Unraveling the KRAB/KAP1 control of transposable elements in pluripotent and somatic cells

PAR

Gabriela ECCO

acceptée sur proposition du jury:

Prof. M. Lütolf, président du jury
Prof. D. Trono, Dr P. Turelli, directeurs de thèse
Prof. C. Feschotte, rapporteur
Prof. M. Branco, rapporteur
Prof. B. Deplancke, rapporteur

EPFL

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2016

"O correr da vida embrulha tudo.
A vida é assim: esquenta e esfria,
aperta e daí afrouxa,
sossega e depois desinquieta.
O que ela quer da gente é coragem."

Guimarães Rosa

Table of Contents

Acknowledgements

The first person I would like to thank is my thesis director Didier Trono. Thank you for welcoming me in your lab first as a Summer student and then as a PhD candidate. I have learned so much over these years and, for all of it, I am thankful. Thanks for always being available and supportive, for the freedom you gave me to work, for helping me grow personally and professionally, and for teaching me to see the bigger picture in my projects. Your passion for knowledge and science are inspiring. Thank you for sharing them with me.

I am grateful to my co-supervisor Priscilla Turelli, for her support over these years, for all the advices and discussions, and for always being available to help. Thank you for taking the time to carefully read this thesis and so many other things.

I am thankful to Helen Rowe, who kindly supervised me and guided me through the first steps in the lab, and whose ideas started this project.

I would also like to acknowledge the members of the jury, Prof. Cédric Feschotte, Prof. Miguel Branco, Prof. Bart Deplancke, and Prof. Matthias Lütolf for their availability and for kindly accepting to evaluate my thesis.

This work would not have been possible without the technical and scientific help of Marco Cassano, Annamaria Kauzlaric, Nathaly Castro-Diaz, Julien Duc, Andrea Coluccio, Sandra Offner, Michaël Imbeault, Adamandia Kapopoulou, Benyamin Yazdanpanah, Marc Friedli, and Suk Min Jang, who directly contributed to the manuscripts in this thesis, and whose help I truly appreciate.

I am deeply thankful to past and present members of the Trono lab. Thank you for the help, discussions, for the apéros, and for the great moments shared inside

Summary

Transposable elements (TEs) are DNA sequences able to change position in the genome, and represent more than 40% of mammalian genetic material. TEs can have positive or detrimental effects on the host, being both important motors of evolution and genomic threats, and are associated with diseases such as cancer and diabetes. Growing evidence indicates the host co-opts TEs for its benefit, with examples such as syncytins – genes important for placental formation that derive from TEs. Given their potential damaging role, the host needs to repress TEs, especially during early embryogenesis. KRAB-containing zinc finger proteins (KRAB-ZFPs) are important regulators of TEs, which they repress with their cofactor KAP1. In stem cells, the KRAB/KAP1 complex irreversibly silence TEs and is believed to be dispensable in adult cells. KRAB-ZFPs constitute one of the largest families of transcriptional regulators encoded by higher vertebrates, but functional information is missing for most of its members.

To shed light on this important family of transcription factors, we first developed a large-scale functional screen matching murine TEs with their cognate KRAB-ZFP. The screen identified KRAB-ZFP809 as the ligand of its previously mapped DNA target in the murine leukemia virus genome. Our method further singled out two novel KRAB-ZFPs binding to TE sequences, as confirmed by repression and binding assays. One of them, Gm6871, was identified to regulate LINE-1 elements together with KAP1 in embryonic stem cells. KAP1 regulation followed a chronologically conditioned pattern, repressing elements that entered the mouse genome between 5.6 and 3.8 million years ago, suggestive of an evolutionary arms race between host KRAB-ZFPs and TEs. Secondly, we found the paralogs ZFP932 and Gm15446 to bind overlapping but distinguishable subsets of ERVK (endogenous retrovirus K), to repress these elements in embryonic stem cells, and to regulate secondarily the expression of neighboring

genes. Furthermore, we uncovered that these KRAB-ZFPs and KAP1 control TEs in adult tissues, in cell culture and *in vivo*, where they partner up to modulate cellular genes.

Our results establish an efficient screening method to identify KRAB-ZFPs DNA targets, opening the way to functional analyses of this major class of transcriptional repressors. Most importantly, our study strongly suggests that *KRAB-ZFP* genes are not only part of an evolutionary arms race with TEs, but also participate in the domestication of these elements for the benefit of the host. Given the abundance and high degree of species-specificity of TEs and KRAB-ZFPs, these results have important implications for understanding the biology of higher vertebrates, including humans.

Sommario

I trasposoni sono sequenze di DNA capaci di cambiare la propria posizione all'interno del genoma, e costituiscono più del 40% dell'informazione genetica nei mammiferi. La presenza di trasposoni nel genoma può avere effetti sia benefici che deleteri per l'organismo ospite; essi rappresentano importanti motori dell'evoluzione ma possono anche minare all'integrità del genoma e sono spesso associati a patologie quali cancro e diabete. Un numero crescente di studi suggerisce che l'organismo ospite coopti i trasposoni a proprio vantaggio, come nel caso delle sincizine, geni coinvolti nella formazione della placenta che sono derivati da proteine virali presenti in alcuni trasposoni. Il potenziale effetto deleterio che questi elementi hanno sul genoma, richiede che l'ospite mantenga i trasposoni repressi, soprattutto durante i primi stadi dello sviluppo embrionale. Una famiglia di proteine a dita di zinco contenenti un dominio KRAB (KRAB-ZFPs) costituisce un gruppo d'importanti fattori di trascrizione coinvolti nella regolazione dei trasposoni, mantenendoli repressi grazie all'azione del loro co-fattore KAP1. In cellule staminali embrionali, il complesso proteico KRAB/KAP1 silenzia irreversibilmente la trascrizione dei trasposoni e si ritiene che l'azione repressiva di questi fattori non sia necessaria in cellule differenziate. Le KRAB-ZFPs rappresentano una delle più grandi famiglie di fattori di trascrizione presenti nei vertebrati, ma per gran parte di queste proteine mancano ancora informazioni funzionali.

Per comprendere meglio questa importante famiglia di fattori di trascrizione, abbiamo sviluppato uno screening funzionale su larga scala che permetta di associare trasposoni murini alla corrispettiva KRAB-ZFP in grado di legarli e reprimerli. Lo screening ha correttamente identificato il legame precedentemente descritto della KRAB-ZFP809 al genoma del virus della leucemia murino. Il nostro metodo ha permesso di identificare il legame di due

nuove KRAB-ZFP a specifiche sequenze di trasposoni, come confermato anche da esperimenti di legame e repressione sul genoma. Una di queste, Gm6871, è in grado di legare trasposoni della famiglia LINE-1 e di regolarne l'espressione in cellule staminali embrionali tramite KAP1. La regolazione operata da Gm6871/KAP1 segue un andamento cronologico, mantenendo repressi elementi che si sono integrati nel genoma murino tra 5.6 e 3.8 milioni di anni fa, suggerendo un braccio di ferro evolutivo tra le KRAB-ZFP dell'ospite e la replicazione dei trasposoni. Abbiamo inoltre caratterizzato due proteine paraloghe, ZFP932 e Gm15446 come fattori responsabili per il legame a la repressione di due sottogruppi sovrapposti ma distinti di ERVK (retrovirus endogeni K). Questi fattori mantengono repressi questi elementi in cellule staminali embrionali, e regolano anche l'espressione di geni situati nelle vicinanze. Grazie a esperimenti condotti sia su cellule in coltura che *in vivo*, abbiamo dimostrato come questi fattori, insieme a KAP1, controllino i trasposoni in cellule differenziate in tessuti adulti e collaborino alla modulazione dell'espressione di geni cellulari.

In questo lavoro descriviamo uno screening efficiente per identificare sequenze di DNA legate da KRAB-ZFP, aprendo la strada a più estese analisi funzionali su questa importante famiglia di repressori della trascrizione. Inoltre i nostri risultati suggeriscono che le KRAB-ZFP non solo partecipano al braccio di ferro evolutivo tra i trasposoni e l'ospite, ma sono anche coinvolte nell'addomesticare questi elementi a favore dell'organismo ospite. Data l'abbondanza e specificità di trasposoni e KRAB-ZFP in ogni singola specie, questo studio rivela forti implicazioni nella comprensione della biologia dei vertebrati, tra cui anche l'uomo.

Parole chiave: regolazione della trascrizione, epigenetica, trasposoni, retrovirus endogeni, KAP1 o TRIM28, KRAB-ZFP, evoluzione del genoma.

I. Introduction

1. Epigenetics and gene regulation

In the development of multicellular organisms, one zygote gives rise to different cell types. The DNA sequences present in this cell generate the most different cellular phenotypes, coding for a specialized set of molecules allowing a blood cell or a neuron to be exactly what they are. If their genome is the same, how are the cell-specific gene expression patterns established and maintained? The answer is "epigenetics". Epigenetics refers to stable alterations in gene expression that are heritable but do not involve changes of the underlying DNA sequence (Jaenisch and Bird, 2003).

