proper repression of pluripotency marker genes. With that, differentiation-related genes became upregulated during the time course of EB formation. However, the temporal dynamics were impaired and late stages marker genes became expressed earlier, while early stage genes were activated later. One reason for this could be that the loss of both repressive layers formed by DNAme and H3K9me3 deposition at once facilitates expression of poised genes. Besides that, it was shown that some poised genes marked by H3K27me3 and H3K4me3 in mESC have a different poised state in preadipocytes, where they are composed by H3K4/H3K9me3 bivalent chromatin. H3K9me3 deposition at those regions depends on MBD1-SETDB1-ATF7IP complex and they only get activated in later stages of adipocyte development, when this repressive complex is not present anymore and its absence allows for the binding of transcription factors (Matsumura et al., 2015). On the other hand, considering that this mechanism depends on lineage-specific gene body DNAme, release from 2i treatment must be enough to permit normal *Dnmt* activity and establishment of DNAme at those poised regions after induction of EB formation. Now, with the knowledge that prolonged 2i exposure reduces developmental potential, the experimental design could be reformulated to account for that. The cells could be kept in 2i+vitC medium for only a few passages before induction of differentiation, allowing for loss of DNAme without irreversible changes that may alter the outcome of the experiment.

4.4. SETDB1-ATF7IP complex plays a role in the regulation of gene expression

The first reports of regulation of gene expression by SETDB1 described the repression of artificial promoters in the context of KRAB-ZFP and TRIM28, where their targeting led to transcriptional inactivation and chromatin compaction (Schultz et al., 2002; Ayyanathan et al., 2003; Yang et al., 2003). Evidences for transcriptional repression of endogenous genes started arising, for instance at the p53BP2 locus (Sarraf and Stancheva, 2004; Ichimura et al., 2005), and soon later SETDB1 was implicated in control of developmental genes (Bilodeau et al., 2009; Yuan et al., 2009). Here in this work, some evidences for genome-wide gene regulation by the cooperation between SETDB1 and ATF7IP are presented. First of all, deletion of *Atf7ip* in mESC led to consistent dysregulation of a set of genes which overlapped in part with the mysregulated genes due to *Setdb1* deletion (Figure 3.6). Interestingly, upregulated genes in both phenotypes were enriched for GO terms related to meiosis and spermatogenesis, whereas between the downregulated genes several terms related to development were enriched.

Secondly, around 140 promoter regions were co-bound by SETDB1 and ATF7IP and more than 600 intra- and intergenic regions, which could be important in gene regulation, also showed common peaks (Figure 3.13H). GO term enrichment analysis revealed terms related to meiosis, spermatogenesis, gene silencing and neuronal processes (Figure 3.13F). Of note, common

binding sites were frequently associated with two distinct genes and suggests that the SETDB1-ATF7IP complex might be often controlling regulatory domains and not only single genes (Figure 3.13G). Indeed, it was recently reported that SETDB1 shields chromatin from excessive CTCF binding at the boundaries of topologically associated domains, what implicates in multiple gene regulation simultaneously (Jiang et al., 2017). CTCF is known to regulate the higher-order structure of the chromatin and its accumulation leads to increased domain insulation, what may lead to less enhancer-promoter contacts and transcriptional repression (Ong and Corces, 2014). The increased insulation might explain why so many developmental-related genes undergo downregulation in the absence of either SETDB1 or ATF7IP and implicates the latter in this newly described role of SETDB1. Supporting this, the analysis of CTCF motifs co-occupied by SETDB1 and ATF7IP reveals loss of H3K9me2/3 at these regions after loss of either protein (Figure 3.14J), similar to what was observed after *Setdb1* deletion in neurons (Jiang et al., 2017).

This work also presents some evidences for ATF7IP involvement in regulating transcription at regions of parental imprinting. Investigation of the *H19* locus showed partial loss of transcriptional repression after *Atf7ip* deletion in mESC (Figure 3.5F). Furthermore, the normal upregulation that occurs during differentiation was further enhanced in the knockout cells. This phenotype was even stronger when DNAme was depleted from mESC (Figure 3.9C) and used in differentiation experiments (Figure 3.10F). Besides that, several imprinted genes harbor common SETDB1 and ATF7IP peaks around their TSS and are enriched for SETDB1-dependent H3K9me3 mark (Figure 3.14L). Some of those genes even slightly lose H3K9me3 enrichment following *Atf7ip* deletion (Figure 3.14M). The fact that parental imprinting is often regulated by SETDB1-dependent H3K9me3 has been described by others and was shown to take place in the context of TRIM28 (Yuan et al., 2009; Quenneville et al., 2011) involving DNAme mechanisms (Cruvinel et al., 2014; Leung et al., 2014). However, this is the first time that this repressive transcriptional control has been associated with ATF7IP activity along with SETDB1.

4.5. Repression of ERV is largely dependent on SETDB1-ATF7IP complex

The accumulation of TE throughout the host genomes is responsible to shape the genomic architecture and, with that, it drives the constant evolution of eukaryotic gene-regulatory networks (Chuong et al., 2017). Due to their ability to transpose and integrate in any genomic region cells developed several mechanisms to defend themselves against endogenous retroelements, viral infection and mutations and other threats posed by possible insertions. These mechanisms depend mainly on trans-acting restriction factors which form many layers of defense either in the cytoplasm or in the nucleus. The nuclear factors usually act by epigenetically suppressing transcription and may be cell-type specific, of which the H3K9MT are good examples (Goodier, 2016). The roles of SETDB1 for the transcriptional restriction of ERV has been well studied in the past years and this protein is of extreme importance to keep the ERV genome transcriptionally silent, especially during early stages of development, when DNAme patterns are reprogrammed

(Leung and Lorincz, 2012; Groh and Schotta, 2017). Recently, ATF7IP has also been implicated in the negative control of ERV transcription in the context of TRIM28-SETDB1 in mESC, as its depletion led to similar TE derepression (Thompson et al., 2015; Yang et al., 2015). By investigating stable *Atf7ip* knockout mESC some ERV families showed transcriptional derepression (Figure 3.3A) and often overlapped with the ones derepressed in *Setdb1* KO cells (Figure 3.7A-C). The effect of compensation previously discussed (Chapter 4.3 Adaptation to *Atf7ip* deletion allows proper cell differentiation in presence of DNAme) is also observed at ERV elements, as acute KO cells show stronger upregulation of some ERV or even derepression of families which are not dysregulated in stable KO cell lines (Figure 3.8D). Similar results are observed in cells where *Atf7ip* expression was knocked down (Figure 3.8F). The mechanism of DNAme must be one of the parallel pathways which functions as a second layer of repression to the one established by the SETDB1-ATF7IP complex, considering that the phenotype was even more pronounced when the stable mutant cell lines were depleted from DNAme by transferring them to medium complemented with 2i+vitC (Figure 3.9B).

Supporting these findings, ATF7IP co-occurred with SETDB1 at several ERV families from classes I and II (Figure 3.15C). Of note, almost half of their common binding sites localized at genomic regions containing ERV annotations (Figure 3.13H). Interestingly, even though not all ERV families from classes I and II were bound by both SETDB1 and ATF7IP, the presence of both proteins correlated with transcriptional derepression in *Atf7ip* or *Setdb1* KO cells (Figure 3.15D,E). Investigation of ERV regulation at the level of individual TE elements uncovered a bias towards the binding to MLV, IAPEz and ETn retrotransposons (Figure 3.16A). When some of these elements were assessed for the position where these factors bound along their structure, the enrichment was localized mainly at 5' LTR region and, more strongly, at a subregion of the gag portion which was previously described as a signal for recruitment of heterochromatin-inducing factors and induction of compaction (Sadic et al., 2015). This was the exact region used in that work in reporter assays for exogenous retrotransposon silencing and which identified the cooperation between SETDB1, DAXX and ATRX for the silencing in mESC. Strikingly, deletion of *Atf7ip* also led to impaired silencing kinetics using this same reporter system (Figure 3.4). Thus, repression of proviral genome involves a long list of proteins which act in concert or constitute different layers of protection, even completely independent, with the same goal.

This work also provided some insights into the regulation of ERV in the context of development and cell differentiation. It is a fact that the expression of several factors controlling ERV inactivation are differentially expressed throughout development and tend to be predominantly present in germ cells, mESC and early embryos. In differentiated cells most ERV do not rely anymore on H3K9me3 and H4K20me3 for their silencing, probably due to the presence of other independent mechanisms (Rowe and Trono, 2011). One speculation is that ERV are also brought into the large organized chromatin K9 modifications (LOCKs), which are large regions of G9a-dependent H3K9me2 involved in gene silencing, highly conserved between mouse and human

and that seldom occur in undifferentiated cells, but can cover more than 30 % of the genome in differentiated cells (Wen et al., 2009). Cell differentiation experiments using control and *Atf7ip* KO mESC showed the independency of ATF7IP for ERV repression at later stages of EB formation (Figure 3.5D). Interestingly, ERV families which lost complete repression in the mutants reached the latest time point of differentiation with expression levels similar to the control. On the other hand, the elements where no derepression occurred in the mutant cells due to compensation by DNAme mechanisms reach the last time point with transcriptional levels lower than the ones in control samples. Most probably the mechanisms ensuring repression of these TE in absence of ATF7IP lead to a more robust silencing that hinders activation later in differentiated cells.

Following the same line of reasoning, when DNAme influence was precluded and mESC grown in 2i+vitC medium were submitted to differentiate into EB, derepression of some ERV happened even at an earlier time point in control cells compared to cells in normal condition without 2i+vitC (Figure 3.10D). Besides that, mutant cells devoid of DNAme, which also leads to slightly reduced H3K9me3 levels (Figure 3.16D,H), displayed an even stronger reactivation of transcription from some of the ERV families after differentiation. This indicates that ATF7IP presence limits deposition of H3K9me3 by SETDB1 at ERV in a DNAme-dependent manner, allowing for proper reactivation of those regions during differentiation. Such reactivation may be important for ERV regions to act as enhancers in different developmental contexts. These results also suggest that either DNAme mechanisms or the activity of SETDB1-ATF7IP complex is important to signalize for proper transcriptional control of TE at later stages of cellular differentiation. Of note, while repression of ERV classes I and II (Matsui et al., 2010) and of exogenous retroviral sequences (Sadic et al., 2015) is not dependent on SETDB1 in immortalized MEF, the same is not true for primary MEF. Conditional deletion of either *Setdb1* or *Atf7ip* in pMEF by expressing Cre recombinase resulted in upregulation of some ERV belonging to classes I and II (Figure 3.11). This demonstrates that the SETBD1-ATF7IP complex can still play a role in TE regulation in differentiated cells to some extent.

4.6. ATF7IP conserved domains are essential for its function as co-repressor

In various studies the structure of ATF7IP and their respective functions have been dissected. Apparently, the most relevant regions are the well conserved Domains 1 and 2, which are responsible for the interaction with different proteins (Fujita et al., 2003; Ichimura et al., 2005; Liu et al., 2009). However, some other interesting regions are present, for instance, a SIM, a NLS and an ABS. All these regions were shown to be functional as introduced before (Chapter 1.9.a Structural features), even though ABS seemed not to be required for ATF7IP repressive activity (De Graeve et al., 2000) and the SIM was only associated with ATF7IP accumulation to PML-NB so far (Sasai et al., 2013). Here in this work all those regions mentioned were assessed for their requirement to repress some of the TE families in mESC by expressing different mutants and truncations. First of all, to avoid any biases that could be introduced by the overexpression

of those mutants using lentiviral vectors and stable knockout cell lines, a platform for exchanging the expression of the endogenous loci by that of target sequences was developed in this work as previously discussed (Chapter 4.1 Establishment of tools for the investigation of SETDB1 regulation by ATF7IP). In summary, the mutant constructs are integrated via Bxb1-dependent recombination due to the presence of signal sequences in the vector containing the construct and in the *Atf7ip* locus of the knockin cell line generated.

The efficiency of recombination after co-transfection of the cells with the recombinase and the target sequence when cells are selected for the integration of an antibiotic resistance marker is very high for at least one of the alleles (Mulholland et al., 2015). However, this efficiency drops drastically when both alleles are expected to be recombined. Thus, cells were selected which contained only one allele targeted for recombination using an empty vector and constituted a knockout allele. Then, this heterozygous cell line was used for recombination to test the various mutant constructs (Figure 3.18B). Of note, the monoallelic knockout already resulted in slight derepression of most of the ERV tested (Figure 3.18E). This indicates that a biallelic expression of this locus is required for proper function. Interestingly, transcriptional compensation occurs (Eckersley-Maslin et al., 2014), considering that *Atf7ip* expression in heterozygous mutants is similar to that in control cells (Figure 3.18F). Besides that, the transcriptional levels of the Flag tag are higher in homozygous mutants than in heterozygous cells. Deletion of the second allele led to enhanced derepression of some of the ERV, whereas the wild type full-length construct was enough to keep ERV transcriptional levels similar to the heterozygous mutants.

Similar to what was observed in *Atf7ip* KO mESC lines or after acute deletion of this gene (Figure 3.8D), derepression of MusD and ERVK10C elements was more pronounced and the effect of the different mutants was clearer. Strikingly, deletion or mutation in the NLS region led to impaired repression of these elements. Thus, nuclear localization of this protein is of extreme importance for the repressive ability. Unexpectedly, mutation of the SIM region did not disrupt silencing, what indicates that binding to SUMO at least by these specific residues is not required for TE repression. A very interesting observations was that the expression of a truncated version of ATF7IP containing only the Domain 1 preceded by the NLS is enough to keep repression of the TE tested, with the exception of the MusD elements. This probably happens because of the involvement of the Domain 2 for silencing these ERV families, considering that it was the only one affected by its deletion. On the other hand, it is conflicting that the mutant lacking Domain 1 did not impair repression of the other ERV. Unfortunately, the effects on the imprinted gene tested were not clear enough to extract any meaningful conclusion.

Overall, the effect caused by acute *Atf7ip* deletion using this recombination system was not as strong as observed before (Figure 3.8D). One reason for this could be the presence of alternative transcription start sites along the endogenous locus, which is not a rare event in mammalian cells. This could permit the expression of shorter variants capable of compensating for the full-length

version of the transcript (Wang et al., 2016). Whether this expression of alternative transcriptional variants leads to isoform-specific translation would be an interesting question to be addressed in the future, especially if those variants are also detected in normal conditions.

4.7. Exogenous TE repression in the context of SETDB1 requires ATF7IP

An interesting fact is observed regarding the silencing kinetics acting over the IAPEz-gag fragment after its transduction during the reporter assay. Even though the loss of one of the components from the silencing machineries impairs repression, after several days the majority of the cells end up repressing the reporter to levels close to what is seen in the control samples. This was true for stable knockout cell lines, acute knockdown or knockout cell pools for *Setdb1, Atrx*, *Daxx* and *Trim28* (Sadic et al., 2015) and *Morc3* (Groh, Master Thesis, 2015) and was also observed for *Atf7ip* (Figure 3.4E) and those proteins again in this work (Figure 3.20A). This happens most probably due to the fact that these and several other proteins act in concert and may even be part of independent pathways.

To address the context in which ATF7IP acts to induce the silencing of the TE reporter in question, a list of possible factors was assembled based on proteins previously described to be involved in TE transcriptional regulation. In addition, the protein interactors of ATF7IP were identified by co-IP-MS in order that some of the candidates could be assessed for their contribution to the silencing pathway involving the complex formed by SETDB1 and ATF7IP. In this assay, several interesting proteins were identified as strong candidate partners (Figure 3.19C and Table 9.1). For instance, KDM1A has been implicated in transcriptional control of MERVL (Macfarlan et al., 2011) and is required for maintenance of global methylation by demethylating and stabilizing DNMT1 (Wang et al., 2009). This might predict a role for ATF7IP also at the ERV class III MERVL and could explain the derepression of this family when DNAme was depleted from *Atf7ip* KO mESC (Figure 3.10D). UHRF1 binding to ATF7IP may be the link in wild type cells between SETDB1-dependent H3K9me3 and DNMT1-mediated maintenance of DNAme (Sharif et al., 2016) for ERV silencing, whereas binding to DNMT3B may link this histone modification to the de novo DNAme machinery.

ATF7IP interaction with some of the other proteins reinforces the different contexts where the SETDB1-ATF7IP complex plays a role. Its interaction with the PcG subunit Eed might associate the change from the poised state marked by H3K27me3/H3K4me3 in pluripotent cells to the poised state marked by H3K9me3/H3K4me3 in differentiating cells (Matsumura et al., 2015). Binding to PCNA most probably happens in the context of DNA replication-coupled maintenance of H3K9me3 (Sarraf and Stancheva, 2004). Interaction with PML implicates also ATF7IP in gene expression regulation mediated by SETDB1 in PML-NB (Cho et al., 2011). Lastly, SAP18 is a component of the histone deacetylase complex and this association suggests ATF7IP to be a

component of the multiprotein complex containing mSin3A/B, HDAC1/2, SETDB1 and other proteins (Yang et al., 2003).

The genes gathered in this list were then evaluated for their genetic interaction with *Atf7ip* in the silencing of the reporter IAPEz-gag fragment. Of all the genes tested leading to impaired reporter repression, only *Setdb1* did not show genetic interaction with *Atf7ip* (Figure 3.20A,B). This result indicates that SETDB1 silences exogenous TE in the context of its complex with ATF7IP. Whether H3K9me3 is still deposited by SETDB1 in absence of ATF7IP remains unclear. This could be tested by ChIP-qPCR, however, it should be controlled by the activity of other H3K9MT. Even though TRIM28 is the anchor for targeting SETDB1 to these regions (Yang et al., 2015), its knockout enhanced impaired repression. Thus, it must recruit other downstream factors leading to SETDB1-independent H3K9me3-mediated repression, for instance, SUV39H1/2, which also plays a role in silencing ERV in mESC (Bulut-Karslioglu et al., 2014) and may become essential in the absence of SETDB1. As expected, *Daxx* and *Atrx* also showed genetic interaction with *Atf7ip* (Figure 3.20C), considering that they also further impair silencing in *Setdb1* KO mESC (Sadic et al., 2015). Interestingly, *Morc3* also showed strong genetic interaction and as it is independent of SETDB1 silencing, unraveling the pathway in which this protein acts could be an exciting topic.

4.8. SETDB1-ATF7IP complex regulates promoters and ERV by different means

In normal conditions, common SETDB1 and ATF7IP binding sites in mESC tend to not be marked by H3K9me3 and have rather H3K9ac when they occur in the vicinity of a TSS (Figure 3.14C). On the other hand, the opposite picture is observed for common peaks occurring more than 2 kb away from a TSS. Independently of being promoter regions or ERV, enrichment for H3K9me2 is always low. Thus, the epigenetic landscape of these regions is different between each another, even though they are both probably under control of SETDB1-ATF7IP complex. Consistently, many of those promoter regions lost H3K9me2 to accumulate more H3K9ac when mESC were depleted of ATF7IP, whereas the opposite trend was observed at ERV (Figure 3.14B). Most of the changes in H3K9me3 were not as strong as 2-fold. However, a cumulative plot showed that this mark became more enriched at ERV in mutant cells (Figure 3.14F), while it was slightly lost at promoters (Figure 3.14G). The most reasonable explanation is that each of these regions are regulated by different mechanisms which have SETDB1-ATF7IP complex as a common effector. As introduced previously (Chapter 1.6 Silencing of euchromatic genes), there are several mechanism governing euchromatic gene silencing which involve SETDB1. The same is true for the transcriptional repression of ERV and was also reviewed before (Chapter 1.7 Repression of transposable elements).