In order to understand epigenetics, the DNA has to be considered in the context of chromatin. The genome of one diploid human cell is about 2 meters long, so it needs to be compacted to fit inside the cell and to be used (**Figure 1**) (Bloom and Joglekar, 2010). In eukaryotes, the DNA is first folded into nucleosomes, consisting of approximately 147bp of DNA wrapped around a histone octamer (with two copies of H2A, H2B, H3, and H4) (Margueron and Reinberg, 2010). Each core histone can be chemically modified and exchanged with variants. The nucleosome and the histone modifications form the primary structure of chromatin, which can be further packed three-dimensionally in the nucleus, originating higher-order chromatin structures. Heterochromatin, for instance, is a highly compacted form of chromatin with generally no active transcription and enriched for repeats, while euchromatin is decondensed and associated with active or regulatory regions. This packaging and the underlying epigenetic modifications ensure the correct DNA information is accessible at the right time in a given cell.
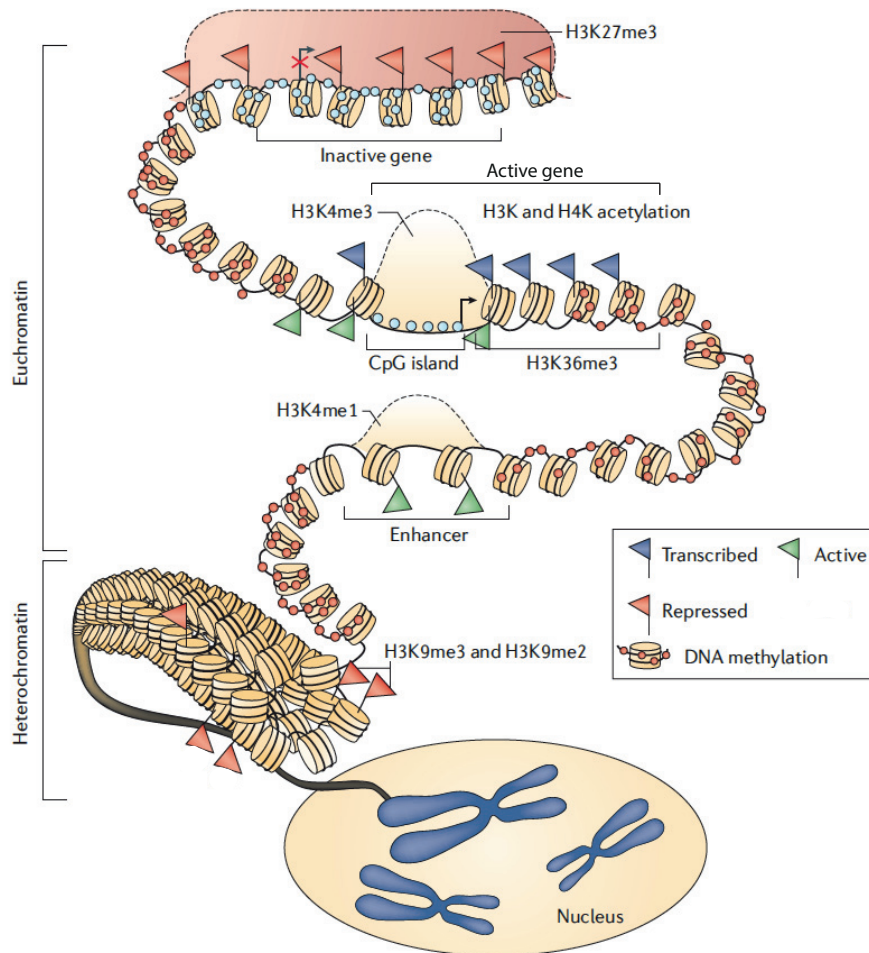
Figure 1. Schematic representation of eukaryotic chromatin structure. Chromatin compaction is represented, as well as DNA methylation and histone modifications. Genomic structures such as enhancers and active/inactive genes and the associated epigenetic marks are also depicted. *Modified from (Baylin and Jones, 2011)*.

The first layer of epigenetic regulation is methylation of DNA. In eukaryotes, DNA methylation normally occurs at cytosines at CpG sites, which are converted to 5-methylcytosine (5mC). DNA methylation is an inheritable mark on the DNA, present at about 70-80% of CpGs, and it is normally associated with transcriptional silencing. The addition of a methyl group to the cytosine alters DNA 3D structure and either inhibits or allows accessibility to the DNA by other molecules. In mammals, DNA methylation is catalyzed by the DNA

methyltransferases (DNMTs) DNMT1, DNMT3a, and DNMT3b (Goll and Bestor, 2005). Establishment of DNA methylation is mediated by *de novo* DNMTs 3a and b, while DNMT1 is a maintenance DNMT that ensures perpetuation of methylated cytosines upon DNA replication.

While DNA methylation was once perceived as a permanent silencing mark established in the early embryo, it is not as static as previously believed. DNA can be demethylated not only passively (through DNA replication), but also actively. Evidence for rapid and active DNA methylation has been observed even before the discovery of one of the key players in this process, the ten-eleven translocation (TET) proteins (Franchini et al., 2012). TETs oxidize 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC), 5-folmylcytosine (5fC), and/or 5-carboxylcytosine (5caC), intermediates in the removal of 5mC (Ito et al., 2010; Kohli and Zhang, 2013). These proteins have been implicated in the rapid decline of methylation observed in in primordial germ cells and early embryogenesis, although recent evidence suggests that TET3 is not required for paternal genome demethylation (Amouroux et al., 2016; Branco et al., 2012). Oxidized 5-mC bases are further lost by dilution or actively removed via DNA-repair mechanisms, involving the thymine DNA glycosylase (TDG) protein and the base-excision repair (BER) pathway (Kohli and Zhang, 2013; Weber et al., 2016).

Another important layer of epigenetic regulation comprises histone modifications. Histones have many aminoacids that undergo post-translational modifications with direct impacts (such as affecting interactions between nucleosomes) or indirect effects (such as driving the recruitment of different effector complexes and transcription factors). The most common and well-studied histone modifications are methylation and acetylation of histone tails (Kouzarides, 2007). The first histone modification to be discovered was histone acetylation, and it is associated with active transcription, DNA replication, and repair. Histone methylation at lysine residues is often associated with transcriptional repression, especially on histone 3 lysine 9 (H3K9), histone 3 lysine 27 (H3K27), and histone 4 lysine 20 (H4K20). Extra complexity arises from the fact that residues can be mono-, di-, or tri- methylated. Most histone

modifications are highly dynamic and reversible, as there are several enzymes that can deposit or remove these marks (writers and erasers, respectively). The diversity of these marks allows "crosstalk" between histone modifications in the establishment of different chromatin environments, as illustrated in **Figure 1** (Venkatesh and Workman, 2015). Besides methylation and acetylation, histones tails and globular domains can undergo a variety of other modifications, such as phosphorylation and ubiquitylation (Lawrence et al., 2016).

Finally, the three-dimensional architecture of chromatin also impacts gene regulation. For instance, insulators are boundary elements defined by their ability to prevent heterochromatin spreading. In mammals, the binding of CTCF protein to certain regions of the genome blocks enhancer activity and heterochromatin spreading (Zhou et al., 2011). More interestingly, metazoan genomes are organized in chromatin compartments called topologically associating domains (TADs), largely conserved between species and cell types (Ciabrelli and Cavalli, 2015). TADs are 3D chromatin folds that tend to favor interactions internally in the domain, rather than externally. Normally, a defined chromatin type can be assigned to each TAD (e.g. active or repressive chromatin), and the chromatin identity of a TAD guides its nuclear positioning.

2. Transposable elements

Transposable elements (TEs) are "jumping genes", or DNA sequences capable of moving from one location to another in the genome, and are present in all organisms, from bacteria to humans. TEs were discovered in the maize by the pioneer geneticist Barbara McClintock (McClintock, 1950). At the time, she described them as mobile genetic elements, named "controlling elements", not only moving in the genome, but also controlling gene expression. This notion was later developed by Britten and Davidson, who suggested that during evolution new structures and functions could arise from the co-option of TEs into regulatory elements (Britten and Davidson, 1971).

These ideas were met with a lot of skepticism, and for much of the 20th century the genome was considered as a static entity, with genes ordered in a linear fashion in the chromosomes. When acknowledged, these elements were still much referred to as "junk", "selfish", or "parasitic" (Doolittle and Sapienza, 1980; Orgel and Crick, 1980). Fortunately, the recognition of McClintock's work with the 1983 Nobel Prize in Physiology or Medicine and the development of high-throughput sequencing techniques have reinforced her view of TEs as "controlling elements".