It was demonstrated that ATF7IP is not required for SETDB1 recruitment to ERV, as knockdown of the first does not lead to reduced enrichment of the latter (Thompson et al., 2015). In this case, ATF7IP is a downstream effector and could be dispensable for H3K9me3 deposition

by SETDB1. Interestingly, a careful inspection of their results reveals that SETDB1 not only still binds chromatin in the absence of ATF7IP, but also becomes more enriched at the ERV sites. The increased accumulation of SETDB1 in absence of ATF7IP may be the reason why stronger H3K9me3 enrichment is observed at ERV in mutant cells. Conversely, ATF7IP could be an upstream effector at promoters, which is required for SETDB1 recruitment. In concert with that is SETDB1 targeting in the context of MBD1. MBD1 is capable of binding directly to methylated DNA and interacts with ATF7IP, which in turn targets SETDB1 to gene promoters (Ichimura et al., 2005; Matsumura et al., 2015). In this context, ATF7IP-dependent recruitment of SETDB1 would result in loss of H3K9me3 in the absence of the former. This process is an attractive explanation for the observed loss of H3K9me2/3 at co-bound promoters in *Atf7ip* KO mESC. In any case, the proposed explanations would mean that ATF7IP does not have the role to stimulate SETDB1 catalytic activity in the in vivo context. Nevertheless, one could argue that accumulation of H3K9me3 at ERV in absence of ATF7IP is exclusively due to activity of other H3K9MT and that SETDB1 loses its catalytic ability in the absence of its cofactor. The enhanced accumulation of SETDB1 after ATF7IP depletion speaks against it, though. Furthermore, the strong H3K9me3 loss at ERV in *Setdb1* KO mESC makes this alternative hypothesis unlikely.

Of note, the distribution of SETDB1 and ATF7IP binding around the TSS occurring at genomic regions annotated as ERV shows that this combination tends to occur more often in the vicinity of TSS (Figure 3.13J). The involvement of TE in establishing regulatory networks for gene transcription has been already demonstrated (Rebollo et al., 2011; Rebollo et al., 2012a; Rebollo et al., 2012b). Thus, it would be interesting to see the distribution of TE around TSS to verify whether they accumulate themselves in the vicinity of TSS or there is a preference for SETDB1-ATF7IP binding to them. This would implicate SETDB1-ATF7IP complex in regulating either the heterochromatin invasion into active genes or the spreading of euchromatic marks into ERV and is supported by the chimeric transcripts observed in *Setdb1* KO mESC (Karimi et al., 2011).

4.9. Future directions

Even though this work could deepen the knowledge regarding the implications of the interaction between SETDB1 and ATF7IP to the epigenetic landscape and transcriptional regulation in mouse embryonic stem cells, some gaps still persist. For instance, it still remains unclear why the phenotype of *Atf7ip* deletion in mouse blastocysts is much stronger than the one in mESC. While these cells survive deletion, embryos do not develop longer than blastocyst stage. Thus, characterization of the in vivo phenotype is of ultimate importance. A proposed experiment is to investigate the transcription of several targets in knockout and control embryos using the digital PCR system Biomark HD (Fluidigm). With this system a large set of targets can be measured, the embryos do not need previous genotyping and single embryos can be used, excluding the requirement of pooling material.

Another question that remains unanswered is the reason for the controversial enrichment of H3K9ac at common SETDB1 and ATF7IP binding sites outside repetitive regions. From the results obtained within this thesis it was not possible to pinpoint any factor which could block SETDB1 activity. However, this may be not the influence of other proteins and may be simply the configuration of the epigenetic landscape. For instance, the strong enrichment of this active mark may inhibit catalysis of H3K9me3. Or recruitment of KDM1A by ATF7IP to these regions might constantly demethylate the H3K9. A last explanation could be that the physical presence alone is what determines the regulation of these regions. A way to start addressing this could be by using a reporter system where mutant SETDB1 lacking its methyltranferase activity is targeted to these regions using a version of this protein which is fused to a catalytically dead Cas9.

Even with all the efforts from this work it is still not clear whether the de novo establishment of the H3K9me3 mark by SETDB1 depends on the presence of ATF7IP. One suggestion of an experimental design is to transiently deplete SETDB1 from mESC by using an inducible shRNA system. After a few days and before it becomes lethal, enough for reduction of most of the SETDB1-dependent H3K9me3, *Atf7ip* deletion is induced in a conditional knockout cell line and the reestablishment of this mark can be evaluated in comparison to *Atf7ip*-positive cells.

In summary, this work contributed to the better understanding of the roles of the SETDB1-ATF7IP complex and the regulation of their functional targets. Some insights in the molecular mechanisms of SETDB1 activity regulation by its cofactor ATF7IP were also provided. Besides that, several tools for further studying the interaction of these proteins were developed and are available for future experiments.

5. MATERIAL

5.1. Antibodies

Target	ID	Weight	Host	Type	Company	Catalogue	Purpose	Dilution
a-Tubulin	220	50 kDa	mouse	monoclonal	Sigma Aldrich	B-5-1-2	Western blot	1000
Atf7ip	258	240 kDa	rabbit	polyclonal	Sigma Aldrich	HPA023505	Western blot	2000
							ChIP-qPCR	5 µg
Flag	132	-	mouse	monoclonal	Sigma Aldrich	F1804	ChIP-Seq	5 µg
							Western blot	2000
H3K9ac	275	17 kDa	rabbit	polyclonal	Active Motif	39138	ChIP-qPCR	3 µg
H3K9me1	-	17 kDa	mouse	monoclonal	Kimura	CMA316	Western blot	2000
H3K9me2	-	17 kDa	mouse	monoclonal	Kimura	CMA317	Western blot	2000
H3K9me2	255	17 kDa	mouse	monoclonal	Abcam	ab1220	ChIP-qPCR	3 µg
H3K9me3	-	17 kDa	mouse	monoclonal	Kimura	CMA318	Western blot	2000
H3K9me3	249	17 kDa	rabbit	polyclonal	Active Motif	39161.39	ChIP-qPCR	3 µg
Lamin B1	251	70 kDa	rabbit	polyclonal	Active Motif	39095	Western blot	2000
Setdb1	28	180 kDa	mouse	monoclonal	Elizabeth Kremmer	2D2-111	Western blot	5
Suz12	253	83 kDa	rabbit	monoclonal	Cell signaling	3737	Western blot	1000
Trim28	19	110 kDa	rabbit	polyclonal	Abcam	ab10484-50	Western blot	2000

5.2. Data and analyses

All raw data, processed data and analyses, together with the scripts used for processing and analyzing the data are stored in the internal server and are available upon request to Prof. Dr. Gunnar Schotta (gunnar.schotta@med.uni-muenchen.de).

5.3. High-throughput sequencing libraries

Library Name	ID	Experiment	Antibody	Sequencing	Index
RNA_ES_wt26_ct25-1	GS114	RNA-Seq	none	50 bp pair-end	CGATGT
RNA_ES_wt26_ct25-2	GS115	RNA-Seq	none	50 bp pair-end	TTAGGC
RNA_ES_wt26_ct25-3	GS116	RNA-Seq	none	50 bp pair-end	TGACCA
RNA_ES_wt26_ko25-1	GS117	RNA-Seq	none	50 bp pair-end	ACAGTG
RNA_ES_wt26_ko25-2	GS118	RNA-Seq	none	50 bp pair-end	GCCAAT
RNA_ES_wt26_ko25-3	GS119	RNA-Seq	none	50 bp pair-end	CAGATC
ChIP_ES_wt26_input	GS133	ChIP-Seq	none	50 bp pair-end	ATCACG
ChIP_ES_wt26_ki13-1_Flag-Atf7ip	GS134	ChIP-Seq	Flag	50 bp pair-end	TTAGGC
ChIP_ES_A9_k9g1_Flag-Setdb1	GS180	ChIP-Seq	Flag	50 bp pair-end	CTTGTA
ChIP_ES_ct25-1_input	GS232	ChIP-Seq	none	50 bp pair-end	CGATGT
ChIP_ES_ko25-1_input	GS233	ChIP-Seq	none	50 bp pair-end	TTAGGC
ChIP_ES_ct25-1_H3K9me3	GS234	ChIP-Seq	H3K9me3	50 bp pair-end	TGACCA
ChIP_ES_ko25-1_H3K9me3	GS235	ChIP-Seq	H3K9me3	50 bp pair-end	ACAGTG
ChIP_ES_ct25-1_H3K9me2	GS236	ChIP-Seq	H3K9me2	50 bp pair-end	GCCAAT
ChIP_ES_ko25-1_H3K9me2	GS237	ChIP-Seq	H3K9me2	50 bp pair-end	CAGATC
ChIP_ES_ct25-1_H3K9ac	GS238	ChIP-Seq	H3K9ac	50 bp pair-end	ACTTGA
ChIP_ES_ko25-1_H3K9ac	GS239	ChIP-Seq	H3K9ac	50 bp pair-end	GATCAG

5.4. Oligonucleotides

Target	Direction	ID	Sequence	Purpose
Atf7ip ko check	fw	GS2775	ATCGATAAGCTTGATTGGATAGTGTAGAAGAACCTCAGA	genotyping
	rw	GS2776	CTGCAGGAATTCGATGAGTCAGCTCACCAGAAGCC	
attL	fw	GS4051	CCGGCTTGTCGACGACG	genotyping
Flag-Atf7ip ki check	fw	GS3542	TGGGGAAACAGTACCGGATCGAG	genotyping
	rw	GS3521	GGCTTGCTTTATGGGCTGAATTT	
Min-Atf7ip ki check	fw	GS3999	AATATGAGAACTCCGACCGAG	genotyping
	rw	GS3915	GACTCTGTGCACAGCATCGA	
Min-Atf7ip ki check 2	fw	GS4002	TGGGTTTGTCTGGTCAACCA	genotyping
	rw	GS3797	GCACAATCATCCAAGGCTGAC	
ETn/MusD	fw	GS3752	ATAGAGGCCGCTTCTTTGC	qPCR
	rw	GS3753	TGAGACTCCACCAAATGTCC	
Gapdh	fw	GS3050	CTCTGCTCCTCCCTGTTCC	qPCR
	rw	GS3051	TCCCTAGACCCGTACAGTGC	
H19	fw	GS928	AGCTTTGAGTACCCCAGGTTCA	qPCR
	rw	GS929	GCCTCTGCTTTTATGGCTATGG	
IAP	fw	GS2512	CGGGTCGCGGTAATAAAGGT	qPCR
	rw	GS2513	ACTCTCGTTCCCCAGCTGAA	
Nnat	fw	GS1652	TGCTGCTGCAGGTGAGTATGTA	qPCR
	rw	GS1653	TTGCGGCAATTGGGATAGGA	
Ptch1	fw	GS1689	TGCTGATGCATGTGAGTGTG	qPCR
	rw	GS1690	AAGCCGACAAAGCTTACAAATG	
Tia1	fw	GS477	GCTCGCCGCCATCTTGGAT	qPCR
	rw	GS478	GGCTATGGCTGCGGAAGAGC	
Atf7ip (exon 2)	fw	GS3794	TGATAACCCAGGCTGTGGTAC	RT-qPCR
	rw	GS3795	TGTACTGTAGTGGCCAGCTC	
Atf7ip (exon 3)	fw	GS2975	GCTCACCCTCGAAACAAGAAAG	RT-qPCR
	rw	GS2976	TCAGCTTTGCCGTCCTTCTC	
CD55	fw	GS2152	CAAGTACAGGAACCCCCTCA	RT-qPCR
	rw	GS2153	CCACCTGTGTTAGGCTCTCC	
Cre	fw	GS31.1	CATCGTCGGTCCGGGCTGCC	RT-qPCR
	rw	GS31.2	CCCCCAGGCTAAGTGCCTTC	
Dnmt1	fw	GS1540	GCTTCAACTCCCGCACTTAC	RT-qPCR
	rw	GS1541	CACGAAGTTCCTGACGTTCTC	
ETn/MusD	fw	GS3752	ATAGAGGCCGCTTCTTTGC	RT-qPCR
	rw	GS3753	TGAGACTCCACCAAATGTCC	
Foxa2	fw	GS2757	TGGCTGCAGACACTTCCTAC	RT-qPCR
	rw	GS2758	AAGCTCTCCCAAAGTCTCCA	
Gapdh	fw	GS276.1	TCAAGAAGGTGGTGAAGCAG	RT-qPCR
	rw	GS276.2	GTTGAAGTCGCAGGAGACAA	
H19	fw	GS809	GGAGACTAGGCCAGGTCTC	RT-qPCR
	rw	GS810	GCCCATG GTGTTCAAGAAGGC	
H19	fw	GS3246	AGTCGATTGCACTGGTTTGG	RT-qPCR
	rw	GS3247	GCCAAAGAGGTTTACACACTCG	
Hnf1b	fw	GS3062	AGGGAGGTGGTCGATGTCA	RT-qPCR
	rw	GS3063	TCTGGACTGTCTGGTTGAACT	

Target	Direction	ID	Sequence	Purpose
Hprt	fw	GS279.1	ATGAGCGCAAGTTGAATCTG	RT-qPCR
	rw	GS279.2	CAGATGGCCACAGGACTAGA	
IAP	fw	GS2512	CGGGTCGCGGTAATAAAGGT	RT-qPCR
	rw	GS2513	ACTCTCGTTCCCCAGCTGAA	
IAPEz	fw	GS3908	CACGCTCCGGTAGAATACTTACAAAT	RT-qPCR
	rw	GS3909	CCTGTCTAACTGCACCAAGGTAAAAT	
Igf2r	fw	GS781	GCTGCCAGCCTTCAGATTCACA	RT-qPCR
	rw	GS782	TTCTTGCAGGCTGCAGTAGTCCTC	
MERVL	fw	GS3455	CTTCCATTCACAGCTGCGACTG	RT-qPCR
	rw	GS3456	CTAGAACCACTCCTGGTACCAAC	
Mixl	fw	GS3040	AACCGACGGGCCAAGTC	RT-qPCR
	rw	GS3041	TCCCCGCCTTGAGGATAAG	
MLV	fw	GS2651	TGGGTCCTCAGGGTCATAAG	RT-qPCR
	rw	GS2652	CAGTTGCTCTTAGCGGGTCT	
MMERVK10C	fw	GS3754	GATTGTCAGCAGCTCTTGCA	RT-qPCR
	rw	GS3755	AAAGCCTCGTCCACAAGGTT	
Nanog	fw	GS534	AACCAAAGGATGAAGTGCAAGCG	RT-qPCR
	rw	GS535	GCTGCAATGGATGCTGGGATACT	
Nodal	fw	GS3060	TTCAAGCCTGTTGGGCTCTAC	RT-qPCR
	rw	GS3061	TCCGGTCACGTCCACATCTT	
Oct4	fw	GS410.1	GCTCACCCTGGGCGTTCTC	RT-qPCR
	rw	GS410.2	GGCCGCAGCTTACACATGTTC	
Reep6	fw	GS2144	ATACCCCGCATATGCTTCAG	RT-qPCR
	rw	GS2145	ACTTGCCCGCGTAGTAGAAA	
Rex2	fw	GS2032	TGTTTGCTCCTCACCAAGAA	RT-qPCR
	rw	GS2033	CCTCCGATGAGAAGTCCAAA	
Setdb1	fw	GS958	AGCAGAACTCCAAAAGACCAGAAGC	RT-qPCR
	rw	GS959	TCTTGCCCAGAATCCGCATG	
T	fw	GS2761	GCTTCAAGGAGCTAACTAACGAG	RT-qPCR
	rw	GS2762	CCAGCAAGAAAGAGTACATGGC	
VL30	fw	GS3457	TTCGCCTCTGCAATCAAGCTCTC	RT-qPCR
	rw	GS3458	TCGCTCRTGCCTGAAGATGTTTC	
WPRE	fw	GS3711	CCGTTGTCAGGCAACGTG	RT-qPCR
	rw	GS3712	AGCTGACAGGTGGTGGCAAT	

Target	ID fw	ID rw	Sequence	Purpose
sgAtf7ip	GS3822	GS3823	CACCGTCTTCATCCATATATCGC	sgRNA
sgAtf7ip-1	GS2767	GS2768	GTCAGAGGCTGGGTTATCGG	sgRNA
sgAtf7ip-2	GS2769	GS2770	GCCACTACAGTACAGGCTAC	sgRNA
sgAtf7ip-ABS-1	GS3814	GS3815	GTCGTCTTCACTCTTGCC	sgRNA
sgAtf7ip-ABS-2	GS3816	GS3817	GAGTCTTTACTTTGAATT	sgRNA
sgAtf7ip-ATG	GS3888	GS3889	GCTCAGAAAAAGTCTTCA	sgRNA
sgAtf7ip-Domain1-1	GS3820	GS3821	GCCATATATCGCCGGCGTT	sgRNA
sgAtf7ip-Domain1-2	GS3822	GS3823	GTCTTCATCCATATATCGC	sgRNA
sgAtf7ip-Domain1-3	GS3824	GS3825	GCAGTGCTCACCGAGCTGC	sgRNA
sgAtf7ip-Domain2-1	GS3830	GS3831	GTCGTGCTAACTTCAAGTG	sgRNA