TEs account for about half of the human and murine genomes, far more than the 1-2% portion encoding for proteins (Lander et al., 2001; Waterston et al., 2002). TEs are classified into two classes according to their transposition mechanism: Class I and Class II transposons (**Figure 2**) (Feschotte and Pritham, 2007; Rebollo et al., 2012c; Warren et al., 2015). Class II elements, or DNA transposons, replicate via a DNA intermediate, either by a cut-and-paste mechanism (classic DNA transposons harboring transposases), by rolling-circle DNA replication (*Helitrons*), or by mechanisms not yet fully understood (*Mavericks*) (likely through a self-encoded DNA polymerase). DNA transposons represent about 3% of the human genome and are not active in humans, with the last transposition-competent element in primates dating back to 37 million years (Pace and Feschotte, 2007; Padeken et al., 2015).

Class I elements, or retrotransposons, replicate via an RNA intermediate, using a copy-and-paste mechanism. Analogously to exogenous retroviruses such as HIV, the RNA intermediate is transcribed from the genome, it is reverse-transcribed into DNA by a reverse transcriptase, and a complete or partial copy is integrated into the genome. This mechanism allows expansion of retroelements in the genome, explaining why they encompass about 40% of both mouse and human DNA. They can be further divided into long terminal repeat (LTR)-containing retrolements and non-LTR elements.

Transposable elements	Structure (example)

Class I (Retrotransposons)

LTR retroelements
(ex: IAP, MusD, HERVK)

ERV

non-LTR retroelements

Autonomous
(ex: LINE-1)

LINE

Nonautonomous
(ex: SINE, SVA)

SINE

Class II (DNA transposons)
(ex: Mariner, Helitron)

Mariner
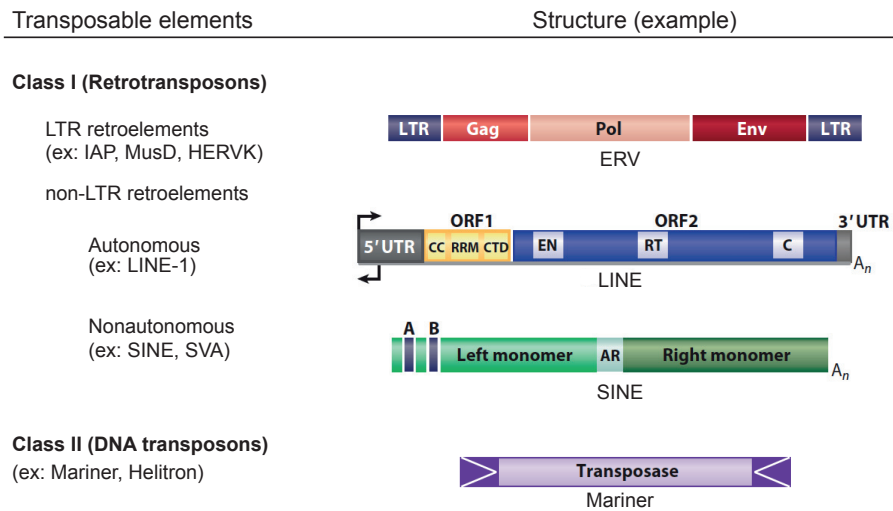
Figure 2. Classification of transposable elements. The different classes of transposable elements and examples of their structures are shown. LTR, long terminal repeat; Gag, group-specific antigen; Pol, polymerase; Env, envelope protein; UTR, untranslated region; CC, coiled coil; RRM, RNA recognition motif; CTD, carboxyl-terminal domain; EN, endonuclease; RT, reverse transcriptase; C, cysteine-rich domain; A_n, poly(A) tail; A and B, component sequences of the RNA polymerase III promoter; AR, the adenosine-rich segment separating the 7SL monomers. *Adapted from (Beck et al., 2011).*

The main representatives of LTR retroelements are endogenous retroviruses (ERVs), comprising about 8% and 12% of the human and mouse genome, respectively (Mager and Stoye, 2015). ERVs are reminiscent of ancient retroviral infections and, as such, are normally composed of prototypic retroviral *gag*, *pol*, and sometimes *env* genes, flanked by two LTRs. Despite their abundance, only a few hundreds are active and capable of retrotransposition in mouse – mostly IAPs and MusD/ETns –, but they are still responsible for up to 10% of spontaneous mutations observed in inbred mice (Dewannieux et al., 2004; Maksakova et al., 2006; Ribet et al., 2004). In humans, ERVs are inactive, with few polymorphisms in the population suggestive of a recent activity of HERVK elements (Marchi et al., 2014).

Non-LTR elements are further divided into the autonomous long interspersed elements (LINE), and the non-autonomous short interspersed elements (SINE)

and, in primates, the SINE-R, VNTR, Alu (SVA) elements. LINEs alone encompass about 20% of mouse and human genomes (Lander et al., 2001; Waterston et al., 2002). Their structure is composed of a 5' promoter driving the expression of the transposition machinery, containing of a reverse transcriptase and an endonuclease, and a terminal 3'UTR with a poly(A) tail (Baudino et al., 2010). LINE-1s are the only known currently active autonomous transposons in humans, with about a hundred copies still capable of retrotransposition, accounting for 0.1% of de novo mutations in humans (Antony et al., 2011; Maksakova et al., 2006). SINEs are non-autonomous elements, meaning they are dependent on LINE-encoded factors in order to replicate. Most SINEs derive from tRNAs, 7SL RNA, or 5s RNA, thus harboring an RNA polymerase III promoter (Padeken et al., 2015). SVAs also rely on LINEs for replication, and are active hominoid-specific TEs consisting of a composite of different retroelements (Hancks and Kazazian, 2010; Raiz et al., 2012).

3. Genomic impact of transposable elements

Vertebrates are extremely diverse and have developed innovative adaptations to adjust to a range of habitats, from the bottom of the sea to high mountains. This adaptability can be greatly attributed to their genomic diversity, which is influenced by different factors, including the high TE content in their genomes (Chalopin et al., 2015; Warren et al., 2015). Due to their mobility and high copy number, TEs provide a great source of genetic diversity and can alter vertebrate genomes in many ways. As mutagenic effects, these impacts that can be either beneficial or detrimental, at the individual or population level.

Since TEs bear promoters, enhancers, open reading frames, splicing sites, and often poly(A) signals, they are true transcriptional landmines. Upon landing on new genomic locations they can impact on genome structure and neighboring genes transcription (**Figure 3**). They can disrupt genes (providing new exons, termination, or splicing sites), alter their transcription (acting as new promoters or enhancers), serve as ground for recombination (deletions, duplications,

rearrangements, and translocations), alter 3D genomic structure (creating new insulator-binding sites or new 3D chromatin interactions), or provide novel open reading frames entirely (reviewed in Friedli and Trono, 2015a; Rebollo et al., 2012c; Warren et al., 2015).
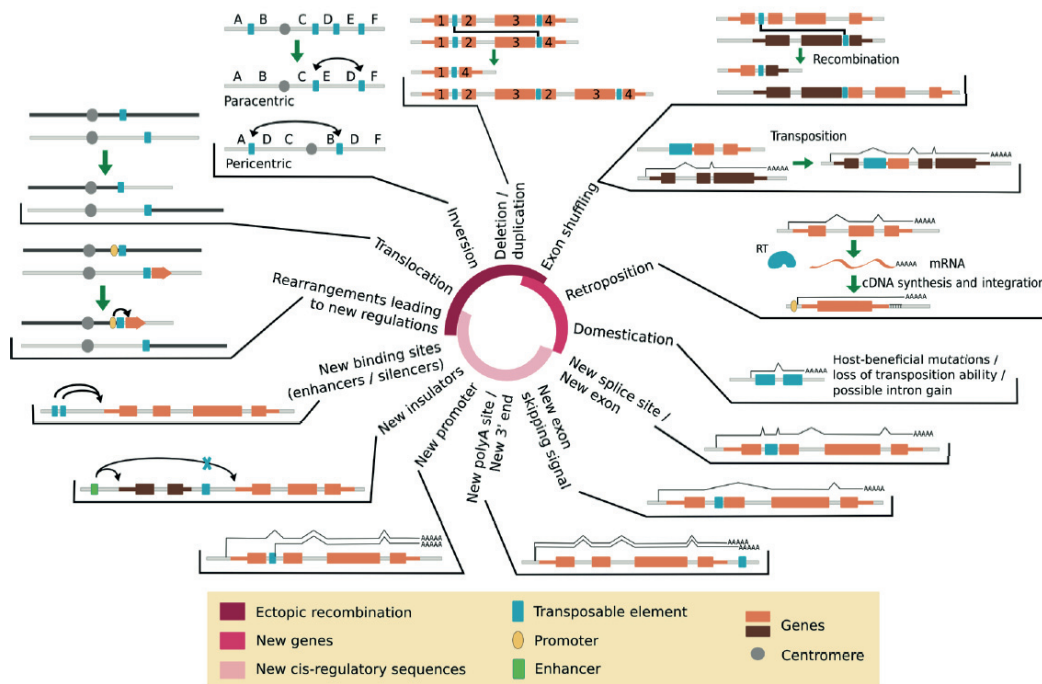
Figure 3. Genomic impact of transposable elements. Summary of different ways in which transposable elements can cause recombination or affect neighboring transcription. TEs are source of genomic alterations and transcriptional perturbations. *Source: (Warren et al., 2015).*

One of the main impacts of TEs on genomes is the generation of new or alternative promoters. LTR transposons are particularly represented in this category as they bear promoters with binding sites for several transcription factors. One of such example is the regulation of the salivary amylase genes (Samuelson et al., 1996; Ting et al., 1992). Mammals generally produce amylase in the pancreas, with the exception of some species that also produce it in the saliva, such as primates. In hominids, the insertion of a HERV-E upstream one of the copies of the amylase gene drives the specific expression of the protein in the saliva. Many other cases of LTR promoter exaptation have been documented, and

15

they normally act as alternative promoters, conferring new or altered tissue-specificity (Cohen et al., 2009; Rebollo et al., 2012a; Stavenhagen and Robins, 1988).