Target	ID fw	ID rw	Sequence	Purpose
sgAtf7ip-Domain2-2	GS3832	GS3833	GCAGCTTCGGTCCACTTCC	sgRNA
sgAtf7ip-NLS	GS3818	GS3819	GTTCTAAGTCAGAAGACA	sgRNA
sgAtf7ip-SIM-1	GS3826	GS3827	GTGTCATTGATCTCACGA	sgRNA
sgAtf7ip-SIM-2	GS3828	GS3829	GATGGACGATGAAGAGAG	sgRNA
sgAtf7ip-Start1	GS2767	GS2768	GTCAGAGGCTGGGTTATCGG	sgRNA
sgAtf7ip-Start2	GS2769	GS2770	GCCACTACAGTACAGGCTAC	sgRNA
sgAtf7ip-stop1	GS3390	GS3392	GTCTAGTGCAGGCACCTGACA	sgRNA
sgAtf7ip-stop2	GS3392	GS3393	GGAGGGTCTTGTCAATCTGGA	sgRNA
sgAtf7ip2-Domain1	GS3836	GS3837	GAAAATGCTTGTGGGGCTC	sgRNA
sgAtf7ip2-Domain2	GS3840	GS3841	GCACAGTTGGGGTTGACCT	sgRNA
sgAtf7ip2-SIM	GS3838	GS3839	GCCCGTCCAACTACAGCAT	sgRNA
sgAtf7ip2-Start	GS3834	GS3835	GATTCTAGGATATGAAATA	sgRNA
sgAtrx	GS2621	GS3106	GCTGTTGCACGCAGTCACCAAGTCCAGTAG	sgRNA
sgAtrx	GS3106	GS3107	GACAACTCCTTTCGACCA	sgRNA
sgDaxx	GS2659	GS3108	AGTACAATGATGCTGTCATCGG	sgRNA
sgDaxx	GS3108	GS3109	GTACAATGATGCTGTCAT	sgRNA
sgDnmt1	GS3526	GS3527	GCTTCGTGAAGTGAGCCGTGA	sgRNA
sgDnmt3a	GS3965	GS3966	GACGATGAGCCTGAGTATG	sgRNA
sgDnmt3b	GS3967	GS3968	GGAAACAGCTTCCCGGCA	sgRNA
sgDnmt3l	GS3969	GS3970	GAAGTCAAAGTGAACCGA	sgRNA
sgEed	GS3979	GS3980	GCGTATTTGTGGGCGTGTC	sgRNA
sgEzh2	GS3981	GS3982	GAACTTCATCCCCCATATA	sgRNA
sgHdac1	GS3973	GS3974	GTACCGACAGAGCCTCCCG	sgRNA
sgHdac2	GS3975	GS3976	GGGGCTGTGAAATTAAAC	sgRNA
sgHnrnpk	GS3989	GS3990	GACCTACCTCTTCCAAGGT	sgRNA
sgHnrnpl	GS3987	GS3988	GCCTGAACTCCGTTCTTC	sgRNA
sgKdm1a	GS3971	GS3972	GTGGCTTCAAACGTCAGC	sgRNA
sgMbd1	GS3960	GS3961	GTTGTGCAAAGATTGTCG	sgRNA
sgMcm5	GS3995	GS3996	GAGTGGGCACCGATCGCA	sgRNA
sgMorc3	GS3439	GS3440	GTGAATGCTAAACAGATC	sgRNA
sgNeg	GS3126	GS3127	GTTTGGCTCTACAAAGGC	sgRNA
sgPcgf2	GS3985	GS3986	GAGGCGAGGTCCTGGAAC	sgRNA
sgPml	GS3977	GS3978	GCAGCGCATCCGCACTAG	sgRNA
sgPspc1	GS3993	GS3994	GCGAAACCTCCTGCACGAA	sgRNA
sgSetdb1	GS3110	GS3111	GCTGAGCTGCAGCAGGCGG	sgRNA
sgSnrnp200	GS3991	GS3992	GTCCGGTCAATGAGAGAG	sgRNA
sgSuz12	GS3983	GS3984	GCTTCGGGGGTTCGGCGG	sgRNA
sgTrim28	GS3112	GS3113	GCCGCGTCGTCCCCTGCG	sgRNA
sgUhrf1	GS3528	GS3529	GTCACAGTGCGAGCACGAGCA	sgRNA
shAtf7ip-1	GS2777	GS2778	CGAGGCCCAATACAGATGAAA	shRNA
shAtf7ip-2	GS2779	GS2780	GACGAACACTTAGACCAAATT	shRNA
shScrambled	GS1450	GS1451	CAACAAGATGAAGAGCACCAA	shRNA

5.5. Organisms

Line	Organism	Type	Source	Parent	Generation	Plasmid
E126-5	Mus musculus	Setdb1 p/p pMEF	Gunnar Schotta	-	-	-

Line	Organism	Type	Source	Parent	Generation	Plasmid
E699-1	Mus musculus	Atf7ip p/p pMEF	this work	-	-	-
E700-5	Mus musculus	Atf7ip p/p pMEF	this work	-	-	-
E701-1	Mus musculus	Atf7ip p/p pMEF	this work	-	-	-
wt26	Mus musculus	wild type mESC	Gunnar Schotta	-	-	-
k9g1	Mus musculus	Flag-Setdb1 KI mESC	Gunnar Schotta	A9	electroporation	p902
T37	Homo sapiens	HeLa mCat	Gunnar Schotta	HeLa	transduction	p1055
min5h-1	Mus musculus	Min-Atf7ip/+ KI mESC	this work	min5	transfection	p1584, p1615
ct25-1	Mus musculus	control mESC	this work	wt26	transfection	p963
ct25-2	Mus musculus	control mESC	this work	wt26	transfection	p963
ct25-3	Mus musculus	control mESC	this work	wt26	transfection	p963
ki13-1	Mus musculus	Atf7ip KI mESC	this work	wt26	transfection	p1467, p1516
ki13-1	Mus musculus	Flag-Atf7ip KI mESC	this work	wt26	transfection	p1467, p1516
ki13-2	Mus musculus	Flag-Atf7ip KI mESC	this work	wt26	transfection	p1467, p1516
ko25-1	Mus musculus	Atf7ip KO mESC	this work	wt26	transfection	p963, p1309, p1310
ko25-2	Mus musculus	Atf7ip KO mESC	this work	wt26	transfection	p963, p1309, p1310
ko25-3	Mus musculus	Atf7ip KO mESC	this work	wt26	transfection	p963, p1309, p1310
min10	Mus musculus	Min-Atf7ip KI mESC	this work	wt26	transfection	p1597, p1646
min2	Mus musculus	Min-Atf7ip KI mESC	this work	wt26	transfection	p1597, p1646
min5	Mus musculus	Min-Atf7ip KI mESC	this work	wt26	transfection	p1597, p1646
min8	Mus musculus	Min-Atf7ip KI mESC	this work	wt26	transfection	p1597, p1646
T89	Mus musculus	Cas9+ mESC	Gunnar Schotta	wt26	transduction	p1368

5.6. Plasmids

Name	ID	Marker	Resistance	Source	Purpose
pCAG/NLS-HA-Bxb1	1584	none	ampicilin	Addgene (51271)	Bxb1 recombination
pattB/3xFLAG-HA-Atf7ip(dD1)pA-PuroR	1607	puroR	kanamycin	this work	Bxb1 recombination
pattB/3xFLAG-HAAtf7ip(dD2)pA-PuroR	1608	puroR	kanamycin	this work	Bxb1 recombination
pattB/3xFLAG-HA-Atf7ip(dSIM)pA-PuroR	1609	puroR	kanamycin	this work	Bxb1 recombination
pattB/3xFLAG-HA-Atf7ip(dABS)pA-PuroR	1610	puroR	kanamycin	this work	Bxb1 recombination
pattB/3xFLAG-HA-Atf7ip(dNLS)pA-PuroR	1611	puroR	kanamycin	this work	Bxb1 recombination
pattB/3xFLAG-HA-Atf7ip(D1)pA-PuroR	1612	puroR	kanamycin	this work	Bxb1 recombination
pattB/3xFLAG-HA-Atf7ip(dD1+NLS)pA-PuroR	1613	puroR	kanamycin	this work	Bxb1 recombination
pattB/3xFLAG-HAPoly(A)-PuroR	1614	puroR	kanamycin	this work	Bxb1 recombination
pattB/3xFLAG-HAPoly(A)-ZeoR	1615	zeoR	kanamycin	this work	Bxb1 recombination
pattB/3xFLAG-HA-Atf7ippA-PuroR	1649	puroR	kanamycin	this work	Bxb1 recombination
pBSII-SK+/5'-Atf7ip-Flag-Stop-3'Atf7ip	1467	none	ampicilin	this work	CRISPR knockin
pX330/sgAtf7ip-stop	1516	none	ampicilin	this work	CRISPR knockin
pX459/sgAtf7ip-Start	1597	puroR	ampicilin	this work	CRISPR knockin
pBSII-SK+/5'-Atf7ip-Start-Min-3'-Atf7ip	1646	none	ampicilin	this work	CRISPR knockin
pX330/U6-Chimeric_BB-CBh-hSpCas9	1238	none	ampicilin	Addgene (42230)	CRISPR knockout
pX330/sgATRX-2	1244	none	ampicilin	Gunnar Schotta	CRISPR knockout
pX330/sgDAXX-4	1271	none	ampicilin	Gunnar Schotta	CRISPR knockout
pX330/sgAtf7ip-1	1309	none	ampicilin	this work	CRISPR knockout
pX330/sgAtf7ip-2	1310	none	ampicilin	this work	CRISPR knockout
pX459v2/EF1a-pSpCas9(BB)-2A-Zeo	1650	zeoR	ampicilin	Heinrich Leonhardt	CRISPR knockout
pX459v2/sgNeg	1652	zeoR	ampicilin	this work	CRISPR knockout

Name	ID	Marker	Resistance	Source	Purpose
pX459v2/sgDaxx	1653	zeoR	ampicilin	this work	CRISPR knockout
pX459v2/sgSetdb1	1654	zeoR	ampicilin	this work	CRISPR knockout
pX459v2/sgAtrx	1655	zeoR	ampicilin	this work	CRISPR knockout
pX459v2/sgAtf7ip	1656	zeoR	ampicilin	this work	CRISPR knockout
pX459v2/sgDnmt1	1657	zeoR	ampicilin	this work	CRISPR knockout
pX459v2/sgMbd1	1658	zeoR	ampicilin	this work	CRISPR knockout
pX459v2/sgMorc3	1659	zeoR	ampicilin	this work	CRISPR knockout
pX459v2/sgTrim28	1660	zeoR	ampicilin	this work	CRISPR knockout
pX459v2/sgUhrf1	1661	zeoR	ampicilin	this work	CRISPR knockout
pX459v2/sgDnmt3a	1662	zeoR	ampicilin	this work	CRISPR knockout
pX459v2/sgDnmt3b	1663	zeoR	ampicilin	this work	CRISPR knockout
pX459v2/sgDnmt3l	1664	zeoR	ampicilin	this work	CRISPR knockout
pX459v2/sgKdm1a	1665	zeoR	ampicilin	this work	CRISPR knockout
pX459v2/sgHdac1	1666	zeoR	ampicilin	this work	CRISPR knockout
pX459v2/sgHdac2	1667	zeoR	ampicilin	this work	CRISPR knockout
pX459v2/sgPml	1668	zeoR	ampicilin	this work	CRISPR knockout
pX459v2/sgEed	1669	zeoR	ampicilin	this work	CRISPR knockout
pX459v2/sgEzh2	1670	zeoR	ampicilin	this work	CRISPR knockout
pX459v2/sgSuz12	1671	zeoR	ampicilin	this work	CRISPR knockout
pX459v2/sgPcgf2	1672	zeoR	ampicilin	this work	CRISPR knockout
pX459v2/sgHnrnpl	1673	zeoR	ampicilin	this work	CRISPR knockout
pX459v2/sgHnrnpk	1674	zeoR	ampicilin	this work	CRISPR knockout
pX459v2/sgSnrnp200	1675	zeoR	ampicilin	this work	CRISPR knockout
pX459v2/sgPspc1	1676	zeoR	ampicilin	this work	CRISPR knockout
pX459v2/sgMcm5	1677	zeoR	ampicilin	this work	CRISPR knockout
pLenti6/EFEGT-neo-GAG2.22	1074	neoR	ampicilin	Gunnar Schotta	lentiviral expression
pLenti6/EF1a-CRE-IRES-PURO	1087	puroR	ampicilin	Gunnar Schotta	lentiviral expression
pLenti6/hPGK-puro-ccdB	1348	puroR	ampicilin	Gunnar Schotta	lentiviral expression
pLenti6/hPGK-puro-sgNeg	1349	puroR	ampicilin	Gunnar Schotta	lentiviral expression
pLenti6/hPGK-puro-sgDaxx	1351	puroR	ampicilin	Gunnar Schotta	lentiviral expression
pLenti6/hPGK-puro-sgSetdb1	1353	puroR	ampicilin	Gunnar Schotta	lentiviral expression
pLenti6/hPGK-puro-sgAtf7ip-Start1	1439	puroR	ampicilin	this work	lentiviral expression
pLenti6/hPGK-puro-sgAtf7ip-Start2	1440	puroR	ampicilin	this work	lentiviral expression
pLenti6/hPGK-puro-sgDnmt1	1449	puroR	ampicilin	Gunnar Schotta	lentiviral expression
pLenti6/hPGK-puro-sgAtf7ip-ABS-1	1565	puroR	ampicilin	this work	lentiviral expression
pLenti6/hPGK-puro-sgAtf7ip-ABS-2	1566	puroR	ampicilin	this work	lentiviral expression
pLenti6/hPGK-puro-sgAtf7ip-NLS	1567	puroR	ampicilin	this work	lentiviral expression
pLenti6/hPGK-puro-sgAtf7ip-Domain1-1	1568	puroR	ampicilin	this work	lentiviral expression
pLenti6/hPGK-puro-sgAtf7ip-Domain1-2	1569	puroR	ampicilin	this work	lentiviral expression
pLenti6/hPGK-puro-sgAtf7ip-Domain1-3	1570	puroR	ampicilin	this work	lentiviral expression
pLenti6/hPGK-puro-sgAtf7ip-SIM-1	1571	puroR	ampicilin	this work	lentiviral expression
pLenti6/hPGK-puro-sgAtf7ip-SIM-2	1572	puroR	ampicilin	this work	lentiviral expression
pLenti6/hPGK-puro-sgAtf7ip-Domain2-1	1573	puroR	ampicilin	this work	lentiviral expression
pLenti6/hPGK-puro-sgAtf7ip-Domain2-2	1574	puroR	ampicilin	this work	lentiviral expression
pLenti6/hPGK-puro-sgAtf7ip2-Start	1575	puroR	ampicilin	this work	lentiviral expression
pLenti6/hPGK-puro-sgAtf7ip2-Domain1	1576	puroR	Respicilin	this work	lentiviral expression
pLenti6/hPGK-puro-sgAtf7ip2-SIM	1577	puroR	ampicilin	this work	lentiviral expression

Name	ID	Marker	Resistance	Source	Purpose
pLenti6/hPGK-puro-sgAtf7ip2-Domain2	1578	puroR	ampicilin	this work	lentiviral expression
pX459/sgNeg	1618	puroR	ampicilin	this work	lentiviral expression
pX459/sgDaxx	1620	puroR	ampicilin	this work	lentiviral expression
pX459/sgAtf7ip	1623	puroR	ampicilin	this work	lentiviral expression
pX459/sgAtf7ip-2	1624	puroR	ampicilin	this work	lentiviral expression
pBS/U6-sgAtf7ip	1647	none	ampicilin	this work	lentiviral expression
pBS/U6-sgAtf7ip-2	1648	none	ampicilin	this work	lentiviral expression
pLenti6/EFEGT-neo	940	G418	ampicilin	Gunnar Schotta	lentiviral expression
pLenti6/EF1a-3xFLAG-IRES-PURO	963	puroR	ampicilin	Gunnar Schotta	lentiviral expression
pLKO1mod/shAtf7ip-1	1315	puroR	ampicilin	this work	lentiviral knockdown
pLKO1mod/shAtf7ip-2	1316	puroR	ampicilin	this work	lentiviral knockdown
pLKO1mod/shScrambled	849	puroR	ampicilin	Gunnar Schotta	lentiviral knockdown
psPAX2	183	none	ampicilin	Didier Trono	lentivirus packaging
pLP/eco-env	811	puroR	ampicilin	Gunnar Schotta	lentivirus packaging

6. METHODS

6.1. Cell culture

Bacteria strains were grown either on LB Agar plates or in LB medium at 37 °C and 160 rpm with proper antibiotics.

Feeder independent mESC were cultured at 37 °C and 5 % CO_2 in high glucose DMEM with L-glutamine (Sigma Aldrich) supplemented with 15 % v/v FCS (batch tested for mESC), 1x MEM non-essential aminoacids, 1x Pen/Strep, 0.175 mM β-mercaptoethanol (Sigma Aldrich) and 0.4 % v/v LIF (2.10^3 U/mL). Dishes suitable for tissue culture were coated with 1x PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 2 mM KH_2PO_4) 0.2 % w/v gelatin for 20 min before seeding cells. mESC were maintained at naïve state by adding 3 µM CHIR and 1 µM PD together with 50 mg of L-ascorbic acid 2-phosphate sesquimagnesium salt hydrate (Sigma Aldrich).

pMEF were cultured in mESC medium without LIF at 37 °C, 5 % CO_2, 3 % O_2. iMEF, Hela and HEK293T were cultured at 37 °C, 5 % CO_2 in high glucose DMEM with L-glutamine (Sigma Aldrich) supplemented with 10 % v/v FCS, 1x MEM non-essential aminoacids, 1x Pen/Strep, 0.175 mM β-mercaptoethanol (Sigma Aldrich). HEK293T cell were grown to a maximum of fifteen passages and the medium was always complemented with 500 µg/mL of G418 (PAA Laboratories).

For passaging of all adherent mammalian cells, they were first washed with 1x PBS. After that, they were trypsinized with 1x Trypsin/EDTA in 1x PBS for 5 min. Trypsin was neutralized with the proper fresh medium and cells were pelleted 3 min at 400 g and resuspended in fresh growth medium. Cell number was counted using CASY Counter Model TT (Roche Innovatis). Cell stocks were frozen in medium specific to each cell line containing 10 % v/v DMSO and 40 % v/v FCS using Nalgene Cryo 1 °C freezing containger (Thermo Fischer Scientific).

6.2. Chromatin immunoprecipitation of histone modifications

After harvesting, chromatin of 3 million cells was crosslinked under mild agitation for 10 min at RT in 1 ml of 1 % v/v methanol-free formaldehyde (Thermo Fischer Scientific) in mESC medium. Formaldehyde was quenched under mild agitation for 5 min at RT with 125 mM of 2.5 M glycine in 1x PBS. Cells were washed twice with 1x PBS 10 % v/v FCS after pelleting for 3 min at 400 g and cell pellets were snap-frozen or freshly processed. Frozen cell pellets were resuspended in 130 µL of lysis buffer 10 mM Tris/HCl pH 7.5, 140 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1 % w/v Triton X-100, 0.1 % w/v Na-Deoxycholate, 0.1 % w/v SDS, 1x cOmplete, EDTA-free Protease Inhibitor Cocktail (Roche Diagnostics) and incubated for 5 min on ice. Samples were transferred to microTUBE AFA Fiber Pre-Slit Snap-Cap 6x16 mm and sonicated for 25 min at 4 °C using Covaris S220 series Focused.ultrasonicator (Thermo Fischer Scientific) with the following settings: duty factor of 2 %, peak incident power of 105 W, 200 cycles per burst and continuous

degassing. SDS of the lysis buffer was diluted in 870 µL of ChIP buffer (50 mM Tris/HCl pH 8.0, 10 mM EDTA, 1 % w/v SDS, 1x cOmplete, EDTA-free Protease Inhibitor Cocktail). Diluted sheared chromatin was centrifuged for 15 min at 4 °C and 17000 g and divided in 6 aliquots of 150 µL (chromatin of 450 thousand cells) and either snap-frozen or used freshly.

In a 0.2 mL PCR tube, 11 µL of magnetic Dynabeads Protein G (Thermo Fischer Scientific) were washed 2 times with ChIP buffer. A total amount of 3 µg of antibody was bound to the beads for 2 h at 4 °C and 35 rpm vertically. Beads-Ab complexes were washed once with ChIP buffer and incubated with 150 µL of sheared chromatin for 4 h at 4 °C and 35 rpm vertically. Immunoprecipitated chromatin was washed 4 times with ChIP buffer and once with washing buffer (10 mM Tris/HCl pH 8.0, 10 mM EDTA) for 10 min at 4 °C and 35 rpm vertically. Chromatin was eluted from the beads-Ab complexes with 70 µL of elution buffer (0.5 % SDS, 300 mM NaCl, 5 mM EDTA, 10 mM Tris HCl pH 8.0) containing 0.3 µg/mL RNase A (Thermo Fischer Scientific) for 30 min at 37 °C and 900 rpm horizontally. Proteins were digested with 0.6 µg/µL proteinase K (Thermo Fischer Scientific) for 2 h at 55 °C and 900 rpm horizontally. Crosslink was reversed by incubating samples for 8 h at 65 °C and 900 rpm horizontally. Supernatant was transferred to a new tube and another step of elution was carried out with 30 µl of elution buffer for 1 min. After combining eluates, remaining proteins were digested with 0.2 µg/µL proteinase K (Thermo Fischer Scientific) for 1 h at 55 °C and 600 rpm horizontally. Purification of DNA was performed using Agencourt AMPure XP beads (Beckman Coulter) at a sample-to-beads ratio of 1:2, according to the manufacturer's instructions, eluting in 12 µL of 10 mM Tris/HCl pH 8.5. For ChIP-qPCR, enrichment on specific regions was analyzed by qPCR using the LightCycler 480 Real Time PCR system (Roche Diagnostics).

6.3. Chromatin immunoprecipitation of transcription factors

The ChIP protocol for transcription factors is a modification of the one for histone modifications. The starting material was 25 million cells and the fixation procedure was scaled up proportionally and performed in 1x PBS 10 % v/v FCS (not in medium). The fresh pellets of fixed cells were resuspended in 10 mL of lysis buffer 1 (50 mM Hepes-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 10 % v/v glycerol, 0.5 % v/v NP-40, 0.25 % v/v Triton X-100) and the samples were left rocking gently at 4 °C for 10 min. After centrifuging samples for 5 min at 4 °C and 1350 g, nuclei were extracted by resuspending cell pellets in 10 mL of lysis buffer 2 (10 mM Tris/HCl pH 7.5, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA) and rocking gently at RT for 10 min. Nuclei were pelleted and resuspended in 1 mL of shearing buffer (10 mM Tris/HCl pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1 % w/v Na-Deoxycholate, 1 % w/v SDS) and incubated on ice for 5 min. Both lysis buffers and the shearing buffer were complemented with 1x cOmplete, EDTA-free Protease Inhibitor Cocktail (Roche Diagnostics) right before use. Samples were transferred to tube AFA Fiber & Cap 12x12 mm and sonicated for 25 min at 4 °C using the following settings: duty factor of 20 %, peak incident power of 140 W, 200 cycles per burst and continuous degassing. After

adding 110 µL of 10 % v/v Triton X-100, sheared chromatin was centrifuged for 15 min at 4 °C and 17000 g and divided in 110 µL aliquots (chromatin of 2.5 million cells) used freshly.

A volume of 30 µL of magnetic beads were coated with 5 µg of Ab. Two aliquots of sheared chromatin were diluted in 890 µL of ChIP buffer and used for overnight immunoprecipitation. The washes were scaled up to 1 mL of buffer and the two samples were pooled together for elution from the beads. After removal of RNA and proteins, samples were transferred to a Phase Lock Gel Heavy 2 mL tube (5 PRIME) and 400 µL of Phenol/Chloroform/Isoamyl alcohol (Carl Roth) was added for DNA purification. Samples were centrifuged for 10 min at 17000 g and aqueous phase transferred to a new Eppendorf tube containing 16 µL of 5 M NaCl and 1.5 µL of glycogen 20 mg/mL. DNA was precipitated with 1 mL of ultrapure EtOH for 1 h at -80 °C. After centrifugation for 30 min at 17000 g, DNA pellets were washed with 70 % v/v EtOH. DNA was dissolved in 12 µL of 10 mM Tris/HCl pH 8.5.

6.4. ChIP-Sequencing

ChIP-Seq libraries were prepared with MicroPlex Library Preparation kit v2 (Diagenode) according to manufacturer's instructions, using at least 1 ng of purified DNA. The purification of the libraries was performed with Agencourt AMPure XP beads (Beckman Coulter). Sample concentration was measured using the Qubit dsDNA HS Assay kits in a Qubit 3.0 Fluorometer (Invitrogen). Library quality was assessed with the Agilent DNA 1000 Kit on an Agilent 2100 Bioanalyzer (Agilent Technologies). Samples were sequenced with a read length of 50 bp in paired-end mode using Illumina's HiSeq 1500 sequencer at the Laboratory for Functional Genome Analysis (LAFUGA) within the Gene Center (LMU Munich). Data demultiplexing was done in the Galaxy platform (Afgan et al., 2016). Reads were mapped against the mm10 genome using bowtie (Langmead et al., 2009) and bowtie2 (Langmead and Salzberg, 2012), excluding or including multimapped reads, respectively. Identification of ChIP-Seq peaks, tag counting, peak merging, genomic annotation at peaks and analysis of repetitive regions were performed with Homer v4.9 (Heinz et al., 2010). Further downstream analysis was carried out in RStudio (RStudio-Team, 2016).

6.5. Co-immunoprecipitation

After harvesting 50 million cells by trypsinization, they were resuspended in 2 mL of 1x PBS and slowly transferred onto 4 mL of freshly prepared ice cold Ficoll gradient 80 mM Tris/HCl pH 7.4, 8 mM MgCl$_2$, 8 mM CaCl$_2$, 1.6 % w/v NP40, 1.28 % w/v Triton X-100, 20 % v/v Ficoll (LSM 1077 Lymphocyte separation medium, PAA Laboratories), 0.1 % v/v DMSO. Nuclei were isolated by centrifuging with steps of 30 s at 400rpm, 500 rpm, 600 rpm and 700 rpm and a final step of 6 min at 800 rpm. This centrifugation was performed in a Heraeus TX 1000 rotor (Thermo Fischer Scientific) at 4 °C.

Nuclei were resuspended in 500 μL of IP buffer 50 mM Hepes-KOH ph 7.5, 150 mM NaCl, 0.05 % v/v NP40, 20 % v/v glycerol, 10 mM N-maleimide, 1x phosSTOP (Roche Diagnostics), 1x cOmplete, EDTA-free Protease Inhibitor Cocktail (Roche Diagnostics). Nucleic acids were digested with 250 U of benzonase (Merck Millipore) for 15 min at 37 C. Salt concentration was adjusted to 300 mM NaCl and samples were incubated for 30 min on ice. Nuclei were disrupted by sonication twice under 10 % amplitude for 10 s using Branson Digital Sonifier S-250D with a 3 mm tapered microtip (Branson Ultrasonics Corporation). Lysates were cleared by centrifugation for 30 min at 4 °C and 17000 g, transferred to a new Eppendorf tube and 25 μL were saved as 5 % input. To the 5 % input and the pellet containing insoluble proteins, 25 μL of 2x Roti-Load 1 sample buffer (Carl Roth) were added and samples were boiled for 7 min at 95 °C.

After washing 50 μL of Anti-Flag M2 affinity gel (Sigma Aldrich) twice with 1 mL of IP buffer, nuclear protein extract was transferred to the agarose beads. Immunoprecipitation was carried out for for 6 h at 4 °C and 6 rpm vertically. Beads and co-immunoprecipitated complexes were washed 4 times with 1 mL of IP buffer with 300 mM NaCl for 5 min at RT and 6 rpm vertically. Proteins were eluted in 25 μL of 2x Roti-Load 1 sample buffer (Carl Roth) for 7 min at 95 °C. Samples containing the 5 % input, insoluble proteins and immunoprecipitated proteins were analyzed by western blot.

6.6. Embryoid body formation

After harvest, 0.5 million mESC were resuspended in 10 mL of EB medium (mESC medium without LIF), to obtain 500 drops of 20 μL each containing 1000 cells. They were dropped onto the inner part of the lid of a 15 cm dish, while the bottom was filled with 15 mL of 1x PBS. The dishes were incubated at 37 °C and 5 % CO_2. Three days later, the drops were flushed with 20 mL of EB medium and split into 2 bacteriological grade dishes. At the fifth and tenth day of differentiation, EB of one of the dishes was harvested. After washing once with 1x PBS, EB were trypsinized for 5 min and triturated by pipetting 60 times. Trypsin was diluted 10 times in 1x PBS and cells were washed once with 1x PBS. Cell pellets were used for RNA extraction.

6.7. Gene targeting for knockout

The sgRNA were designed using an online tool (crispr.mit.edu). They were cloned in the pX330-U6-Chimeric_BB-CBh-hSpCas9 Addgene plasmid # 42230 (Cong et al., 2013). To facilitate the screening process, a pair of sgRNA were used, that were complementary to two sites in the second exon of the gene around 100 bp away from each other. This way, deletion of the region between the targeted sites would be induced and it was used to easily distinguish mutants during genotyping by PCR. One day before transfection, 0.3 million mESC were seeded on a 6-well dish. The sgRNA-containing plasmids also expressing Cas9 were co-transfected with a plasmid expressing EF1α-puromycin (p963) using jetPRIME Transfection Reagent (Polyplus-transfection) for 16 h. Cells transfected with a high load of plasmids were selected with 1.5 μg/mL

of puromycin for 24 h. The described steps were also performed without transfection of Cas9-containing plasmids to generate control cell lines. For acute knockout experiments, the knockout cell pool was analyzed around 6 days after transfection. For generation of stable isogenic cell lines, single colonies were picked around 7 days after transfection and expanded to isolate cells containing the same type of mutation. The target loci of each clonal cell line were analyzed in agarose gel after PCR amplification and by Sanger sequencing (Eurofins Genomics). Knockout at protein level was confirmed by western blot.

6.8. Gene targeting for knockin

For Flag-*Atf7ip* KI mESC lines generation, 0.3 million cells were seeded on a 6-well dish one day before transfection. The sgRNA were designed with the help of an online tool (crispr.mit.edu). The plasmid expressing Cas9 and the sgRNA to target the stop codon region of *Atf7ip* (p1516) was co-transfected with a plasmid containing the Flag sequence surrounded by 800 bp homology arms corresponding to the endogenous sequence surrounding the desired insertion site (p1467) and another expressing EF1α-puromycin (p963). The PAM sequence of the HDR template was mutated to avoid recognition and cleavage by the Cas9. Transfection was performed for 16 h using jetPRIME Transfection Reagent (Polyplus-transfection). Cells transfected with a high load of plasmids were selected with 1.5 µg/mL of puromycin for 24 h. For generation of stable isogenic cell lines, single colonies were picked around 7 days after transfection. The target loci of each clonal cell line were analyzed in agarose gel after PCR amplification and by Sanger sequencing (Eurofins Genomics). Knockin at protein level was confirmed by western blot. To generate Min-*Atf7ip* KI mESC lines, a similar procedure was carried out, with the difference that the plasmid expressing Cas9 and the sgRNA targeting the start codon region of *Atf7ip* also expressed puromycin (p1597). In this case the donor plasmid contained the Min-tag sequence surrounded by homology arms (p1646) of 800 bp. The target loci of each clonal cell line were analyzed in agarose gel after PCR amplification and by Sanger sequencing (Eurofins Genomics). For expressing the different mutants, a heterozygous *Atf7ip* knockout cell line was obtained by co-transfecting the Min-*Atf7ip* KI mESC line with an empty *pattB* vector coding for hygromicin B resistence (p1615) and the plasmid coding for the Bxb1 integrase (p1584). After antibiotics selection for 6 days to make sure all surviving cells suffered the targeted recombination, single clones were isolated and screened for the heterozygous genotype by PCR. For the rescue experiments, a heterozygous clone was again co-transfected to integrate the different mutants of interest separately, which also expressed zeocin resistance.

6.9. Immunofluorescence microscopy

Coverslips were placed each in one well of a 24-well tissue culture grade dish. They were washed once with 100 % ethanol and once with sterile ddH$_2$O. After drying at RT for 5 min 10 % Poly-L-Lysine (Sigma Aldrich) in water was added to each well and incubated for 15 min at RT.

The coverslips were washed three times with sterile ddH$_2$O and air-dried for 5 min. After that, they were incubated 30 min at 37 °C. Around 200 thousand cells in suspension were seeded on each well and incubated 60 min, to let them adhere. The medium was aspirated and cells were carefully washed once in 1 mL of 1x PBS. Fixation was carried out with 500 µL of 3.7 % formaldehyde (Carl Roth) in 1x PBS for 10 min at RT: Fixed cells were washed twice in 1 mL of 1x PBS for 5 min and permeabilized with 500 µL of 3 mM sodium citrate tribasic dehydrate (Merck), 0.1 % v/v Triton X-100. Permeabilized cells were washed twice in 1 mL of washing solution 1x PBS and twice in 1x PBS, 0.1 % v/v Tween 20, 0.2 % w/v BSA for 5 min. Cells were blocked with 300 µL of blocking solution 1x PBS, 0.1 % v/v Tween 20, 2.5 % w/v BSA for 30 min and incubated overnight at 4 °C with 200 µL of blocking solution containing the desired primary Ab. The dish was warmed to RT for 30 min and stained cells were washed three times in 1 mL of washing solution for 10 min. They were incubated with proper secondary Ab diluted 1:1000 in blocking solution, 10 % normal goat serum (Dianova-Jackson Immuno Research) at RT for 1 h. After washing three times in 1x PBS, 0.1 % Tween 20 for 10 min, cells were imbedded in Vectashield Antifade Mounting Medium with DAPI (Vector Laboratories) on standard microscope slides (Carl Roth) and sealed with nail polish. The immunofluorescent staining was examined with Axiovert 200 M inverted microscope for transmitted light and epifluorescence (Carl Zeiss Microscopy) with the help of the AxioVision Special Edition Software (Carl Zeiss Microscopy).

6.10. Mass spectrometry

In the last step of co-IP, instead of eluting the proteins from agarose beads, the excess of NP-40 detergent and protease inhibitors was removed by washing beads and co-immunoprecipitated complexes 3 times with 300 µL of 100 mM Tris/HCl pH 7.5. To avoid contamination with Ab peptides, proteins were eluted from the agarose beads by digesting with 1 µg of endoproteinase LysC in 50 µl of 100 mM Tris/HCl buffer, 3.7 M Urea. Samples were incubated for 50 min at 28 °C and 1200 rpm horizontally. After spinning down the samples, supernatants were transferred to fresh low binding eppendorf tubes. Beads were washed 3 times with 75 µL of 100 mM ammonium bicarbonate and the supernatants were combined with the first one of that sample. DDT was added to a final concentration of 10 mM and proteins were digested overnight with 1.5 µg of trypsin at 28 °C and 1200 rpm horizontally. Digestion was blocked with 30 mM of iodoacetamide for 40 min at RT and 550 rpm in the dark.

Samples were acidified with 1 µl of 100 % v/v trifluoroacetic acid and 2 µl of 100 % v/v formic acid. Formic acid was added until pH of samples reached between 2 and 3. Purification of peptides was carried out using C18 stage tips. Tips were first activated by washing twice with 50 µl of methanol, once with 50 µl of elution solution and three times with 65 µl of 0.1 % v/v trifluoroacetic acid in. Samples were loaded into the activated tips by centrifugation for 20 min at 55 g. The stage tips were washed twice with 70 µL of 0.2 % v/v formic acid in ultra-pure H$_2$O and dried completely by centrifuging for 2 min at 170 g. Peptides were eluted twice by adding 65 µl of

elution buffer and centrifuging for 4 minutes at 70 g. The flow through was vacuum-dried and re-suspended in 10 µl 2% v/v formic acid in ultra-pure H_2O. High-performance liquid chromatography-tandem mass spectrometry (HPLC-MS/MS) was performed by injecting purified peptides in an Ultimate 3000 nanoLC system coupled to a QExactive HF hybrid quadrupole-Orbitrap mass spectrometer (Thermo Fischer Scientific). Measurements were done with a 90 min program of data-dependent acquisition.

The raw data files were analyzed by MaxQuant Software Suit v.1.5.5.1 (Cox and Mann, 2008). Peptides were searched against the Swiss-Prot mouse database (October 2016) using the Andromeda search algorithm allowing for a maximum of three missed cleavage sites. Carbamidomethylation (C) was set as fixed modification and oxidation (M) and acetylation at the protein N-terminus as variable modifications for peptide identification. Mass deviation for precursor ions used was 20 ppm for the first search and 5 ppm for the main search. MS/MS mass deviation was set to 25 ppm. Peptide spectrum matches were filtered for 1% FDR rate and proteins filtered for 5%. Log-fitted iBAQ values were calculated for protein quantitation. For quantitative analysis of IP experiments, reversed protein hits were removed prior to further analysis. Protein hits with log 2 transformed iBAQ values lower than 3 quantitation values were removed. All samples were normalized by the median from the sample with the highest median. Subsequently, normalized IP-MS data were analyzed in Perseus (Tyanova et al., 2016) to determine proteins specifically enriched in the FLAG-ATF7IP pulldown experiment. A modified t-test was performed employing a permutation-based FDR rate. For analysis of PTM of ATF7IP, lysine acetylation, lysine ubiquitination (GG), lysine and arginine monomehtylation, as well as phosphorylation of serine, threonine, and tyrosine were enabled as variable modifications.

6.11. Molecular cloning

Bacteria were transformed using heat shock method. Plasmid preparation was performed using NucleoBond Xtra Midi kit and NucleSpin Plasmid (noLid) kit (Macherey-Nagel), according to manufacturer's instructions. For restriction digestion-based cloning, restriction endonuclease products, T4 DNA ligase and Antarctic Phosphatase from New England BioLabs were used. For PCR-based cloning, Gibson Assembly Master Mix (New England BioLabs) was used. DNA fragments of interest were amplified with Q5 High Fidelity DNA Polymerase (New England BioLabs). Sequence of the constructs was verified by Sanger sequencing (Eurofins Genomics).

6.12. Protein extracts

Nuclear protein extracts were obtained using the Nuclear Extract Kit (Active Motif). Whole cell proteins extracts were prepared by resuspending cell pellets in freshly prepared lysis buffer 50 mM Tris/HCl pH 7.5, 2 % w/v SDS, 1 % v/v Triton X-100, 1 mM PMSF, 0.5x cOmplete, EDTA-free Protease Inhibitor Cocktail (Roche Diagnostics). Samples were vortexed for 10 s at max speed and boiled for 10 min at 95 °C. They were sonicated twice under 10 % amplitude for 10 s

using Branson Digital Sonifier S-250D with a 3 mm tapered microtip (Branson Ultrasonics Corporation). After a 15 min centrifugation at RT and 17000 g, supernatant containing soluble protein extracts were mixed with 1x Roti-Load 1 sample buffer (Carl Roth) and stored at -20 °C. Protein concentration was quantified with the Pierce BCA Protein Assay kit (Thermo Fischer Scientific). For acid extraction of histones, nuclei of 20 million cells were isolated as described for the co-immunoprecipitation protocol. After that, they were resuspended in 100 µL of ice cold 0.4 N HCl and incubated overnight at 4 °C and 20 rpm vertically. Samples were centrifuged for 20 min at 17000 g and transferred to a fresh eppendorf tube on ice. They were dialyzed 3 times in 1 L of 100 mM acetic acid for 1 h each. After adding 1x Roti-Load 1 sample buffer (Carl Roth) samples were stored at -20 °C.