Another important implication of TE activity is the emergence of new enhancers, sometimes leading to emergence of new regulatory networks. Examples in mammals include MER130 elements that act as neocortex-specific enhancers, a SINE integrant that functions as a distal enhancer of the *Fgf8* gene expression in the diencephalon, and RLTR13D5 ERVs that were co-opted as placental-specific enhancers (Chuong et al., 2013; Nakanishi et al., 2012; Notwell et al., 2015). Retroelements are also associated with pluripotency regulatory networks, as many binding sites for pluripotency factors (such as Oct4 and Nanog) reside within mobile DNA elements in the human genome, and LTR elements are implicated in the regulation of specific genes in early embryogenesis (Bourque et al., 2008; Fort et al., 2014; Macfarlan et al., 2012; Peaston et al., 2004). TEs further correlate with p53 binding sites, with more than 1/3 of the binding sites of this protein overlapping with primate-specific ERV elements (Wang et al., 2007).

Interestingly, TEs are not only co-opted by transcriptional networks but also add a layer of species-specificity to it, as these elements are largely species- or lineage-specific. A nice example of such is found in the immune system (Chuong et al., 2016). In humans, MER41 ERV elements function as interferon-inducible enhancers, and removal of some of those TEs by CRISPR/Cas9-mediated depletion impairs expression of neighboring genes involved in immune functions. Additionally, MER41-like elements independently colonized multiple mammalian lineages, including mouse, dog, and cow, where they also act as interferon-induced enhancers.

TE-driven ectopic gene expression can lead to measurable phenotypes in mice, such as the agouti and the axin-fused mice. In the agouti mouse, insertion of an IAP upstream the *agouti* gene drives ectopic expression, leading to yellow fur, obesity, diabetes, and tumors (Morgan et al., 1999). Furthermore, the differential epigenetic control of the IAP results in a variation of phenotypes among

individuals, ranging, for instance, from yellow fur when the IAP is completely silenced to brown when the retroelement is active. In the case of the axin-fused mouse, the insertion of an IAP in the middle of the *Axin* gene results in a characteristic kinked-tail phenotype. *Axin* is an inhibitor of the Wnt signaling pathway that regulates embryonic axis formation. Different insertions of the IAP into the gene lead to exonization of the retroelement, exon skipping, and initiation of transcripts from the 3'LTR of the IAP, disrupting protein function (Vasicek et al., 1997; Zeng et al., 1997).

TEs can also influence gene expression via production of long non-coding RNAs (lncRNAs). Transposons are overrepresented in known lncRNA sequences, with two-thirds of lncRNA containing at least one TE sequence in zebrafish, mouse, or human, and about 10% of human lncRNAs initiating within an ERV LTR (Kapusta et al., 2013). Functional examples can be found in embryonic development – with HERVH/LTR7-derived lncRNAs being required to pluripotency maintenance – and in the betaglobin gene cluster – in which an ERV-9 LTR appears to affect downstream globin genes via lncRNA production (Fort et al., 2014; Ng et al., 2012; Ohnuki et al., 2014; Pi et al., 2010).

ERV-derived proteins themselves can be direct sources of genetic diversity. A fascinating example is provided by mammalian syncytins. Placental development involves the formation of syncytiotrophoblasts via extensive cellular fusion, a process essentially mediated by ERV-derived proteins called syncytins (Dupressoir et al., 2009; Dupressoir et al., 2011; Mi et al., 2000). Interestingly, across mammals, these proteins are different and derive from the *env* gene of distinct groups of ERVs, indicating a convergent evolution with multiple and independent events of ERV coding sequences co-option (Lavialle et al., 2013).

Certainly, not all TE mutagenic effects are beneficial, and diseases associated with TEs include cancers, hemophilia, and muscular dystrophy (reviewed in Ayarpadikannan et al., 2015; Hancks and Kazazian, 2012; Mager and Stoye, 2015). In mouse, similarly to their exogenous counterparts, replication competent ERVs related to MMTV and MLV can cause cancer via activation of protoncogenes (Mager and Stoye, 2015). Most of the TE-associated diseases in

humans are related to non-LTR transposons, as they are the only currently active elements in this species. A known cause of breast cancer in women is the insertion of the primate-specific *Alu* SINEs into the *BRCA1/2* genes (Miki et al., 1996; Puget et al., 1999), and cases of hemophilia A and B are associated to insertional mutations of LINE-1 or Alu elements into genes coding for proteins important for blood clothing (Kazazian et al., 1988; Li et al., 2001). Nonetheless, some tumors upregulate ERV transcripts, and there are reports of LTRs driving oncogene expression in human lymphomas (Lamprecht et al., 2010; Romanish et al., 2010). Furthermore, expression of ERV proteins can be detrimental to the host and is associated with autoimmune diseases, such as systemic lupus erythematosus in mice and multiple sclerosis in humans (Antony et al., 2011; Baudino et al., 2010).

4. Epigenetic regulation of transposable elements

While there are many beneficial events of TE activity, there can be severe detrimental consequences of their unrestrained spread. Hence, the host has developed several mechanisms to control TE expression and transposition (reviewed in Friedli and Trono, 2015b; Mager and Stoye, 2015; Rowe and Trono, 2011). Silencing happens mostly in early embryogenesis so as to keep TEs in a silent state in the adult organism. TE control in pre-implantation development and in germ cells is of particular importance, as global erasure of DNA methylation marks leads to a wave of reactivation of transposons.

DNA methylation

The best-studied mechanism of control of TEs is DNA methylation. As explained in section 1 of this introduction, methylation of cytosines at CpG sites is a repressive mark, and it is established by the *de novo* methyltransferases DNMT3a and b, and maintained by DNMT1. After the wave of epigenetic

reprogramming in pre-implantation embryos and in primordial germ cells, DNMT3a and b reestablish DNA methylation in the genome, including at most transposon sequences. As expected, knockout (KO) animals for any of the *de novo* DNMTs are not viable or die at 4 weeks of age, with more severe phenotypes in the double KO (Okano et al., 1999). Also, a small decrease in methylation of IAP elements was observed in these mice. Detailed analysis in germ cells revealed differential recruitment of DNMTs to different types of repeats, with SINEs being methylated by DNMT3a, IAPs and LINEs by both DNMT3a and b, and satellite repeats by DNMT3b (Kato et al., 2007).

Maintenance DNA methylation is also crucial for TE silencing during development. DNMT1 KO is lethal at E8.5, and KO animals display strong hypomethylation of MLV and IAP ERVs, with 100-fold increased expression of IAP transcripts (Li et al., 1992; Walsh et al., 1998). The specific upregulation of IAPs may be due to the fact that, differently from other TEs, IAPs are one of the few sequences that resist global DNA methylation in pre-implantation development, increasing the importance of maintenance DNA methylation in their transcriptional control. The importance of DNMTs in TE regulation is more evident in DNMT triple KO cells. In ES cells lacking the three DNMTs, upregulation of IAPs and LINEs was observed, consistently with their hypomethylated status in these cells (Tsumura et al., 2006).

DNA methylation is particularly important in differentiated cells, while in pluripotent cells other mechanisms, such as histone modifications, also take place (Leung and Lorincz, 2012). Silencing of IAPs is maintained upon *Dnmt1* KO ES cells, but not when they are differentiated (Hutnick et al., 2010). Depletion of Dnmt1 or treatment with the nucleoside analog 5-azacytidine (which inhibits DNMTs activity) induces IAP expression in post-implantation embryos, MEFs, and other somatic cells (Hutnick et al., 2010; Rowe et al., 2013a; Walsh et al., 1998). Of note, most of the accumulated data on DNA methylation role in silencing retrotransposons in somatic cells is on IAPs, and DNA demethylation appears to affect TE classes differently, with LINE-1s loosing more 5mC than IAP, for example (Bulut-Karslioglu et al., 2014; Wissing et al., 2012).