6.13. Quantification of RNA levels

The RNA levels were quantified by two-step RT-qPCR. For that, harvested cells were washed once with 1x PBS and cell pellet was snap-frozen in RLT buffer (Qiagen), 1% v/v β-mercaptoethanol (Sigma Aldrich). Total RNA was extracted using RNeasy Mini Kit (Qiagen) coupled to DNA on-column digestion using RNase-free DNase I (Qiagen). Synthesis of cDNA was carried out using SuperScript III Reverse Transcriptase (Invitrogen), Random Primer 6 (New England BioLabs) and RNasin Ribonuclease Inhibitor (Promega). Quantification of cDNA was performed with Fast SYBR Green Master Mix (Applied Biosystems) in a LightCycler 480 (Roche Diagnostics), according to the MIQE Guidelines (Bustin et al., 2009). Reactions were performed in a volume of 10 µL in a 384-well LightCycler plate PP (Sarstedt). C_q values were calculated by the LightCycler 480 Software (Roche Diagnostics) using the Fit Points method of absolute quantification. *Gapdh* and *Hprt* were used as reference genes and data normalization was performed as described elsewhere (Vandesompele et al., 2002).

6.14. RNA-Sequencing

The integrity of the purified total RNA was verified using the Agilent RNA 6000 Pico Kit on an Agilent 2100 Bioanalyzer (Agilent Technologies). Ribosomal RNA was removed from 4 µg of purified total RNA with Ribo-Zero Magnetic Kit (Human/Mouse/Rat, Epicentre) using RiboGuard RNase Inhibitor. rRNA-depleted samples were purified using the RNA Clean & Concentrator-5 columns (Zymo Research). Depletion was checked on the Bioanalyzer. RNA-Seq libraries were prepared following the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (New England BioLabs). Library quality was assessed with the Agilent DNA 1000 Kit (Agilent Technologies) on the Bioanalyzer. Before every Bioanalyzer run, sample concentration was measured using either the Qubit RNA HS Assay or the Qubit dsDNA HS Assay kits in a Qubit 3.0 Fluorometer (Invitrogen). Libraries were sequenced with a read length of 50 bp in paired-end mode using Illumina's HiSeq 1500 sequencer at the Laboratory for Functional Genome Analysis (LAFUGA) within the Gene Center (LMU Munich). Data demultiplexing was done in the Galaxy platform

(Afgan et al., 2016). Reads were mapped to mm10 mouse genome using Tophat (Trapnell et al., 2009). Tags were counted with Homer v4.9 (Heinz et al., 2010). Differential expression was calculated using Cufflinks v2.2.1 (Trapnell et al., 2012). Further downstream analysis was carried out in RStudio (RStudio-Team, 2016).

6.15. RNA interference

One day before transfection for virus packaging, around 4 million HEK293T cells were seeded in a 10 cm cell culture dish. Transfection was carried out using the calcium phosphate method 1 h after changing the medium to a fresh one without G418. For that, a total of 24 µg of plasmid DNA consisting of equimolar ratios of the lentiviral transfer vector, psPax2 (p183) and pLP-eco-env (p811) were mixed with 120 µL of 2.5 M $CaCl_2$ and diluted to 1.2 mL of sterile ddH_2O. While gently vortexing the solution, 1,2 mL of 50 mM HEPES, 280 mM NaCl, 1.5 mM Na_2HPO_4, adjusted to pH 7.05 with NaOH was added dropwise. The 2.4 mL transfection mix was quickly added to the medium, to ensure no changes in pH. After 5-6 h, cells were washed once in 1x PBS and 10 mL of fresh medium was added. The supernatant containing viruses were harvested 30 h and 60 h after transfection, centrifuged for 5 min at RT and 3000 g to pellet debris, snap frozen in liquid nitrogen and stored at -80 °C. For transduction of mESC, 0.3 million cells were seeded per well on a 6-well dish. The next day, the medium was removed and 1 ml of a fresh one was added to the cells. Viruses were quickly thawed and 2 ml of viral supernatant were added to the cells containing 6 µl of 4 mg/ml of polybrene (final concentration of 8 mg/ml). Plates were spun down in the Thermo Multifuge 4KR, for 1 h, at 1000 g and 34° C. Polybrene was diluted by adding 3 ml of fresh medium. Then, the cells were incubated overnight at 37° C + 5% CO_2. One day after transduction medium was removed and the cells were washed in 1xPBS, trypsinized and transferred to a 10 cm dish for antibiotic selection with 1.5 µg/ml of puromycin.

6.16. Exogenous TE silencing reporter assay

Viral production for control (p940) and test viruses (p1074) was carried out as described in the previous section (Chapter 6.15 RNA interference). Virus titer was calculated by transducing HeLa T37 cells. For that, 30 thousand cells were seeded into wells of a 12-well dish. The day after, a serial dilution of the viruses was added to the cells (containing 1 µl to 64 µl of viruses) together with 8 µg/ml of polybrene. Transduction and data acquiring was performed as described below for mESC. For the reporter assay in mESC, 0.1 million cells were seeded per well in two wells of a 12-well dish for each genotype tested and volume of medium was completed to 700 µl. The cells were incubated for around 3 h to let them attach to the bottom of the well. Control and test viruses were quickly thawed and added to the cells containing enough polybrene 4 mg/ml to a final concentration of 8 mg/ml in the well. Usually between 300 and 500 µl of viral supernatant were used, depending on the titer, to obtain a multiplicity of infection of only 15 %. Of note, mESC are normally 60-fold harder to transduce than HeLa T37 cells. One well should contain the control

virus used for normalization and the second well should contain the test virus for evaluating the silencing efficiency. Plates were spun down in the Thermo Multifuge 4KR, for 1 h, at 1000 g and 34° C. Polybrene was diluted by adding 1 volume of fresh medium and the cells were incubated at 37° C + 5% CO_2. Two days after transduction, cells were washed in 1x PBS and 100 µl of trypsin was added. After 3 min of incubation at 37° C, 100 µl of 1x PBS was added to each well and the plates were placed on ice. After pipetting up and down several times, the cells were filtered and EGFP-positive cells were counted by flow cytometry using BD FACSCanto flow cytometer (BD Biosciences) with the help of the BD FACSDiva Software. Data were further processed using the FlowJo Software (TreeStar). Percentage of remaining EGFP-positive cells were calculated by dividing the values obtained for the test virus by the values from the control after correcting by the virus titer. This protocol was based on a previously published method (Sadic et al., 2015).

6.17. Silver staining

Polyacrylamide gels were fixed at RT in 10 % v/v acetic acid, 30 % v/v isopropanol for at least 2 h. Gels were washed 3 times for 20 min in 30 % v/v ethanol and incubated in 0.02% sodium thiosulfate at RT for 1 min. They were washed with ddH_2O 3 times for 20 s. Staining was performed at RT in 0.1 % w/v silver nitrate for 1 h with mild agitation in the dark. Stained gels were washed 3 times for 30 s with ddH_2O and developed in 3 % w/v Na_2CO_3, 0.05 % v/v formaldehyde, 0.0004 % $Na_2S_2O_3$ for 5 to 10 min, until the bands were clear. Reduction was quenched by incubating the gels in 5 % w/v glycine for 10 min. Developed gels were stored in ddH_2O at 4°C until scanning.

6.18. Western blot

Protein extracts were denatured by adding 2x Roti-Load 1 sample buffer (Carl Roth) and boiling at 95 °C for 5 min. Samples were separated through SERVAGel TG PRiME 4-12 % precast SDS Page (SERVA Electrophoresis) in running buffer 25 mM Tris, 200 mM glycine, 1 % (m/v) SDS at RT for 1 h and 25 mA per gel. They were blotted onto methanol activated-PVDF membranes in a wet-blotting chamber (Bio-Rad Laboratories) containing blotting buffer 50 mM Tris, 40 mM glycine, 10 % v/v methanol, 5 µM SDS for 1.5 h at 4 °C and 400 mA. The blotting was performed at 400 mA under 4°C for 1.5 hours. Membranes were incubated in blocking buffer 1x PBS, 2.5 % w/v BSA and 2.5 % w/v skim milk at RT for 1 h under mild agitation. Blocked membranes were incubated with primary Ab in blocking buffer at 4°C for 16 hours. After that, they were washed 3 times with PBST buffer 1x PBS, 0.1 % v/v Tween 20 for 20 min. Incubation with secondary Ab was done in a dilution of 1:3000 in blocking buffer at RT for 1.5 h. The probed membranes were washed 3 times in PBST for 20 min. Chemioluminescence was detected in X-ray films (Fujifilm) using ECL Western blot detection reagent (Amersham Biosciences) or Immobilon Western Chemiluminescent HRP Substrate (Merck Millipore).

7. ABBREVIATIONS

2i+vitC	two inhibitors medium complemented with vitamin C
aa	amino acid(s)
Ab	antibody(ies)
Atf7ip	Activating transcription factor 7-interacting protein 1 (also mAM, Mcaf1)
bp	base pair(s)
Cas9	CRISPR-associated endonuclease 9
cDNA	complementary DNA
ChIP	chromatin immunoprecipitation
ChIP-qPCR	chromatin immunoprecipitation followed by quantitative PCR
ChIP-Seq	chromatin immunoprecipitation followed by high throughput sequencing
cKO	conditional knockout
co-IP	co-immunoprecipitation
CpG	cytosine nucleotide followed by a guanine in the 5' to 3' direction
CRISPR	clustered regularly interspaced short palindromic repeats
ddH$_2$O	Double-distilled water
dGene-ID	*Drosphila* gene
dKO	double knockout
DNAme	deoxyribonucleic acid methylation
dPROTEIN-ID	*Drosphila* protein
EB	embryoid body(ies)
EGFP	enhanced green fluorescent protein
EGFP+	EGFP-positive
GO	gene ontology
H3K9ac	Histone 3 Lysine 9 acetylation
H3K9me/1/2/3	Histone 3 Lysine 9 -, mono-, di- or tri- methylation
H3K9MT	Histone 3 Lysine 9 methyltransferase(s)
HDR	homology-directed repair
HKMT	Histone lysine methyltransferase(s)
HEK293T	human embryonic kidney cell line 293 transformed with the large-T-antigen of SV40 virus
HeLa	human cervix carcinoma cell line
HP1	Heterochromatic protein 1
hPROTEIN-ID	human protein
hygroR	hygromycin B resistance
IAP	Internal A-type Particle
ICM	inner cell mass
IRES	internal ribosome entry site
kb	kilobases
KO	knockout
KI	knockin
LIF	leukemia inhibitor factor
MBD	methyl-CpG binding domain
MEF	mouse embryonic fibroblast(s)
mESC	mouse embryonic stem cell(s)
MIN	multifunctional integrase
mRNA	messenger ribonucleic acid

NHEJ	non-homologous end-joining
p	flox allele
PCA	principal component analysis
attB	bacterial attachment site
attP	phage attachment site
Pol-II	RNA polymerase II
PcG	Polycomb group
PML-NB	promyelocytic leukemia nuclear bodies
PTM	post-translational modification/s
puroR	puromycin resistance
qPCR	quantitative polymerase chain reaction
rpm	rotations per minute
RNA-Seq	RNA high throughput sequencing
rRNA	ribosomal RNA
RNP	Ribonucleoprotein(s)
RT	room temperature
RT-qPCR	reverse transcription quantitative polymerase chain reaction
sgRNA	single guide ribonucleic acid(s)
shRNA	short hairpin ribonucleic acid(s)
RNA-Seq	ribonucleic acid-sequencing
SAM	S-adenosyl-L-methyonine
Setdb1	SET domain bifurcated 1 (also ESET)
SIM	Sumo-interaction motif
TE	transposable element(s)
TKO	triple knockout
TRD	transcriptional repression domain
TSS	transcription start site(s)
WB	Western blot
XCI	X chromosome inactivation

8. CURRICULUM VITAE

The CV is not accessible in the public version.

9. APPENDIX

Table 9.1 | Mass spectrometric analysis of anti-FLAG co-IP of Flag-*Atf7ip* mESC versus wild type cells.
The top 80 hits are displayed, together with their peptide counts and the protein coverage obtained in all replicates.

Gene	Description	ct1	ct2	ct3	ip1	ip2	ip3	ip4	% Coverage	p
Atf7ip	Activating transcription factor 7-interacting protein 1	4	0	4	28	53	38	30	50	0.033
Setdb1	Histone-lysine N-methyltransferase SETDB1	1	2	2	26	38	27	25	31	0.001
Flnb	Filamin-B	1	3	4	7	37	17	28	21	0.018
Nup214	Nuclear pore complex protein Nup214	2	7	9	12	35	18	16	24	0.028
Ddx39a	ATP-dependent RNA helicase DDX39A	2	6	8	14	19	14	15	54	0.017
Nono	Non-POU domain-containing octamer-binding protein	2	6	7	14	22	9	11	57	0.021
Uhrf1	E3 ubiquitin-protein ligase UHRF1	2	4	3	8	24	14	9	43	0.009
Nup93	Nuclear pore complex protein Nup93	0	1	3	4	17	13	12	29	0.020
Hcfc1	Host cell factor 1	0	1	1	4	21	9	11	15	0.002
Ruvbl1	RuvB-like 1	2	4	4	7	15	11	10	48	0.002
Smc3	Structural maintenance of chromosomes protein 3	0	0	3	2	23	5	5	21	0.026
Nop56	Nucleolar protein 56	1	4	6	10	19	15	15	38	0.009
Kpnb1	Importin subunit beta-1	2	5	8	12	18	14	14	28	0.028
Nop2	Probable 28S rRNA (cytosine-C(5))-methyltransferase	0	1	2	5	12	10	9	24	0.007
Eif4a3	Eukaryotic initiation factor 4A-III	1	3	10	7	13	14	15	42	0.024
Ccar2	Cell cycle and apoptosis regulator protein 2	0	1	1	3	10	6	8	13	0.009
Skiv2l2	Superkiller viralicidic activity 2-like 2	0	2	5	6	17	14	15	25	0.012
Mcm5	DNA replication licensing factor MCM5	0	0	4	4	7	8	8	19	0.007
Polr1c	DNA-directed RNA polymerases I and III subunit RPAC1	0	0	1	1	3	1	1	10	0.028
Ppp1ca	Serine/threonine-protein phosphatase PP1-alpha catalytic subunit	1	2	4	3	8	4	5	28	0.007
Wdr18	WD repeat-containing protein 18	0	0	1	4	8	5	5	31	0.007
Sap18	Histone deacetylase complex subunit SAP18	0	0	1	0	3	1	2	23	0.028
Pcna	Proliferating cell nuclear antigen	0	3	3	4	12	6	6	63	0.010
Rbbp4	Histone-binding protein RBBP4	2	2	6	1	9	7	8	39	0.024
Dkc1	H/ACA ribonucleoprotein complex subunit 4	0	1	0	1	7	3	5	27	0.029
Rae1	mRNA export factor	0	1	0	1	5	2	3	22	0.011
Cpsf1	Cleavage and polyadenylation specificity factor subunit 1	0	0	1	1	5	5	5	9	0.019
Eed	Polycomb protein EED	0	0	1	1	5	3	4	13	0.028
Gatad2a	Transcriptional repressor p66 alpha	0	0	2	3	12	8	9	26	0.011
Pno1	RNA-binding protein PNO1	0	0	0	2	4	2	2	24	0.000
Eif2s3x/y	Eukaryotic translation initiation factor 2 subunit 3	0	1	1	1	7	6	6	24	0.014
Usp39	U4/U6.U5 tri-snRNP-associated protein 2	0	1	0	1	6	2	4	16	0.009
Rpl30	60S ribosomal protein L30	0	1	0	2	4	2	2	47	0.001
Cstf2	Cleavage stimulation factor subunit 2	0	0	1	2	6	3	4	16	0.007
Rpl10a	60S ribosomal protein L10a	1	3	2	7	8	4	4	33	0.001
Ppp1cc	Serine/threonine-protein phosphatase PP1-gamma catalytic subunit	0	1	6	4	6	3	5	21	0.017
Smarcb1	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily B member 1	0	0	1	1	4	3	4	16	0.010

Gene	Description	ct1	ct2	ct3	ip1	ip2	ip3	ip4	% Coverage	p
Zfr	Zinc finger RNA-binding protein	0	0	0	0	4	2	1	4	0.016
Uqcrc1	Cytochrome b-c1 complex subunit 1, mitochondrial	0	0	0	1	2	2	3	10	0.001
Fscn1	Fascin	1	0	2	3	7	12	12	23	0.014
Rps5	40S ribosomal protein S5;40S ribosomal protein S5, N-terminally processed	0	0	0	0	2	2	2	23	0.034
Pml	Protein PML	0	0	1	1	5	5	6	14	0.022
Polr1a	DNA-directed RNA polymerase I subunit RPA1	0	0	1	0	5	3	2	4	0.003
Pabpn1	Polyadenylate-binding protein 2	0	0	1	2	4	3	4	15	0.016
Ppp2r1a	Serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A alpha isoform	0	1	0	1	6	3	1	14	0.015
Sympk	Symplekin	0	0	1	2	8	4	4	10	0.002
Kdm1a	Lysine-specific histone demethylase 1A	0	0	1	1	4	2	3	9	0.001
Wdr43	WD repeat-containing protein 43	0	0	0	2	7	3	6	15	0.003
Lin28a	Protein lin-28 homolog A	0	2	0	1	3	1	1	23	0.017
Ddx39b	Spliceosome RNA helicase Ddx39b	1	5	4	13	15	13	13	44	0.025
Bub3	Mitotic checkpoint protein BUB3	0	0	0	2	4	2	2	17	0.000
Exosc10	Exosome component 10	0	0	0	0	2	2	2	4	0.009
Nup37	Nucleoporin Nup37	0	0	0	0	6	4	3	26	0.018
Rps11	40S ribosomal protein S11	1	1	1	5	5	3	3	33	0.002
Eif3f	Eukaryotic translation initiation factor 3 subunit F	0	0	0	0	5	2	3	13	0.020
Rpl3	60S ribosomal protein L3	0	1	2	3	6	2	4	18	0.031
Polr2e	DNA-directed RNA polymerases I, II, and III subunit RPABC1	0	0	0	0	4	2	3	30	0.025
Emg1	Ribosomal RNA small subunit methyltransferase NEP1	0	0	2	2	5	3	4	23	0.013
Eif3l	Eukaryotic translation initiation factor 3 subunit L	1	0	0	1	4	1	1	8	0.029
Mtco2	Cytochrome c oxidase subunit 2	0	0	0	3	2	2	3	19	0.000
Polr2h	DNA-directed RNA polymerases I, II, and III subunit RPABC3	0	0	0	3	3	2	2	34	0.000
Dnmt3b	DNA (cytosine-5)-methyltransferase 3B	0	0	1	1	5	3	4	8	0.019
Znf518a	Zinc finger protein 518A	0	0	0	0	4	3	3	3	0.002
Cct8	T-complex protein 1 subunit theta	0	0	0	1	3	1	0	9	0.024
Rps14	40S ribosomal protein S14	0	0	3	4	4	3	4	38	0.022
Ddx42	ATP-dependent RNA helicase DDX42	0	0	0	0	4	2	3	6	0.023
Gnb1	Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-1	0	1	2	3	4	3	3	17	0.023
Pes1	Pescadillo homolog	0	0	0	0	3	2	1	10	0.027
Ddx6	Probable ATP-dependent RNA helicase DDX6	0	0	0	1	4	2	3	16	0.007
Rpl8	60S ribosomal protein L8	0	0	2	2	2	2	2	11	0.032
Thoc5	THO complex subunit 5 homolog	0	0	0	0	3	2	3	8	0.016
Ddx1	ATP-dependent RNA helicase DDX1	0	0	1	0	3	1	2	8	0.027
Noc4l	Nucleolar complex protein 4 homolog	0	0	1	1	1	2	3	9	0.030
Exosc9	Exosome complex component RRP45	0	0	2	1	2	2	2	5	0.029
Cstf3	Cleavage stimulation factor subunit 3	0	0	0	1	2	1	1	5	0.002
Exosc5	Exosome complex component RRP46	0	0	0	2	3	2	1	15	0.000
Rpl38	60S ribosomal protein L38	0	1	0	1	3	3	3	46	0.032
Nol6	Nucleolar protein 6	0	0	0	1	3	2	2	4	0.017
Prmt1	Protein arginine N-methyltransferase 1	0	0	0	0	3	1	2	12	0.033
Cdc40	Pre-mRNA-processing factor 17	0	0	0	0	3	1	1	7	0.003

10. REFERENCES

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

**Afgan, E., Baker, D., van den Beek, M., Blankenberg, D., Bouvier, D.*, et al.* (2016). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Res *44*, W3-W10.