Histone modifications

DNA methylation is not the only epigenetic mark important for TE control, histone modifications also play a major role. This is evident in the case of silencing of the exogenous retrovirus MLV. Upon transduction, the provirus is repressed within 2 days, while DNA methylation at the LTR is not detected until 8-14 days, suggesting that other mechanisms mediate the initial silencing of retroviruses (Gautsch and Wilson, 1983; Kempler et al., 1993; Niwa et al., 1983). Indeed, MLV and several TEs are also silenced by histone modifications, particularly H3K9 methylation.

H3K9 methylation is found on several TEs in ES cells. It can be deposited by histone methyltransferases (HMT), comprising SETDB1/ESET, G9a, SUV39H1, SUV39H2, and EuHMTase/GLP, and expression of different TE classes is regulated by different histone modifying enzymes (Kouzarides, 2007; Mikkelsen et al., 2007). Among these ones, SETDB1 is one of the most important for TE control, and catalyzes the addition of one, two, or three methyl groups at H3K9. SETDB1 KO is lethal at E3.5-E5.5, an embryonic period at which most TEs become inactivated. ES cells depleted for this HMT upregulate several TEs (e.g. IAPs and MusD elements), which in turn lose SETDB1 binding and corresponding deposition of H3K9me3 marks (Dodge et al., 2004; Jahner et al., 1982; Matsui et al., 2010). Interestingly, SETDB1 depleted ES cells deregulate a much higher number of TEs than DNMT triple KO cells, indicating that in pluripotent cells H3K9me3 is more important for TE transcriptional control than DNA methylation (Karimi et al., 2011; Matsui et al., 2010). Still, simultaneous depletion of SETDB1 and DNMT1 leads to synergistic effects on the reactivation of some ERVS, including IAPs, confirming that both DNA and histone methylation play a role in TE silencing.

G9a catalyzes H3K9me1 and H3K9me2 and can form a complex with GLP. Depletion of either one of these HMTs is embryonically lethal and they are of particular importance for MERVL silencing (Tachibana et al., 2002; Tachibana et al., 2005). MERVL are bound by G9a, and ES cells depleted for this HMT or its activator GLP display higher expression of these elements (Maksakova et al.,

2013). Conversely, while G9a is required for H3K9me2 and DNA methylation of IAPs and MusDs, no increase of transcription of these ERVs has been observed in murine ES cells depleted for this HMT (Dong et al., 2008). Interestingly, in the case of MLV exogenous retroviruses, G9a is required for *de novo* methylation and establishment of silencing of the provirus, but not for the maintenance of this repression (Leung et al., 2011).

The other HMTs known to be implicated in TE silencing are Suv39h1 and Suv39h2. They deposit H3K9me3 at major satellite repeats, and the double KO of these proteins is lethal (Peters et al., 2001). In murine ES cells, Suv39h double KO leads to the decrease of H3K9me3 in LINEs and ERVs, but to the upregulation only of LINEs, suggesting that they are the main HMTs responsible for the regulation of these non-LTR elements, while SETDB1 is the main responsible for controlling ERVs (Bulut-Karslioglu et al., 2014). Interestingly, this effect was observed mainly for full length LINEs.

RNA-based mechanisms

Although histone and DNA methylation are well known mechanisms implicated in the silencing of TEs, there are still many questions regarding how these marks specifically target certain TE sequences. One of such mechanisms involves small RNAs and RNA interference. In mammals, this mechanism likely helps establishing the silencing of some TEs in early embryogenesis, a period when these elements are transcriptionally activated and then re-silenced.

RNA interference (RNAi) is a general term to define related pathways comprising small RNAs (20-30 nucleotides long) that guide sequence-specific gene regulation. Classes of small RNAs include small interfering RNAs (siRNAs), micro RNA (miRNA), and piwi-interacting RNA (piRNAs), which are classified according to their length, proteins involved in their biogenesis, or mode of regulation (Ghildiyal and Zamore, 2009). RNAi pathways involve proteins such as Dicer and Argonaute, that process different types of small RNAs and

ultimately lead to gene silencing through target cleavage, inhibition of protein synthesis, or directed chromatin modifications (Malone and Hannon, 2009).

Interestingly, many TEs produce small RNAs, and recent studies show that transposons can self-inflict RNA-mediated regulation. In mice *Dicer*-depleted early embryos, IAP expression is increased; and in ES cells KO for *Dicer*, not only IAPs, but also LINE-1s are overexpressed (Kanellopoulou et al., 2005; Svoboda et al., 2004). LINE-1 retrotransposition itself is blocked by Dicer-processed endogenous LINE-1-derived siRNAs, originated from bidirectional transcription (Yang and Kazazian, 2006). Moreover, Microprocessor, a nuclear complex involved in miRNA biogenesis, binds RNAs derived from LINE-1s, SINEs and SVAs (Heras et al., 2013).

piRNAs are also largely implicated in transposon control and DNA methylation in pluripotent and germ cells. The Piwi Argonaute proteins Miwi and Mili2 are involved in *de novo* methylation of TEs, and the depletion of these proteins in the male germ line leads to sterility and LINE-1 activation (Aravin et al., 2008; Carmell et al., 2007; Kuramochi-Miyagawa et al., 2008). Furthermore, recent evidence demonstrates that Piwi-2 partakes in the regulation of young LINE-1 elements in human and ape pluripotent cells (Marchetto et al., 2013). Although it remains unclear how exactly small RNAs trigger chromatin changes and DNA methylation, recent evidence showed that piRNAs drive accumulation of H3K9me3 at LINEs in germ cells, and can also recruit DNMTs at TEs (Aravin and Bourc'his, 2008; Pezic et al., 2014).

KRAB/KAP1 complex

The KRAB/KAP1 repressor complex also directs histone and DNA methylation to TEs. KAP1 is a corepressor that is recruited to a specific DNA locus via Krüppel-associated box domain zinc finger proteins (KRAB-ZFPs). KAP1 then acts as a scaffold for a repressive complex encompassing the HMT SETDB1, the nucleosome remodeling and deacetylation (NuRD) complex, and

heterochromatin protein 1 (HP1) (Nielsen et al., 1999; Schultz et al., 2002; Schultz et al., 2001; Sripathy et al., 2006). Altogether these proteins promote chromatin modifications with deposition of H3K9me3 and DNA methylation, leading to gene silencing (**Figure 4**).
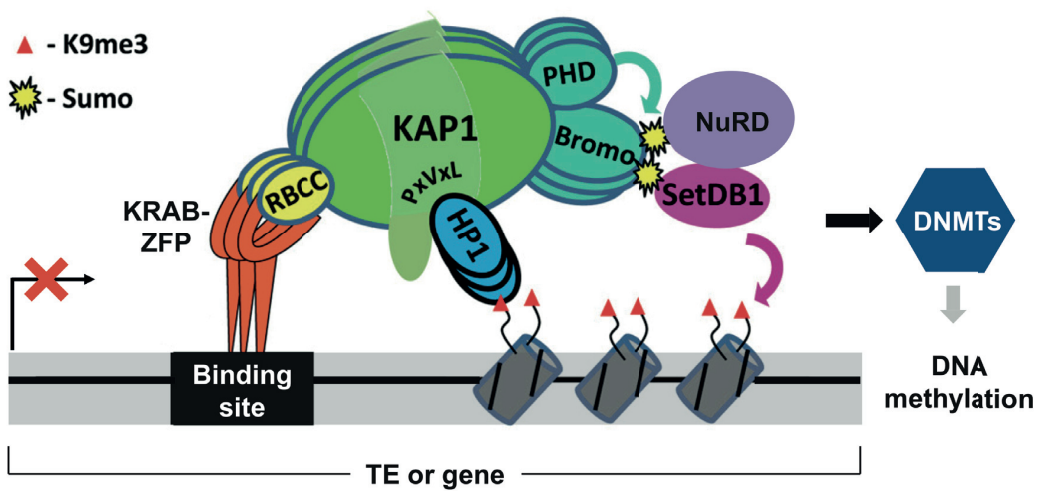
Figure 4. Schematic representation of the KRAB/KAP1 complex. KAP1 is depicted with its different domains and the proteins of the repressive complex are represented. *Adapted from (Iyengar and Farnham, 2011).*

The importance of KRAB/KAP1 regulation is evident by the early embryonic lethality of *Kap1* KO mouse embryos (Cammas et al., 2000). KAP1 has been implicated in many aspects of physiology, such as embryonic development, stem cell pluripotency, genomic imprinting, and differentiation. It has been demonstrated that KAP1 is required for the maintenance of self-renewal in ES cells and for oocyte-to-embryo transition (Hu et al., 2009; Messerschmidt et al., 2012). Together with the KRAB-ZFP ZFP57, KAP1 is implicated in the maintenance of imprinting marks during early embryogenesis through binding to imprinting control regions (ICR) (Quenneville et al., 2011; Zuo et al., 2012). KAP1 further regulates gene networks important for B cells and skeletal muscle

23

differentiation, and it is important for liver homeostasis (Bojkowska et al., 2012; Santoni de Sio et al., 2012; Singh et al., 2015). This protein has also been associated with regulation of apoptosis and DNA damage response (Iyengar and Farnham, 2011). Most importantly, KAP1 is involved in the silencing TEs in embryonic development. The corepressor binds to TEs and controls them in ES cells, with ES cells depleted for KAP1 displaying strong upregulation of different classes of repeats, both in mouse and human (Rowe et al., 2010; Turelli et al., 2014). KAP1 regulates TEs together with its partner SETDB1 and KAP1 or SETDB1 depletion leads to decreasing levels of H3K9me3 at repeats (Matsui et al., 2010).