**Ait-Si-Ali, S., Guasconi, V., Fritsch, L., Yahi, H., Sekhri, R.*, et al.* (2004). A Suv39h-dependent mechanism for silencing S-phase genes in differentiating but not in cycling cells. EMBO J *23*, 605-615.

**Alagoz, M., Katsuki, Y., Ogiwara, H., Ogi, T., Shibata, A.*, et al.* (2015). SETDB1, HP1 and SUV39 promote repositioning of 53BP1 to extend resection during homologous recombination in G2 cells. Nucleic Acids Res *43*, 7931-7944.

**An, J., Zhang, X., Qin, J., Wan, Y., Hu, Y.*, et al.* (2014). The histone methyltransferase ESET is required for the survival of spermatogonial stem/progenitor cells in mice. Cell Death Dis *5*, e1196.

Andersen, E.C., and Horvitz, H.R. (2007). Two C. elegans histone methyltransferases repress lin-3 EGF transcription to inhibit vulval development. Development *134*, 2991-2999.

Atlasi, Y., and Stunnenberg, H.G. (2017). The interplay of epigenetic marks during stem cell differentiation and development. Nat Rev Genet.

**Ayyanathan, K., Lechner, M.S., Bell, P., Maul, G.G., Schultz, D.C.*, et al.* (2003). Regulated recruitment of HP1 to a euchromatic gene induces mitotically heritable, epigenetic gene silencing: a mammalian cell culture model of gene variegation. Genes Dev *17*, 1855-1869.

Bachman, K.E., Rountree, M.R., and Baylin, S.B. (2001). Dnmt3a and Dnmt3b are transcriptional repressors that exhibit unique localization properties to heterochromatin. J Biol Chem *276*, 32282-32287.

**Basavapathruni, A., Gureasko, J., Porter Scott, M., Hermans, W., Godbole, A.*, et al.* (2016). Characterization of the Enzymatic Activity of SETDB1 and Its 1:1 Complex with ATF7IP. Biochemistry.

Bessler, J.B., Andersen, E.C., and Villeneuve, A.M. (2010). Differential localization and independent acquisition of the H3K9me2 and H3K9me3 chromatin modifications in the Caenorhabditis elegans adult germ line. PLoS Genet *6*, e1000830.

**Beyer, S., Pontis, J., Schirwis, E., Battisti, V., Rudolf, A.*, et al.* (2016). Canonical Wnt signalling regulates nuclear export of Setdb1 during skeletal muscle terminal differentiation. Cell Discov *2*, 16037.

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 Dev *23*, 2484-2489.

**Binda, O., LeRoy, G., Bua, D.J., Garcia, B.A., Gozani, O.*, et al.* (2010). Trimethylation of histone H3 lysine 4 impairs methylation of histone H3 lysine 9: regulation of lysine methyltransferases by physical interaction with their substrates. Epigenetics *5*, 767-775.

Bird, A.P., and Wolffe, A.P. (1999). Methylation-induced repression--belts, braces, and chromatin. Cell *99*, 451-454.

Blackburn, M.L., Chansky, H.A., Zielinska-Kwiatkowska, A., Matsui, Y., and Yang, L. (2003). Genomic structure and expression of the mouse ESET gene encoding an ERG-associated histone methyltransferase with a SET domain. Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression *1629*, 8-14.

**Blaschke, K., Ebata, K.T., Karimi, M.M., Zepeda-Martinez, J.A., Goyal, P.*, et al.* (2013). Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. Nature *500*, 222-226.

Brower-Toland, B., Riddle, N.C., Jiang, H., Huisinga, K.L., and Elgin, S.C. (2009). Multiple SET methyltransferases are required to maintain normal heterochromatin domains in the genome of Drosophila melanogaster. Genetics *181*, 1303-1319.

**Bulut-Karslioglu, A., De La Rosa-Velazquez, I.A., Ramirez, F., Barenboim, M., Onishi-Seebacher, M.*, et al.* (2014). Suv39h-dependent H3K9me3 marks intact retrotransposons and silences LINE elements in mouse embryonic stem cells. Mol Cell *55*, 277-290.

**Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J.*, et al.* (2009). The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin Chem *55*, 611-622.

**Ceol, C.J., Houvras, Y., Jane-Valbuena, J., Bilodeau, S., Orlando, D.A.*, et al.* (2011). The histone methyltransferase SETDB1 is recurrently amplified in melanoma and accelerates its onset. Nature *471*, 513-517.

Chang, L.K., Chuang, J.Y., Nakao, M., and Liu, S.T. (2010). MCAF1 and synergistic activation of the transcription of Epstein-Barr virus lytic genes by Rta and Zta. Nucleic Acids Res *38*, 4687-4700.

**Chang, L.K., Chung, J.Y., Hong, Y.R., Ichimura, T., Nakao, M.*, et al.* (2005). Activation of Sp1-mediated transcription by Rta of Epstein-Barr virus via an interaction with MCAF1. Nucleic Acids Res *33*, 6528-6539.

Chase, K.A., Gavin, D.P., Guidotti, A., and Sharma, R.P. (2013). Histone methylation at H3K9: evidence for a restrictive epigenome in schizophrenia. Schizophr Res *149*, 15-20.

Chase, K.A., and Sharma, R.P. (2013). Nicotine induces chromatin remodelling through decreases in the methyltransferases GLP, G9a, Setdb1 and levels of H3K9me2. Int J Neuropsychopharmacol *16*, 1129-1138.

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. J Biol Chem *286*, 41115-41124.

Cho, S., Park, J.S., and Kang, Y.K. (2013). Regulated nuclear entry of over-expressed Setdb1. Genes Cells *18*, 694-703.

Cho, S., Park, J.S., and Kang, Y.K. (2014). AGO2 and SETDB1 cooperate in promoter-targeted transcriptional silencing of the androgen receptor gene. Nucleic Acids Res *42*, 13545-13556.

Cho, S., Park, J.S., Kwon, S., and Kang, Y.K. (2012). Dynamics of Setdb1 expression in early mouse development. Gene Expr Patterns *12*, 213-218.

Choi, J., Huebner, A.J., Clement, K., Walsh, R.M., Savol, A*., et al.* (2017). Prolonged Mek1/2 suppression impairs the developmental potential of embryonic stem cells. Nature *548*, 219-223.

Chuong, E.B., Elde, N.C., and Feschotte, C. (2017). Regulatory activities of transposable elements: from conflicts to benefits. Nat Rev Genet *18*, 71-86.

Clough, E., Moon, W., Wang, S., Smith, K., and Hazelrigg, T. (2007). Histone methylation is required for oogenesis in Drosophila. Development *134*, 157-165.

Collins, P.L., Kyle, K.E., Egawa, T., Shinkai, Y., and Oltz, E.M. (2015). The histone methyltransferase SETDB1 represses endogenous and exogenous retroviruses in B lymphocytes. Proc Natl Acad Sci U S A *112*, 8367-8372.

Collins, R.E., Northrop, J.P., Horton, J.R., Lee, D.Y., Zhang, X*., et al.* (2008). The ankyrin repeats of G9a and GLP histone methyltransferases are mono- and dimethyllysine binding modules. Nat Struct Mol Biol *15*, 245-250.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R*., et al.* (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819-823.

Corley, S.M., MacKenzie, K.L., Beverdam, A., Roddam, L.F., and Wilkins, M.R. (2017). Differentially expressed genes from RNA-Seq and functional enrichment results are affected by the choice of single-end versus paired-end reads and stranded versus non-stranded protocols. BMC Genomics *18*, 399.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol *26*, 1367-1372.

Cruvinel, E., Budinetz, T., Germain, N., Chamberlain, S., Lalande, M*., et al.* (2014). Reactivation of maternal SNORD116 cluster via SETDB1 knockdown in Prader-Willi syndrome iPSCs. Hum Mol Genet *23*, 4674-4685.

Cukier, H.N., Lee, J.M., Ma, D., Young, J.I., Mayo, V*., et al.* (2012). The expanding role of MBD genes in autism: identification of a MECP2 duplication and novel alterations in MBD5, MBD6, and SETDB1. Autism Res *5*, 385-397.

Dambacher, S., de Almeida, G.P., and Schotta, G. (2013). Dynamic changes of the epigenetic landscape during cellular differentiation. Epigenomics *5*, 701-713.

Daxinger, L., Oey, H., Isbel, L., Whitelaw, N.C., Youngson, N.A*., et al.* (2016). Hypomethylation of ERVs in the sperm of mice haploinsufficient for the histone methyltransferase Setdb1 correlates with a paternal effect on phenotype. Sci Rep *6*, 25004.

De Graeve, F., Bahr, A., Chatton, B., and Kedinger, C. (2000). A murine ATFa-associated factor with transcriptional repressing activity. Oncogene *19*, 1807-1819.

Delacote, F., Perez, C., Guyot, V., Mikonio, C., Potrel, P*., et al.* (2011). Identification of genes regulating gene targeting by a high-throughput screening approach. J Nucleic Acids *2011*, 947212.

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*., et al.* (2008). DNA methylation in ES cells requires the lysine methyltransferase G9a but not its catalytic activity. EMBO J *27*, 2691-2701.

Dou, Y., and Gorovsky, M.A. (2000). Phosphorylation of linker histone H1 regulates gene expression in vivo by creating a charge patch. Mol Cell *6*, 225-231.

Eames, H.L., Saliba, D.G., Krausgruber, T., Lanfrancotti, A., Ryzhakov, G*., et al.* (2012). KAP1/TRIM28: an inhibitor of IRF5 function in inflammatory macrophages. Immunobiology *217*, 1315-1324.

Eckersley-Maslin, M.A., Thybert, D., Bergmann, J.H., Marioni, J.C., Flicek, P*., et al.* (2014). Random monoallelic gene expression increases upon embryonic stem cell differentiation. Dev Cell *28*, 351-365.

El-Brolosy, M.A., and Stainier, D.Y.R. (2017). Genetic compensation: A phenomenon in search of mechanisms. PLoS Genet *13*, e1006780.

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

Escamilla-Del-Arenal, M., da Rocha, S.T., Spruijt, C.G., Masui, O., Renaud, O*., et al.* (2013). Cdyl, a new partner of the inactive X chromosome and potential reader of H3K27me3 and H3K9me2. Mol Cell Biol *33*, 5005-5020.

Eymery, A., Liu, Z., Ozonov, E.A., Stadler, M.B., and Peters, A.H. (2016). The methyltransferase Setdb1 is essential for meiosis and mitosis in mouse oocytes and early embryos. Development *143*, 2767-2779.

Falandry, C., Fourel, G., Galy, V., Ristriani, T., Horard, B*., et al.* (2010). CLLD8/KMT1F is a lysine methyltransferase that is important for chromosome segregation. J Biol Chem *285*, 20234-20241.

Fan, R. (PhD Thesis, 2015). Deletion of Setdb1 in Sox17 lineage cells impairs early embryonic development in the mouse. Ludwig-Maximilians-Universität München.

Fei, Q., Shang, K., Zhang, J., Chuai, S., Kong, D*., et al.* (2015a). Histone methyltransferase SETDB1 regulates liver cancer cell growth through methylation of p53. Nat Commun *6*, 8651.

Fei, Q., Yang, X., Jiang, H., Wang, Q., Yu, Y., *et al.* (2015b). SETDB1 modulates PRC2 activity at developmental genes independently of H3K9 trimethylation in mouse ES cells. Genome Res *25*, 1325-1335.

Figueiredo, M.L., Philip, P., Stenberg, P., and Larsson, J. (2012). HP1a recruitment to promoters is independent of H3K9 methylation in Drosophila melanogaster. PLoS Genet *8*, e1003061.

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

Frietze, S., O'Geen, H., Blahnik, K.R., Jin, V.X., and Farnham, P.J. (2010). ZNF274 recruits the histone methyltransferase SETDB1 to the 3' ends of ZNF genes. PLoS One *5*, e15082.

Fritsch, L., Robin, P., Mathieu, J.R., Souidi, M., Hinaux, H., *et al.* (2010). A subset of the histone H3 lysine 9 methyltransferases Suv39h1, G9a, GLP, and SETDB1 participate in a multimeric complex. Mol Cell *37*, 46-56.

Fujita, N., Watanabe, S., Ichimura, T., Ohkuma, Y., Chiba, T., *et al.* (2003). MCAF Mediates MBD1-Dependent Transcriptional Repression. Molecular and Cellular Biology *23*, 2834-2843.

Gao, H., Yu, Z., Bi, D., Jiang, L., Cui, Y., *et al.* (2007). Akt/PKB interacts with the histone H3 methyltransferase SETDB1 and coordinates to silence gene expression. Mol Cell Biochem *305*, 35-44.

Garcia-Bassets, I., Kwon, Y.S., Telese, F., Prefontaine, G.G., Hutt, K.R., *et al.* (2007). Histone methylation-dependent mechanisms impose ligand dependency for gene activation by nuclear receptors. Cell *128*, 505-518.

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. Nat Genet *36*, 94-99.

Garcia-Perez, J.L., Widmann, T.J., and Adams, I.R. (2016). The impact of transposable elements on mammalian development. Development *143*, 4101-4114.

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

Goodier, J.L. (2016). Restricting retrotransposons: a review. Mob DNA *7*, 16.

Gou, D., Rubalcava, M., Sauer, S., Mora-Bermudez, F., Erdjument-Bromage, H., *et al.* (2010). SETDB1 is involved in postembryonic DNA methylation and gene silencing in Drosophila. PLoS One *5*, e10581.

Groh, S. (Master Thesis, 2015). Mechanisms of retrotransposon silencing in mouse embryonic stem cells. Ludwig-Maximilians-Universität München.

Groh, S., and Schotta, G. (2017). Silencing of endogenous retroviruses by heterochromatin. Cell Mol Life Sci *74*, 2055-2065.

Gunther, M., Laithier, M., and Brison, O. (2000). A set of proteins interacting with transcription factor Sp1 identified in a two-hybrid screening. Mol Cell Biochem *210*, 131-142.

Hachiya, R., Shiihashi, T., Shirakawa, I., Iwasaki, Y., Matsumura, Y., *et al.* (2016). The H3K9 methyltransferase Setdb1 regulates TLR4-mediated inflammatory responses in macrophages. Sci Rep *6*, 28845.

Harte, P.J., Wu, W., Carrasquillo, M.M., and Matera, A.G. (1999). Assignment of a novel bifurcated SET domain gene, SETDB1, to human chromosome band 1q21 by in situ hybridization and radiation hybrids. Cytogenet Cell Genet *84*, 83-86.

Hayashi-Takanaka, Y., Yamagata, K., Wakayama, T., Stasevich, T.J., Kainuma, T., *et al.* (2011). Tracking epigenetic histone modifications in single cells using Fab-based live endogenous modification labeling. Nucleic Acids Res *39*, 6475-6488.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., *et al.* (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell *38*, 576-589.

Heitz, E. (1928). Das heterochromatin der moose. I Jahrb Wiss Botanik *69*, 762–818.

Henikoff, S., and Greally, J.M. (2016). Epigenetics, cellular memory and gene regulation. Curr Biol *26*, R644-648.

Herz, H.M., Garruss, A., and Shilatifard, A. (2013). SET for life: biochemical activities and biological functions of SET domain-containing proteins. Trends Biochem Sci *38*, 621-639.

Hong, W., Li, J., Wang, B., Chen, L., Niu, W., *et al.* (2011). Epigenetic involvement of Alien/ESET complex in thyroid hormone-mediated repression of E2F1 gene expression and cell proliferation. Biochem Biophys Res Commun *415*, 650-655.

Hornbeck, P.V., Zhang, B., Murray, B., Kornhauser, J.M., Latham, V., *et al.* (2015). PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. Nucleic Acids Res *43*, D512-520.

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

Hwang, Y.J., Han, D., Kim, K.Y., Min, S.J., Kowall, N.W., *et al.* (2014). ESET methylates UBF at K232/254 and regulates nucleolar heterochromatin plasticity and rDNA transcription. Nucleic Acids Res *42*, 1628-1643.

Ichimura, T., Watanabe, S., Sakamoto, Y., Aoto, T., Fujita, N., *et al.* (2005). Transcriptional repression and heterochromatin formation by MBD1 and MCAF/AM family proteins. J Biol Chem *280*, 13928-13935.

Ideno, H., Shimada, A., Imaizumi, K., Kimura, H., Abe, M., *et al.* (2013). Predominant expression of H3K9 methyltransferases in prehypertrophic and hypertrophic chondrocytes during mouse growth plate cartilage development. Gene Expr Patterns *13*, 84-90.

Inoue, Y., Matsuura, S., Kurabe, N., Kahyo, T., Mori, H., *et al.* (2015). Clinicopathological and Survival Analysis of Japanese Patients with Resected Non-Small-Cell Lung Cancer Harboring NKX2-1, SETDB1, MET, HER2, SOX2, FGFR1, or PIK3CA Gene Amplification. J Thorac Oncol *10*, 1590-1600.