KRAB-ZFPs are the largest family of transcriptional regulators encoded by higher vertebrates. While poly-zinc finger proteins are widely present in eukaryotes, KRAB-ZFPs are specific to tetrapods (Emerson and Thomas, 2009). They represent about one quarter of all human transcription factors with 300-400 KRAB-ZFP genes annotated in both mouse and human (Urrutia, 2003; Vaquerizas et al., 2009). KRAB-ZFPs are characterized by the presence of C-terminal C2H2 DNA-binding zinc fingers and an N-terminal Krüppel-associated box (KRAB) domain, separated by a linker region (**Figure 5**). The C2H2 zinc fingers confer DNA-binding specificity to these transcription factors, and consist in tandem repeats of the pattern $CX_{2-4}CX_{12}HX_{2-6}H$ (where X is any amino acid) (Emerson and Thomas, 2009; Iuchi, 2001). The KRAB domain is associated with the transcriptional activity of KRAB-ZFPs. It consists of approximately 75-aminoacids and mediates the interaction with their universal co-repressor KAP1 (Friedman et al., 1996). Other domains, such as the SCAN can also be found in some members of the family (Urrutia, 2003).
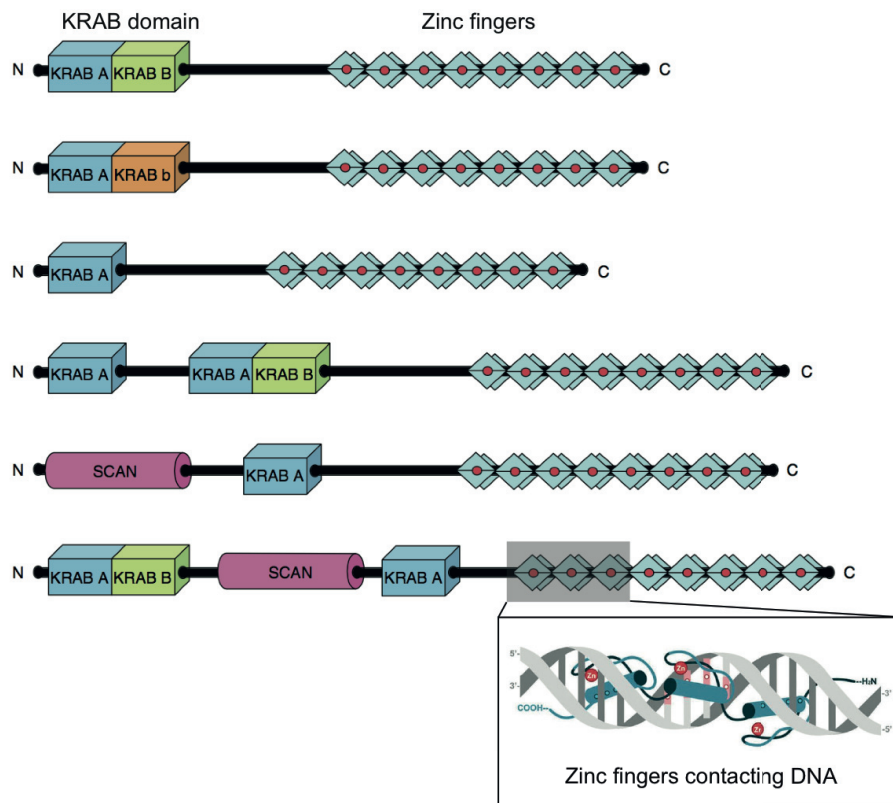
Figure 5. General structure of KRAB-ZFPs. KRAB-ZFPs are composed of a KRAB domain and an array of zinc fingers. The KRAB domain can comprise KRAB A or KRAB B variants, and the number of zinc fingers varies among different members of the family. KRAB-ZFPs can also have other domains such as the SCAN. The lower panel shows a scheme of zinc fingers contacting DNA. N, amino term; C, carboxyl terminus. *Adapted from (Urrutia, 2003) and (Stubbs et al., 2011).*

Interestingly, over time, the expansion of *Krab-zfp* genes mirrored waves of retroviral invasion into the genomes of tetrapods, reinforcing their role in repressing TEs together with KAP1 (Thomas and Schneider, 2011). During this process, KRAB-ZFPs underwent strong positive selection at positions encoding for amino acids predicted to determine DNA binding specificity, consistent with a role in countering rapidly mutating genetic invaders (Emerson and Thomas, 2009). Furthermore the recent study of ZNF93 and its TE targets across primates suggests there is a parallel evolution of restriction factors and TE mutants escaping their inhibition (Jacobs et al., 2014). This argues for a host-TE arms race hypothesis, where KRAB-ZFPs were primarily selected to silence TEs. In this hypothesis, repressors with new DNA binding specificity are selected in

response to new exogenous retroviruses or mutating endogenous transposons (**Figure 6**).
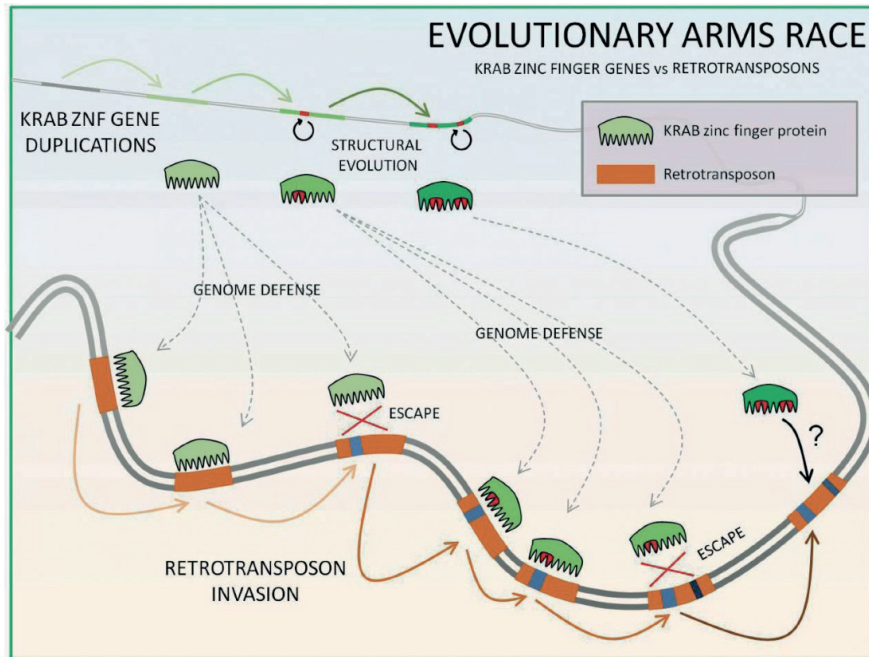
Figure 6. Evolutionary arms race between KRAB-ZFPs and TEs. KRAB-ZFP genes co-evolve to counteract transposons insertions. While KRAB-ZFP genes undergo duplication and structural modifications, TEs transpose and accumulate mutations in a constant evolutionary arms race. *Source: (http://www.frankjacobslab.com/Evolutionary-impact-of-Retrotransposons/).*

Although recent genome-wide binding studies indicate that many KRAB-ZFPs have EREs as their preferential genomic targets, detailed functional data are missing for most members of the family (Najafabadi et al., 2015). The best-studied case is ZFP809, a murine specific KRAB-ZFP responsible for the silencing of exogenous MLV and its endogenous counterparts (Wolf and Goff, 2007, 2009 ; Wolf et al., 2015). ZFP809 binds exactly to the so-called primer binding site (PBS) sequence of the viral genome, recruits KAP1 and the repressive complex, leading to deposition of H3K9me3, DNA methylation, and repression of its exogenous and endogenous retrovirus targets. Curiously, depletion of ZFP809 in ES cells does not lead to increased expression of endogenous MLVs, contrary to what is observed in adult tissues of mice depleted for this protein (Wolf et al.,

2015). Also, depletion of the pluripotency-related ZFP819 in murine ES cells results in the derepression of IAPs and LINEs, although the role of this ZFP in regulating TEs is still not fully understood (Tan et al., 2013). Other recent examples of KRAB-ZFPs controlling TEs in ES cells are ZNF91 and ZNF93, which repress SVAs and LINE-1, respectively, in humans (Jacobs et al., 2014).