Ishibashi, T., Yaguchi, A., Terada, K., Ueno-Yokohata, H., Tomita, O., *et al.* (2016). Ph-like ALL-related novel fusion kinase ATF7IP-PDGFRB exhibits high sensitivity to tyrosine kinase inhibitors in murine cells. Exp Hematol *44*, 177-188 e175.

Ishimoto, K., Kawamata, N., Uchihara, Y., Okubo, M., Fujimoto, R., *et al.* (2016). Ubiquitination of Lysine 867 of the Human SETDB1 Protein Upregulates Its Histone H3 Lysine 9 (H3K9) Methyltransferase Activity. PLoS One *11*, e0165766.

Ivanov, A.V., Peng, H., Yurchenko, V., Yap, K.L., Negorev, D.G., *et al.* (2007). PHD domain-mediated E3 ligase activity directs intramolecular sumoylation of an adjacent bromodomain required for gene silencing. Mol Cell *28*, 823-837.

Jaenisch, R., and Bird, A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat Genet *33 Suppl*, 245-254.

James, T.C., Eissenberg, J.C., Craig, C., Dietrich, V., Hobson, A., *et al.* (1989). Distribution patterns of HP1, a heterochromatin-associated nonhistone chromosomal protein of Drosophila. Eur J Cell Biol *50*, 170-180.

Jenuwein, T., Laible, G., Dorn, R., and Reuter, G. (1998). SET domain proteins modulate chromatin domains in eu- and heterochromatin. Cell Mol Life Sci *54*, 80-93.

Jiang, Y., Jakovcevski, M., Bharadwaj, R., Connor, C., Schroeder, F.A., *et al.* (2010). Setdb1 histone methyltransferase regulates mood-related behaviors and expression of the NMDA receptor subunit NR2B. J Neurosci *30*, 7152-7167.

Jiang, Y., Loh, Y.E., Rajarajan, P., Hirayama, T., Liao, W., *et al.* (2017). The methyltransferase SETDB1 regulates a large neuron-specific topological chromatin domain. Nat Genet.

Jiang, Y., Matevossian, A., Guo, Y., and Akbarian, S. (2011). Setdb1-mediated histone H3K9 hypermethylation in neurons worsens the neurological phenotype of Mecp2-deficient mice. Neuropharmacology *60*, 1088-1097.

Jin, G., Zheng, S.L., Lilja, H., Kim, S.-T., Tao, S., *et al.* (2013). Genome-wide Association Study Identifies Loci at ATF7IP and KLK2 Associated with Percentage of Circulating Free PSA. Neoplasia *15*, 95-IN30.

Kalinava, N., Ni, J.Z., Peterman, K., Chen, E., and Gu, S.G. (2017). Decoupling the downstream effects of germline nuclear RNAi reveals that H3K9me3 is dispensable for heritable RNAi and the maintenance of endogenous siRNA-mediated transcriptional silencing in Caenorhabditis elegans. Epigenetics Chromatin *10*, 6.

Kanetsky, P.A., Mitra, N., Vardhanabhuti, S., Vaughn, D.J., Li, M., *et al.* (2011). A second independent locus within DMRT1 is associated with testicular germ cell tumor susceptibility. Hum Mol Genet *20*, 3109-3117.

Kang, H.C., Kim, H.K., Lee, S., Mendez, P., Kim, J.W., *et al.* (2016). Whole exome and targeted deep sequencing identify genome-wide allelic loss and frequent SETDB1 mutations in malignant pleural mesotheliomas. Oncotarget.

Kang, Y.K. (2015). SETDB1 in Early Embryos and Embryonic Stem Cells. Curr Issues Mol Biol *17*, 1-10.

Karanth, A.V., Maniswami, R.R., Prashanth, S., Govindaraj, H., Padmavathy, R., *et al.* (2017). Emerging role of SETDB1 as a therapeutic target. Expert Opin Ther Targets *21*, 319-331.

Karimi, M.M., Goyal, P., Maksakova, I.A., Bilenky, M., Leung, D., *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.

Karlsson, R., Andreassen, K.E., Kristiansen, W., Aschim, E.L., Bremnes, R.M., *et al.* (2013). Investigation of six testicular germ cell tumor susceptibility genes suggests a parent-of-origin effect in SPRY4. Hum Mol Genet *22*, 3373-3380.

Kauzlaric, A., Ecco, G., Cassano, M., Duc, J., Imbeault, M., *et al.* (2017). The mouse genome displays highly dynamic populations of KRAB-zinc finger protein genes and related genetic units. PLoS One *12*, e0173746.

Keniry, A., Gearing, L.J., Jansz, N., Liu, J., Holik, A.Z., *et al.* (2016). Setdb1-mediated H3K9 methylation is enriched on the inactive X and plays a role in its epigenetic silencing. Epigenetics Chromatin *9*, 16.

Kera, Y., Katoh, Y., Ohta, M., Matsumoto, M., Takano-Yamamoto, T., *et al.* (2013). Methionine adenosyltransferase II-dependent histone H3K9 methylation at the COX-2 gene locus. J Biol Chem *288*, 13592-13601.

Kerr, S.C., Ruppersburg, C.C., Francis, J.W., and Katz, D.J. (2014). SPR-5 and MET-2 function cooperatively to reestablish an epigenetic ground state during passage through the germ line. Proc Natl Acad Sci U S A *111*, 9509-9514.

Kim, J., Zhao, H., Dan, J., Kim, S., Hardikar, S., *et al.* (2016). Maternal Setdb1 Is Required for Meiotic Progression and Preimplantation Development in Mouse. PLoS Genet *12*, e1005970.

Kobayashi, K., Mitsui, K., Ichikawa, H., Nakabayashi, K., Matsuoka, M., *et al.* (2014). ATF7IP as a novel PDGFRB fusion partner in acute lymphoblastic leukaemia in children. Br J Haematol *165*, 836-841.

Koch, C.M., Honemann-Capito, M., Egger-Adam, D., and Wodarz, A. (2009). Windei, the Drosophila homolog of mAM/MCAF1, is an essential cofactor of the H3K9 methyl transferase dSETDB1/Eggless in germ line development. PLoS Genet *5*, e1000644.

Koester-Eiserfunke, N., and Fischle, W. (2011). H3K9me2/3 binding of the MBT domain protein LIN-61 is essential for Caenorhabditis elegans vulva development. PLoS Genet *7*, e1002017.

**Koide, S., Oshima, M., Takubo, K., Yamazaki, S., Nitta, E.*, et al.* (2016). Setdb1 maintains hematopoietic stem and progenitor cells by restricting the ectopic activation of nonhematopoietic genes. Blood *128*, 638-649.

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

**Kostaki, M., Manona, A.D., Stavraka, I., Korkolopoulou, P., Levidou, G.*, et al.* (2014). High-frequency p16(INK) (4A) promoter methylation is associated with histone methyltransferase SETDB1 expression in sporadic cutaneous melanoma. Exp Dermatol *23*, 332-338.

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. Biochem Biophys Res Commun *337*, 901-907.

**Lafuente-Sanchis, A., Zuniga, A., Galbis, J.M., Cremades, A., Estors, M.*, et al.* (2015). Prognostic value of ERCC1, RRM1, BRCA1 and SETDB1 in early stage of non-small cell lung cancer. Clin Transl Oncol.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357-359.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25.

Larson, B.J., Van, M.V., Nakayama, T., and Engebrecht, J. (2016). Plasticity in the Meiotic Epigenetic Landscape of Sex Chromosomes in Caenorhabditis Species. Genetics *203*, 1641-1658.

**Lawson, K.A., Teteak, C.J., Gao, J., Li, N., Hacquebord, J.*, et al.* (2013a). ESET histone methyltransferase regulates osteoblastic differentiation of mesenchymal stem cells during postnatal bone development. FEBS Lett *587*, 3961-3967.

**Lawson, K.A., Teteak, C.J., Zou, J., Hacquebord, J., Ghatan, A.*, et al.* (2013b). Mesenchyme-specific knockout of ESET histone methyltransferase causes ectopic hypertrophy and terminal differentiation of articular chondrocytes. J Biol Chem *288*, 32119-32125.

**Lee, J., Hagerty, S., Cormier, K.A., Kim, J., Kung, A.L.*, et al.* (2008). Monoallele deletion of CBP leads to pericentromeric heterochromatin condensation through ESET expression and histone H3 (K9) methylation. Hum Mol Genet *17*, 1774-1782.

Lee, J.K., and Kim, K.C. (2013). DZNep, inhibitor of S-adenosylhomocysteine hydrolase, down-regulates expression of SETDB1 H3K9me3 HMTase in human lung cancer cells. Biochem Biophys Res Commun *438*, 647-652.

**Lessel, D., Gamulin, M., Kulis, T., Toliat, M.R., Grgic, M.*, et al.* (2012). Replication of genetic susceptibility loci for testicular germ cell cancer in the Croatian population. Carcinogenesis *33*, 1548-1552.

**Leung, D., Du, T., Wagner, U., Xie, W., Lee, A.Y.*, et al.* (2014). Regulation of DNA methylation turnover at LTR retrotransposons and imprinted loci by the histone methyltransferase Setdb1. Proc Natl Acad Sci U S A *111*, 6690-6695.

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

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

**Lin, T.Y., Chu, Y.Y., Yang, Y.C., Hsu, S.W., Liu, S.T.*, et al.* (2014). MCAF1 and Rta-activated BZLF1 transcription in Epstein-Barr virus. PLoS One *9*, e90698.

**Liu, L., Ishihara, K., Ichimura, T., Fujita, N., Hino, S.*, et al.* (2009). MCAF1/AM is involved in Sp1-mediated maintenance of cancer-associated telomerase activity. J Biol Chem *284*, 5165-5174.

**Liu, S., Brind'Amour, J., Karimi, M.M., Shirane, K., Bogutz, A.*, et al.* (2014). Setdb1 is required for germline development and silencing of H3K9me3-marked endogenous retroviruses in primordial germ cells. Genes Dev *28*, 2041-2055.

**Lohmann, F., Loureiro, J., Su, H., Fang, Q., Lei, H.*, 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.

**Loyola, A., Tagami, H., Bonaldi, T., Roche, D., Quivy, J.P.*, et al.* (2009). The HP1alpha-CAF1-SetDB1-containing complex provides H3K9me1 for Suv39-mediated K9me3 in pericentric heterochromatin. EMBO Rep *10*, 769-775.

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

Lundberg, L.E., Stenberg, P., and Larsson, J. (2013). HP1a, Su(var)3-9, SETDB1 and POF stimulate or repress gene expression depending on genomic position, gene length and expression pattern in Drosophila melanogaster. Nucleic Acids Res *41*, 4481-4494.

**Luzzani, C., Solari, C., Losino, N., Ariel, W., Romorini, L.*, et al.* (2011). Modulation of chromatin modifying factors' gene expression in embryonic and induced pluripotent stem cells. Biochem Biophys Res Commun *410*, 816-822.

Lyon, M.F. (1961). Gene action in the X-chromosome of the mouse (Mus musculus L.). Nature *190*, 372-373.

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

**Macfarlan, T.S., Gifford, W.D., Agarwal, S., Driscoll, S., Lettieri, K.*, et al.* (2011). Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A. Genes Dev *25*, 594-607.

**Macgregor, S., Montgomery, G.W., Liu, J.Z., Zhao, Z.Z., Henders, A.K.*, et al.* (2011). Genome-wide association study identifies a new melanoma susceptibility locus at 1q21.3. Nat Genet *43*, 1114-1118.

**Maksakova, I.A., Goyal, P., Bullwinkel, J., Brown, J.P., Bilenky, M.*, et al.* (2011). H3K9me3-binding proteins are dispensable for SETDB1/H3K9me3-dependent retroviral silencing. Epigenetics Chromatin *4*, 12.

**Maksakova, I.A., Thompson, P.J., Goyal, P., Jones, S.J., Singh, P.B.*, et al.* (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.

Mani, R., St Onge, R.P., Hartman, J.L.t., Giaever, G., and Roth, F.P. (2008). Defining genetic interaction. Proc Natl Acad Sci U S A *105*, 3461-3466.

**Mao, H., Zhu, C., Zong, D., Weng, C., Yang, X.*, et al.* (2015). The Nrde Pathway Mediates Small-RNA-Directed Histone H3 Lysine 27 Trimethylation in Caenorhabditis elegans. Curr Biol *25*, 2398-2403.

**Martin, F.J., Xu, Y., Lohmann, F., Ciccone, D.N., Nicholson, T.B.*, et al.* (2015). KMT1E-mediated chromatin modifications at the FcgammaRIIb promoter regulate thymocyte development. Genes Immun *16*, 162-169.

**Matsui, T., Leung, D., Miyashita, H., Maksakova, I.A., Miyachi, H.*, et al.* (2010). Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. Nature *464*, 927-931.

**Matsumura, Y., Nakaki, R., Inagaki, T., Yoshida, A., Kano, Y.*, et al.* (2015). H3K4/H3K9me3 Bivalent Chromatin Domains Targeted by Lineage-Specific DNA Methylation Pauses Adipocyte Differentiation. Mol Cell *60*, 584-596.

**Mazzoni, E.O., Mahony, S., Iacovino, M., Morrison, C.A., Mountoufaris, G.*, et al.* (2011). Embryonic stem cell-based mapping of developmental transcriptional programs. Nat Methods *8*, 1056-1058.

**McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T.*, et al.* (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol *28*, 495-501.

**McMurchy, A.N., Stempor, P., Gaarenstroom, T., Wysolmerski, B., Dong, Y.*, et al.* (2017). A team of heterochromatin factors collaborates with small RNA pathways to combat repetitive elements and germline stress. Elife *6*.

**Melcher, M., Schmid, M., Aagaard, L., Selenko, P., Laible, G.*, et al.* (2000). Structure-function analysis of SUV39H1 reveals a dominant role in heterochromatin organization, chromosome segregation, and mitotic progression. Mol Cell Biol *20*, 3728-3741.

**Minkovsky, A., Sahakyan, A., Rankin-Gee, E., Bonora, G., Patel, S.*, et al.* (2014). The Mbd1-Atf7ip-Setdb1 pathway contributes to the maintenance of X chromosome inactivation. Epigenetics Chromatin *7*, 12.

Mozzetta, C., Boyarchuk, E., Pontis, J., and Ait-Si-Ali, S. (2015). Sound of silence: the properties and functions of repressive Lys methyltransferases. Nat Rev Mol Cell Biol *16*, 499-513.

**Mulholland, C.B., Smets, M., Schmidtmann, E., Leidescher, S., Markaki, Y.*, et al.* (2015). A modular open platform for systematic functional studies under physiological conditions. Nucleic Acids Res *43*, e112.

**Musri, M.M., Carmona, M.C., Hanzu, F.A., Kaliman, P., Gomis, R.*, et al.* (2010). Histone demethylase LSD1 regulates adipogenesis. J Biol Chem *285*, 30034-30041.

Na, H.H., Noh, H.J., Cheong, H.M., Kang, Y., and Kim, K.C. (2016). SETDB1 mediated FosB expression increases the cell proliferation rate during anticancer drug therapy. BMB Rep.

Nakayama, J., Rice, J.C., Strahl, B.D., Allis, C.D., and Grewal, S.I. (2001). Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. Science *292*, 110-113.

Noh, H.J., Kim, K.A., and Kim, K.C. (2014). p53 down-regulates SETDB1 gene expression during paclitaxel induced-cell death. Biochem Biophys Res Commun *446*, 43-48.

**Nomura, N., Nagase, T., Miyajima, N., Sazuka, T., Tanaka, A.*, et al.* (1994). Prediction of the coding sequences of unidentified human genes. II. The coding sequences of 40 new genes (KIAA0041-KIAA0080) deduced by analysis of cDNA clones from human cell line KG-1 (supplement). DNA Res *1*, 251-262.

Nuber, A. (Diploma thesis, 2017). Die Phänotypisierung der Mausmutante Atf7ip-bgal und die charakterisierung von Atf7ip in murinen embryonalen Stammzellen und Fibroblasten. DIW-MTA Diplomarbeit.

**O'Carroll, D., Scherthan, H., Peters, A.H., Opravil, S., Haynes, A.R.*, et al.* (2000). Isolation and characterization of Suv39h2, a second histone H3 methyltransferase gene that displays testis-specific expression. Mol Cell Biol *20*, 9423-9433.

**Olcina, M.M., Leszczynska, K.B., Senra, J.M., Isa, N.F., Harada, H.*, et al.* (2016). H3K9me3 facilitates hypoxia-induced p53-dependent apoptosis through repression of APAK. Oncogene *35*, 793-799.

Ong, C.T., and Corces, V.G. (2014). CTCF: an architectural protein bridging genome topology and function. Nat Rev Genet *15*, 234-246.

**Otte, A., Gohring, G., Steinemann, D., Schlegelberger, B., Groos, S.*, et al.* (2012). A tumor-derived population (SCCOHT-1) as cellular model for a small cell ovarian carcinoma of the hypercalcemic type. Int J Oncol *41*, 765-775.

**Pasquarella, A., Ebert, A., Pereira de Almeida, G., Hinterberger, M., Kazerani, M.*, et al.* (2016). Retrotransposon derepression leads to activation of the unfolded protein response and apoptosis in pro-B cells. Development *143*, 1788-1799.

**Peters, A.H., Kubicek, S., Mechtler, K., O'Sullivan, R.J., Derijck, A.A.*, et al.* (2003). Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. Mol Cell *12*, 1577-1589.

**Peters, A.H., O'Carroll, D., Scherthan, H., Mechtler, K., Sauer, S.*, et al.* (2001). Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. Cell *107*, 323-337.

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

Poleshko, A., Einarson, M.B., Shalginskikh, N., Zhang, R., Adams, P.D._, et al._ (2010). Identification of a functional network of human epigenetic silencing factors. J Biol Chem *285*, 422-433.

Ponting, C.P. (1997). Tudor domains in proteins that interact with RNA. Trends Biochem Sci *22*, 51-52.

Popp, M.W., and Maquat, L.E. (2016). Leveraging Rules of Nonsense-Mediated mRNA Decay for Genome Engineering and Personalized Medicine. Cell *165*, 1319-1322.

Poulin, G., Dong, Y., Fraser, A.G., Hopper, N.A., and Ahringer, J. (2005). Chromatin regulation and sumoylation in the inhibition of Ras-induced vulval development in Caenorhabditis elegans. EMBO J *24*, 2613-2623.

Quenneville, S., Verde, G., Corsinotti, A., Kapopoulou, A., Jakobsson, J._, et al._ (2011). In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. Mol Cell *44*, 361-372.

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B.D., Sun, Z.W._, et al._ (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature *406*, 593-599.

Rebollo, R., Karimi, M.M., Bilenky, M., Gagnier, L., Miceli-Royer, K._, et al._ (2011). Retrotransposon-induced heterochromatin spreading in the mouse revealed by insertional polymorphisms. PLoS Genet *7*, e1002301.

Rebollo, R., Miceli-Royer, K., Zhang, Y., Farivar, S., Gagnier, L._, et al._ (2012a). Epigenetic interplay between mouse endogenous retroviruses and host genes. Genome Biol *13*, R89.