Recent studies on KAP1 and the few well-characterized ZFPs revealed that the KRAB/KAP1 complex not only regulates TEs in ES cells, but can also impact neighboring gene transcription in this setting (Jacobs et al., 2014; Rowe et al., 2013b; Wolf et al., 2015). Upon KAP1 depletion in ES cells, repressive chromatin marks at TEs are replaced by active chromatin modifications typical of enhancers, with transcriptional impact to neighboring cellular genes. These findings suggest that the KRAB/KAP1 complex is important both in TE silencing and in safeguarding neighboring genes from their transcriptional potential.

While KRAB/KAP1 role in silencing TEs in ES cells is well established, its role in somatic cells is not well defined. In ES cells, the KRAB-ZFP-mediated docking of KAP1 triggers the deposition of H3K9me3 and DNA methylation. Once established, DNA methylation is perpetuated across cell divisions, and its establishment during early embryogenesis is thought to result in permanent silencing, without need for persistent expression of the cognate KRAB-ZFP repressors (Quenneville et al., 2012; Wolf et al., 2015). Cumulated evidences indicate that DNA methylation is an important mechanism of TE control in somatic tissues (Hutnick et al., 2010; Jackson-Grusby et al., 2001; Rowe et al., 2013a; Walsh et al., 1998). However, we previously observed that a significant fraction of TEs bound by KAP1 in human ES cells still carries the corepressor in mature T lymphocytes (Turelli et al., 2014), and that KAP1 deletion in neuronal progenitors activates some EREs (Fasching et al., 2015). A few ERVs are similarly induced in murine B lymphocytes and MEFs deleted for SETDB1 (Collins et al., 2015; Wolf et al., 2015). Moreover, many KRAB-ZFPs are expressed not only in ES cells but also in a variety of tissues (Barde et al., 2013; Corsinotti et al., 2013; Lizio et al., 2015). Altogether, these recent data would suggest a possible role for the KRAB/KAP1 complex in silencing TEs also in differentiated cells.

Aims of the thesis

TEs have dual implications in host biology and can act both as genomic threats and sources of genetic variability. Although these elements represent almost half of our genome and many advances have been made in recent years, we do not fully understand how they impact vertebrate biology and how they are controlled. KAP1 and the associated repressor complex are important for TE regulation in stem cells. In the last couple of years, the large family of transcriptional repressors KRAB-ZFPs has been shown to mediate this process in stem cells, but functional information on most of the members of the family is still missing. While the role of the KRAB/KAP1 complex in ES cells is better defined, it is believed to instate permanent silencing and to be dispensable in adult tissues. However, recent evidences – such as expression of KRAB-ZFPs in different adult tissues and SETDB1-mediated control of retroelements in somatic cells – suggest a possible role for the KRAB/KAP1-mediated regulation of TEs in adult cells.

In view of the physiological importance of this large family of transcription factors and the present lack of knowledge around it, the aim of this thesis was to explore the mechanisms of KRAB/KAP1-mediated regulation of TEs in pluripotent and differentiated cells. Our goal was to identify novel KRAB-ZFPs controlling transposons and to use them as tools to unravel the general mechanisms and implications of KRAB/KAP1 regulation. For that purpose, we first set up a large-scale screen system to identify murine KRAB-ZFPs interacting with specific DNA targets, with emphasis on TE-derived sequences. We then characterized the physiological roles of the candidates obtained in the screen, and defined the molecular mechanisms triggered by the genomic recruitment of these candidates on the local chromatin, both in pluripotent and differentiated cells, and *in vivo*. Understanding how TEs are regulated is important not only to decipher vertebrate evolution and development, but also to shed light on the mechanisms of gene regulation and diseases associated with these elements.

II. Results – Part I: A large-scale functional screen to identify epigenetic repressors of retrotransposon expression

Manuscript title

A large-scale functional screen to identify epigenetic repressors of retrotransposon expression

Authors

Gabriela Ecco, Helen M. Rowe, and Didier Trono

Summary of results and contribution

In the first part of this thesis we developed a method to identify KRAB-ZFPs interacting with specific DNA targets, with emphasis on TE sequences. Our screen is based on the repression of a PGK-GFP cassette, reminiscent of a mammalian one-hybrid system. In sum, a TE-derived DNA sequence of interest is cloned upstream of a PGK-GFP cassette and screened against 211 murine KRAB-ZFPs, initially by plate reader and finally by FACS readout. This screening, which led us to the identification of TE/KRAB-ZFPs pairs, has proven to be useful for

various projects in the laboratory. In this section the methodology of the established screen is described in detail.

My contribution to this manuscript comprised its development from the elaboration of the idea, to the execution, and the writing of the final text, with supervision of Dr. Helen Rowe and Prof. Didier Trono.

A Large-Scale Functional Screen to Identify Epigenetic Repressors of Retrotransposon Expression

Gabriela Ecco, Helen M. Rowe, and Didier Trono

Abstract

Deposition of epigenetic marks is an important layer of the transcriptional control of retrotransposons, especially during early embryogenesis. Krüppel-associated box domain zinc finger proteins (KRAB-ZFPs) are one of the largest families of transcription factors, and collectively partake in this process by tethering to thousands of retroelement-containing genomic loci their cofactor KAP1, which acts as a scaffold for a heterochromatin-inducing machinery. However, while the sequence-specific DNA binding potential of the poly-zinc finger-containing KRAB-ZFPs is recognized, very few members of the family have been assigned specific targets. In this chapter, we describe a large-scale functional screen to identify the retroelements bound by individual murine KRAB-ZFPs. Our method is based on the automated transfection of a library of mouse KRAB-ZFP-containing vectors into 293T cells modified to express GFP from a PGK promoter harboring in its immediate vicinity a KAP1-recruiting retroelement-derived sequence. Analysis is then performed by plate reader and flow cytometry fluorescence readout. Such large-scale DNA-centered functional approach can not only help to identify the trans-acting factors responsible for silencing retrotransposons, but also serve as a model for dissecting the transcriptional networks influenced by retroelement-derived *cis*-acting sequences.

Key words Retroelement control, KRAB-ZFP, Protein-DNA interaction, Epigenetic regulation, Functional screen, Mammalian one-hybrid assay, GFP repression assay

1 Introduction

The sequence-specific deposition of chromatin marks has a profound influence on transcription. DNA methylation and histone modifications shape chromatin structure, guide the genomic recruitment of transcription factors, and dictate expression patterns. Retroelement-derived sequences account for more than half of the human and mouse genome, and epigenetic modifications play a paramount role in their transcriptional control. DNA methylation is a well-established mechanism of retroelement silencing, particularly in differentiated cells. Furthermore, during early embryonic development—when there is a wave of erasure of DNA methylation—covalent histone modifications, such as trimethylation

of lysine 9 on histone H3 (H3K9me3), take over and keep retroelements under control (reviewed in [1, 2]).

Some important mediators in the epigenetic silencing of retroelements are Krüppel-associated box domain zinc finger proteins (KRAB-ZFPs). These represent one of the largest families of transcriptional regulators with 300–400 members in mice and humans [3, 4]. KRAB-ZFPs bind to the DNA and recruit the universal corepressor KAP1 (also known as Trim28, TIF1β, and KRIP-1) [5], which then acts as a scaffold for chromatin-modifying proteins, such as histone methyltransferases and histone deacetylases that alter the chromatin structure and silence the genomic region [6–9]. KAP1 and other proteins of the KAP1-nucleated silencing complex repress a large array of retroelements in mouse and human stem cells, including ERVs, LINEs, and SVAs [10–14]. However, while a few matches between a given KRAB-ZFP and its target retrotransposon have been established, the sequence-specific mediator of KAP1 recruitment to most retroelements remains unidentified [14, 15].

This gap in knowledge largely stems from the difficulty in identifying the proteins binding to specific DNA targets, even more so when these reside in repetitive genomic sequences. Methods to determine protein-DNA complexes from the protein perspective (protein-centered methods, such as ChIP-on-chip, ChIP-seq, SELEX, and MITOMI) have greatly improved in the last years, thanks to the advances in high-throughput DNA sequencing technologies. Techniques taking DNA as the starting point, in contrast, have evolved at a lower pace and rarely allow a large-scale approach.