Rebollo, R., Romanish, M.T., and Mager, D.L. (2012b). Transposable elements: an abundant and natural source of regulatory sequences for host genes. Annu Rev Genet *46*, 21-42.

Regina, C., Compagnone, M., Peschiaroli, A., Lena, A., Annicchiarico-Petruzzelli, M._, et al._ (2016). Setdb1, a novel interactor of DeltaNp63, is involved in breast tumorigenesis. Oncotarget.

Reichmann, J., Crichton, J.H., Madej, M.J., Taggart, M., Gautier, P._, et al._ (2012). Microarray analysis of LTR retrotransposon silencing identifies Hdac1 as a regulator of retrotransposon expression in mouse embryonic stem cells. PLoS Comput Biol *8*, e1002486.

Richards, E.J., and Elgin, S.C. (2002). Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects. Cell *108*, 489-500.

Riddle, N.C., Jung, Y.L., Gu, T., Alekseyenko, A.A., Asker, D._, et al._ (2012). Enrichment of HP1a on Drosophila chromosome 4 genes creates an alternate chromatin structure critical for regulation in this heterochromatic domain. PLoS Genet *8*, e1002954.

Rigbolt, K.T., Prokhorova, T.A., Akimov, V., Henningsen, J., Johansen, P.T._, et al._ (2011). System-wide temporal characterization of the proteome and phosphoproteome of human embryonic stem cell differentiation. Sci Signal *4*, rs3.

Rivera, C., Saavedra, F., Alvarez, F., Diaz-Celis, C., Ugalde, V._, et al._ (2015). Methylation of histone H3 lysine 9 occurs during translation. Nucleic Acids Res *43*, 9097-9106.

Rodriguez-Paredes, M., Martinez de Paz, A., Simo-Riudalbas, L., Sayols, S., Moutinho, C._, et al._ (2014). Gene amplification of the histone methyltransferase SETDB1 contributes to human lung tumorigenesis. Oncogene *33*, 2807-2813.

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

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

RStudio-Team (2016). RStudio: Integrated Development for R. URL: http://www.rstudio.org/.

Ryu, H., Lee, J., Hagerty, S.W., Soh, B.Y., McAlpin, S.E._, et al._ (2006). ESET/SETDB1 gene expression and histone H3 (K9) trimethylation in Huntington's disease. Proc Natl Acad Sci U S A *103*, 19176-19181.

Sadic, D., Schmidt, K., Groh, S., Kondofersky, I., Ellwart, J._, et al._ (2015). Atrx promotes heterochromatin formation at retrotransposons. EMBO Rep *16*, 836-850.

Sanchez, C., Sanchez, I., Demmers, J.A., Rodriguez, P., Strouboulis, J._, et al._ (2007). Proteomics analysis of Ring1B/Rnf2 interactors identifies a novel complex with the Fbxl10/Jhdm1B histone demethylase and the Bcl6 interacting corepressor. Mol Cell Proteomics *6*, 820-834.

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

Sasai, N., Saitoh, N., Saitoh, H., and Nakao, M. (2013). The transcriptional cofactor MCAF1/ATF7IP is involved in histone gene expression and cellular senescence. PLoS One *8*, e68478.

Schek, N., Cooke, C., and Alwine, J.C. (1992). Definition of the upstream efficiency element of the simian virus 40 late polyadenylation signal by using in vitro analyses. Mol Cell Biol *12*, 5386-5393.

Schultz, D.C., Ayyanathan, K., Negorev, D., Maul, G.G., and Rauscher, F.J., 3rd (2002). SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. Genes Dev *16*, 919-932.

Sekiyama, N., Ikegami, T., Yamane, T., Ikeguchi, M., Uchimura, Y._, et al._ (2008). Structure of the small ubiquitin-like modifier (SUMO)-interacting motif of MBD1-containing chromatin-associated factor 1 bound to SUMO-3. J Biol Chem *283*, 35966-35975.

Seum, C., Reo, E., Peng, H., Rauscher, F.J., 3rd, Spierer, P._, et al._ (2007). Drosophila SETDB1 is required for chromosome 4 silencing. PLoS Genet *3*, e76.

Shalem, O., Sanjana, N.E., and Zhang, F. (2015). High-throughput functional genomics using CRISPR-Cas9. Nat Rev Genet *16*, 299-311.

Sharif, J., Endo, T.A., Nakayama, M., Karimi, M.M., Shimada, M.*, et al. (2016). Activation of Endogenous Retroviruses in Dnmt1(-/-) ESCs Involves Disruption of SETDB1-Mediated Repression by NP95 Binding to Hemimethylated DNA. Cell Stem Cell *19*, 81-94.

Shinkai, Y., and Tachibana, M. (2011). H3K9 methyltransferase G9a and the related molecule GLP. Genes Dev *25*, 781-788.

Shinoda, N., Obata, F., Zhang, L., and Miura, M. (2016). Drosophila SETDB1 and caspase cooperatively fine-tune cell fate determination of sensory organ precursor. Genes Cells.

Sim, Y.J., Kim, M.S., Nayfeh, A., Yun, Y.J., Kim, S.J.*, et al. (2017). 2i Maintains a Naive Ground State in ESCs through Two Distinct Epigenetic Mechanisms. Stem Cell Reports *8*, 1312-1328.

Sims, R.J., 3rd, Belotserkovskaya, R., and Reinberg, D. (2004). Elongation by RNA polymerase II: the short and long of it. Genes Dev *18*, 2437-2468.

Singh, S., Ghosh, P., and Hatfull, G.F. (2013). Attachment site selection and identity in Bxb1 serine integrase-mediated site-specific recombination. PLoS Genet *9*, e1003490.

Song, G.G., Kim, J.H., and Lee, Y.H. (2013). Genome-wide pathway analysis in major depressive disorder. J Mol Neurosci *51*, 428-436.

Song, Y.J., Choi, J.H., and Lee, H. (2015). Setdb1 is required for myogenic differentiation of C2C12 myoblast cells via maintenance of MyoD expression. Mol Cells *38*, 362-372.

Spyropoulou, A., Gargalionis, A., Dalagiorgou, G., Adamopoulos, C., Papavassiliou, K.A.*, et al. (2014). Role of histone lysine methyltransferases SUV39H1 and SETDB1 in gliomagenesis: modulation of cell proliferation, migration, and colony formation. Neuromolecular Med *16*, 70-82.

Sridharan, R., Gonzales-Cope, M., Chronis, C., Bonora, G., McKee, R.*, et al. (2013). Proteomic and genomic approaches reveal critical functions of H3K9 methylation and heterochromatin protein-1gamma in reprogramming to pluripotency. Nat Cell Biol *15*, 872-882.

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

Stabell, M., Bjorkmo, M., Aalen, R.B., and Lambertsson, A. (2006). The Drosophila SET domain encoding gene dEset is essential for proper development. Hereditas *143*, 177-188.

Stocking, C., and Kozak, C.A. (2008). Murine endogenous retroviruses. Cell Mol Life Sci *65*, 3383-3398.

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

Suka, N., Suka, Y., Carmen, A.A., Wu, J., and Grunstein, M. (2001). Highly specific antibodies determine histone acetylation site usage in yeast heterochromatin and euchromatin. Mol Cell *8*, 473-479.

Sun, L., and Fang, J. (2016). E3-Independent Constitutive Monoubiquitination Complements Histone Methyltransferase Activity of SETDB1. Mol Cell *62*, 958-966.

Sun, Q.Y., Ding, L.W., Xiao, J.F., Chien, W., Lim, S.L.*, et al. (2015). SETDB1 accelerates tumourigenesis by regulating the WNT signalling pathway. J Pathol *235*, 559-570.

Sun, Y., Wei, M., Ren, S.C., Chen, R., Xu, W.D.*, et al. (2014). Histone methyltransferase SETDB1 is required for prostate cancer cell proliferation, migration and invasion. Asian J Androl *16*, 319-324.

Tachibana, K., Gotoh, E., Kawamata, N., Ishimoto, K., Uchihara, Y.*, et al. (2015). Analysis of the subcellular localization of the human histone methyltransferase SETDB1. Biochem Biophys Res Commun *465*, 725-731.

Tachibana, M., Ueda, J., Fukuda, M., Takeda, N., Ohta, T.*, et al. (2005). Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9. Genes Dev *19*, 815-826.

Takada, I., Mihara, M., Suzawa, M., Ohtake, F., Kobayashi, S.*, et al. (2007). A histone lysine methyltransferase activated by non-canonical Wnt signalling suppresses PPAR-gamma transactivation. Nat Cell Biol *9*, 1273-1285.

Takikita, S., Muro, R., Takai, T., Otsubo, T., Kawamura, Y.I.*, et al. (2016). A Histone Methyltransferase ESET Is Critical for T Cell Development. J Immunol *197*, 2269-2279.

Tan, S.L., Nishi, M., Ohtsuka, T., Matsui, T., Takemoto, K.*, et al. (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. Biosci Biotechnol Biochem *74*, 1302-1305.

Tchasovnikarova, I.A., Timms, R.T., Matheson, N.J., Wals, K., Antrobus, R.*, et al. (2015). Epigenetic silencing by the HUSH complex mediates position-effect variegation in human cells. Science *348*, 1481-1485.

Thompson, P.J., Dulberg, V., Moon, K.M., Foster, L.J., Chen, C.*, et al. (2015). hnRNP K coordinates transcriptional silencing by SETDB1 in embryonic stem cells. PLoS Genet *11*, e1004933.

Tian, Y., Garcia, G., Bian, Q., Steffen, K.K., Joe, L.*, et al. (2016). Mitochondrial Stress Induces Chromatin Reorganization to Promote Longevity and UPR(mt). Cell *165*, 1197-1208.

Timms, R.T., Tchasovnikarova, I.A., Antrobus, R., Dougan, G., and Lehner, P.J. (2016). ATF7IP-Mediated Stabilization of the Histone Methyltransferase SETDB1 Is Essential for Heterochromatin Formation by the HUSH Complex. Cell Rep *17*, 653-659.

**Towbin, B.D., Gonzalez-Aguilera, C., Sack, R., Gaidatzis, D., Kalck, V.*, et al.* (2012). Step-wise methylation of histone H3K9 positions heterochromatin at the nuclear periphery. Cell *150*, 934-947.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105-1111.

**Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D.*, et al.* (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc *7*, 562-578.

**Tschiersch, B., Hofmann, A., Krauss, V., Dorn, R., Korge, G.*, et al.* (1994). The protein encoded by the Drosophila position-effect variegation suppressor gene Su(var)3-9 combines domains of antagonistic regulators of homeotic gene complexes. EMBO J *13*, 3822-3831.

**Turnbull, C., Rapley, E.A., Seal, S., Pernet, D., Renwick, A.*, et al.* (2010). Variants near DMRT1, TERT and ATF7IP are associated with testicular germ cell cancer. Nat Genet *42*, 604-607.

Turner, B.M. (2000). Histone acetylation and an epigenetic code. Bioessays *22*, 836-845.

**Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y.*, et al.* (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat Methods *13*, 731-740.

Tzeng, T.Y., Lee, C.H., Chan, L.W., and Shen, C.K. (2007). Epigenetic regulation of the Drosophila chromosome 4 by the histone H3K9 methyltransferase dSETDB1. Proc Natl Acad Sci U S A *104*, 12691-12696.

**Uchimura, Y., Ichimura, T., Uwada, J., Tachibana, T., Sugahara, S.*, et al.* (2006). Involvement of SUMO modification in MBD1- and MCAF1-mediated heterochromatin formation. J Biol Chem *281*, 23180-23190.

**Ushijima, Y., Inoue, Y.H., Konishi, T., Kitazawa, D., Yoshida, H.*, et al.* (2012). Roles of histone H3K9 methyltransferases during Drosophila spermatogenesis. Chromosome Res *20*, 319-331.

**Valcarcel-Ocete, L., Alkorta-Aranburu, G., Iriondo, M., Fullaondo, A., Garcia-Barcina, M.*, et al.* (2015). Exploring Genetic Factors Involved in Huntington Disease Age of Onset: E2F2 as a New Potential Modifier Gene. PLoS One *10*, e0131573.

**Valle-Garcia, D., Qadeer, Z.A., McHugh, D.S., Ghiraldini, F.G., Chowdhury, A.H.*, et al.* (2016). ATRX binds to atypical chromatin domains at the 3' exons of zinc finger genes to preserve H3K9me3 enrichment. Epigenetics *11*, 398-414.

**Van Duyne, R., Easley, R., Wu, W., Berro, R., Pedati, C.*, et al.* (2008). Lysine methylation of HIV-1 Tat regulates transcriptional activity of the viral LTR. Retrovirology *5*, 40.

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

**Verschure, P.J., van der Kraan, I., de Leeuw, W., van der Vlag, J., Carpenter, A.E.*, et al.* (2005). In vivo HP1 targeting causes large-scale chromatin condensation and enhanced histone lysine methylation. Mol Cell Biol *25*, 4552-4564.

Walter, M., Teissandier, A., Perez-Palacios, R., and Bourc'his, D. (2016). An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. Elife *5*.

Wang, G.Z., and Goff, S.P. (2017). Transcriptional Silencing of Moloney Murine Leukemia Virus in Human Embryonic Carcinoma Cells. J Virol *91*.

**Wang, H., An, W., Cao, R., Xia, L., Erdjument-Bromage, H.*, et al.* (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, J., Hevi, S., Kurash, J.K., Lei, H., Gay, F.*, et al.* (2009). The lysine demethylase LSD1 (KDM1) is required for maintenance of global DNA methylation. Nat Genet *41*, 125-129.

Wang, X., Hou, J., Quedenau, C., and Chen, W. (2016). Pervasive isoform-specific translational regulation via alternative transcription start sites in mammals. Mol Syst Biol *12*, 875.

**Wang, X., Pan, L., Wang, S., Zhou, J., McDowell, W.*, et al.* (2011). Histone H3K9 trimethylase Eggless controls germline stem cell maintenance and differentiation. PLoS Genet *7*, e1002426.

**Watanabe, H., Soejima, K., Yasuda, H., Kawada, I., Nakachi, I.*, et al.* (2008). Deregulation of histone lysine methyltransferases contributes to oncogenic transformation of human bronchoepithelial cells. Cancer Cell Int *8*, 15.

**Waterfield, M., Khan, I.S., Cortez, J.T., Fan, U., Metzger, T.*, et al.* (2014). The transcriptional regulator Aire coopts the repressive ATF7ip-MBD1 complex for the induction of immunotolerance. Nat Immunol *15*, 258-265.

Wei, Y., Yu, L., Bowen, J., Gorovsky, M.A., and Allis, C.D. (1999). Phosphorylation of histone H3 is required for proper chromosome condensation and segregation. Cell *97*, 99-109.

Wen, B., Wu, H., Shinkai, Y., Irizarry, R.A., and Feinberg, A.P. (2009). Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. Nat Genet *41*, 246-250.

**Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P.*, et al.* (2007). A unified classification system for eukaryotic transposable elements. Nat Rev Genet *8*, 973-982.

**Wong, C.M., Wei, L., Law, C.T., Ho, D.W., Tsang, F.H.*, et al.* (2016). Up-regulation of histone methyltransferase SETDB1 by multiple mechanisms in hepatocellular carcinoma promotes cancer metastasis. Hepatology *63*, 474-487.

**Wu, P.C., Lu, J.W., Yang, J.Y., Lin, I.H., Ou, D.L.*, et al.* (2014). H3K9 histone methyltransferase, KMT1E/SETDB1, cooperates with the SMAD2/3 pathway to suppress lung cancer metastasis. Cancer Res *74*, 7333-7343.

Xie, M., Hong, C., Zhang, B., Lowdon, R.F., Xing, X.*, et al.* (2013). DNA hypomethylation within specific transposable element families associates with tissue-specific enhancer landscape. Nat Genet *45*, 836-841.

Xu, Q., Goldstein, J., Wang, P., Gadi, I.K., Labreche, H.*, et al.* (2016). Chromosomal microarray analysis in clinical evaluation of neurodevelopmental disorders-reporting a novel deletion of SETDB1 and illustration of counseling challenge. Pediatr Res *80*, 371-381.

Yahiro, K., Higashihori, N., and Moriyama, K. (2017). Histone methyltransferase Setdb1 is indispensable for Meckel's cartilage development. Biochem Biophys Res Commun *482*, 883-888.

Yang, B.X., El Farran, C.A., Guo, H.C., Yu, T., Fang, H.T.*, et al.* (2015). Systematic identification of factors for provirus silencing in embryonic stem cells. Cell *163*, 230-245.

Yang, L., Lawson, K.A., Teteak, C.J., Zou, J., Hacquebord, J.*, et al.* (2013). ESET histone methyltransferase is essential to hypertrophic differentiation of growth plate chondrocytes and formation of epiphyseal plates. Dev Biol *380*, 99-110.

Yang, L., Mei, Q., Zielinska-Kwiatkowska, A., Matsui, Y., Blackburn, M.L.*, et al.* (2003). An ERG (ets-related gene)-associated histone methyltransferase interacts with histone deacetylases 1/2 and transcription co-repressors mSin3A/B. Biochem J *369*, 651-657.

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

Yeap, L.S., Hayashi, K., and Surani, M.A. (2009). ERG-associated protein with SET domain (ESET)-Oct4 interaction regulates pluripotency and represses the trophectoderm lineage. Epigenetics Chromatin *2*, 12.

Yoon, J., Lee, K.S., Park, J.S., Yu, K., Paik, S.G.*, et al.* (2008). dSETDB1 and SU(VAR)3-9 sequentially function during germline-stem cell differentiation in Drosophila melanogaster. PLoS One *3*, e2234.

Yuan, P., Han, J., Guo, G., Orlov, Y.L., Huss, M.*, et al.* (2009). Eset partners with Oct4 to restrict extraembryonic trophoblast lineage potential in embryonic stem cells. Genes Dev *23*, 2507-2520.

Zeller, P., Padeken, J., van Schendel, R., Kalck, V., Tijsterman, M.*, et al.* (2016). Histone H3K9 methylation is dispensable for Caenorhabditis elegans development but suppresses RNA:DNA hybrid-associated repeat instability. Nat Genet *48*, 1385-1395.

Zhang, H., Cai, K., Wang, J., Wang, X., Cheng, K.*, et al.* (2014). MiR-7, inhibited indirectly by lincRNA HOTAIR, directly inhibits SETDB1 and reverses the EMT of breast cancer stem cells by downregulating the STAT3 pathway. Stem Cells *32*, 2858-2868.

Zhang, X., Guo, C., Chen, Y., Shulha, H.P., Schnetz, M.P.*, et al.* (2008). Epitope tagging of endogenous proteins for genome-wide ChIP-chip studies. Nat Methods *5*, 163-165.

Zhang, Y., Ng, H.H., Erdjument-Bromage, H., Tempst, P., Bird, A.*, et al.* (1999). Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. Genes Dev *13*, 1924-1935.

Zheng, C., Karimzadegan, S., Chiang, V., and Chalfie, M. (2013). Histone methylation restrains the expression of subtype-specific genes during terminal neuronal differentiation in Caenorhabditis elegans. PLoS Genet *9*, e1004017.