The DNA-centric techniques most frequently used to characterize DNA-protein complexes are electrophoretic mobility shift assay (EMSA), yeast one-hybrid assay, and in vitro methods of affinity purification with DNA as bait usually followed by mass spectrometry (reviewed in [16, 17]). These approaches have the disadvantage of studying the DNA-protein interactions in an artificial environment, without a functional readout. In the case of the yeast one-hybrid assay, the interaction occurs in a eukaryotic cell environment, closer to mammalian systems, but it still lacks a functional component.

In order to identify KRAB-ZFPs responsible for the sequence-specific recognition of retrotransposon-derived elements, we have developed a transcriptional repression-based assay reminiscent of a mammalian one-hybrid system. A scheme of the screen is depicted in Fig. 1. A library of KRAB-ZFPs, for instance all mouse family members, is built into a gateway-compatible vector. In parallel, a DNA target of interest, such as a retrotransposon-derived KAP1-recruiting repressive sequence, is cloned upstream of a PGK-GFP reporter and the resulting vector is used to establish a stable cell line. The cell line is then transfected in a large-scale fashion together with the KRAB-ZFP library. The readout is done by the assessment of fluorescence
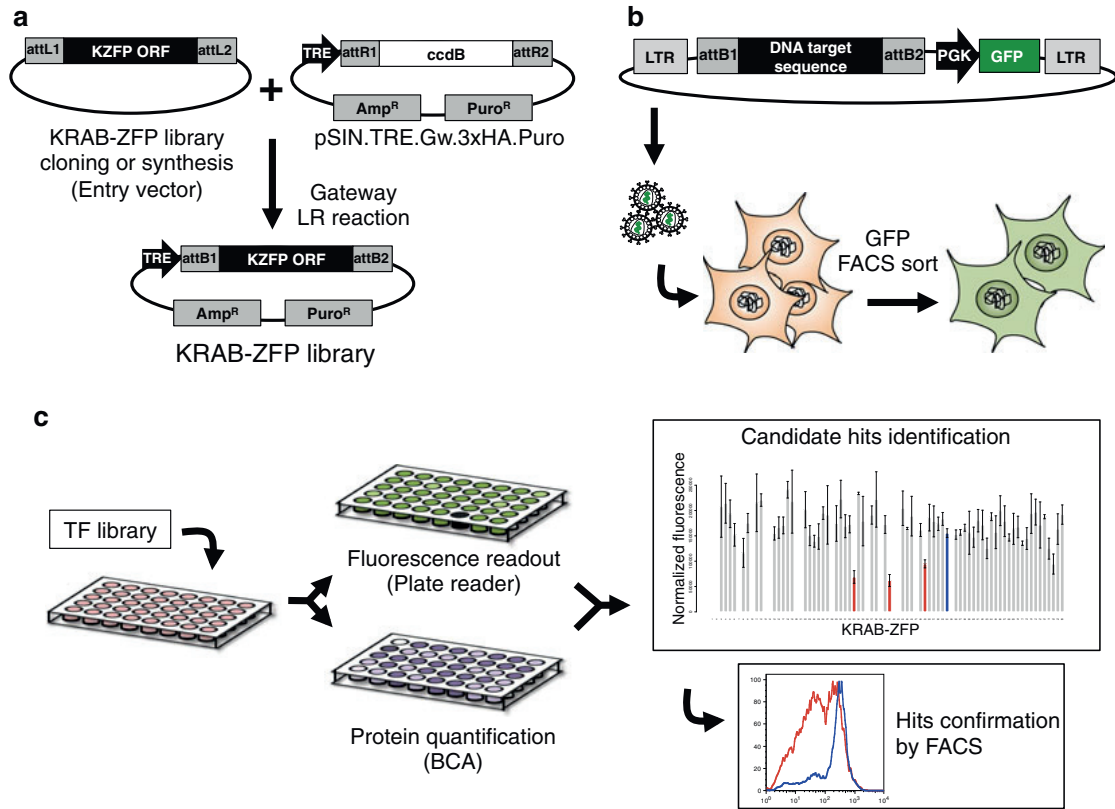
Fig. 1 Screen overview. (**a**) Transcription factor library construction. A library of KRAB-ZFPs is synthesized or PCR-amplified into gateway Entry vectors and used in LR reactions with the pSIN.TRE.Gw.3xHA.Puro vector to generate a final KRAB-ZFP library. (**b**) DNA target sequence cloning and establishment of stable cell lines. The DNA target sequence of interest, in this case a retrotransposon sequence, is cloned into the pRRL.R1R2.PGK. GFP reporter vector using pENTR/D-TOPO cloning. The plasmids are used for lentiviral vector production, which is used to transduce 293T cell. The stable cell line is established after sorting the GFP-positive cells. (**c**) Screening assay. The established cell lines are transfected against each KRAB-ZFP in triplicates in 96-well plates. After 6 days of induction of expression and selection, GFP fluorescence is assessed with a plate reader and protein amount is quantified with BCA. The normalized fluorescence is calculated and candidate hits are identified. As an example, in the graph three candidate hits and the control transfection are highlighted. The candidate hits are tested by flow cytometry (FACS) and the hits are identified when there is specific repression of the DNA target sequence

intensity by plate reader, coupled with protein quantification, followed by the confirmation of the hits by flow cytometry.

We found this screening method, which uses the functional interaction of KRAB-ZFPs with the repressor complex as readout, as capable of identifying a high percentage of DNA target-KRAB-ZFP pairs, with negligible rates of false positivity. Furthermore, such DNA-centered approach can be used to understand other aspects of the regulatory roles of retroelements. Finally, it can be adjusted to different species and to other types of repressors or other modulators of gene expression, and hence help unravel the roles of many *cis*-acting regulatory sequences and transcription

factors. In this chapter, we describe in detail each step of this large-scale screen, here aimed at identifying KRAB-ZFPs binding to specific DNA sequences.

2   Materials

2.1   Transcription Factor Library Construction

1. Library of transcription factors in an Entry plasmid (from synthesis or PCR amplification).
2. Gateway LR Clonase II Enzyme mix (Life Technologies).
3. pSIN.TRE.Gw.3xHA.Puro destination vector, 100 ng/μL.
4. *E. coli* HB101 competent cells.
5. LB medium and LB agar plates.
6. Ampicillin sodium salt.
7. GoTaq Green Master Mix 2× (Promega).
8. CMV1F primer: 5′-GGAGGCCTATATAAGCAGAGCTCGT-3′.
9. PGK3R primer: 5′-GCTGCCTTGGAAAAGGCGCAACC-3′.
10. Agarose, LE, analytical grade.
11. Glycerol 60 % (v/v) in ultrapure water: Autoclave it before use, and store it at room temperature.
12. Nucleospin 96 Plasmid Kit (Macherey-Nagel).
13. Nanodrop 8000 Spectrophotometer.
14. 50 μg/mL Proteinase K solution.

2.2   DNA Target Sequence Cloning

1. dNTP set, 100 mM.
2. Pfu DNA polymerase.
3. pENTR/D-TOPO Cloning Kit (Life Technologies).
4. One Shot TOP10 competent bacteria.
5. Kanamycin sulfate.
6. LB medium and LB agar plates.
7. Restriction enzymes.
8. QIAprep Spin Miniprep Kit or similar.
9. M13F primer: 5′-TGTAAAACGACGGCCAG-3′.
10. M13R primer: 5′-CAGGAAACAGCTATGAC-3′.
11. Gateway LR Clonase II Enzyme mix (Life Technologies).
12. pRRL.R1R2.PGK.GFP vector, 100 ng/μL.

2.3   Establishment of Stable Cell Lines Containing the DNA Target Sequence

1. Lentiviral vectors containing the sequence of interest (for lentiviral vector production protocol, *see* [18]).
2. HEK 293T cells.
3. 10 cm round tissue culture dishes.

4. Dulbecco's modified Eagle medium, DMEM.

5. Fetal calf serum (FCS).

6. Penicillin-streptomycin (PS) solution, 100×.

7. Dulbecco's PBS.

8. 0.05 % Trypsin-EDTA.

9. Flow cytometry cell sorter.

2.4 Screening Assay

1. DMEM.

2. Fetal calf serum (FCS).

3. Penicillin-streptomycin (PS) solution, 100×.

4. Dulbecco's PBS.

5. 0.05 % Trypsin-EDTA.

6. 96-Well tissue culture plate, flat bottom.

7. Fugene 6 Transfection Reagent (Promega).

8. Opti-MEM Reduced Serum Media (Life Technologies).

9. 96-Well conical bottom plates.

10. Sciclone ALH 3000 (Caliper Life Sciences).

11. Multidrop Combi dispenser and standard tube dispensing cassette.

12. Doxycycline Hyclate (Sigma-Aldrich).

13. Protease Inhibitor Cocktail Tablets, cOmplete, Mini, EDTA-free (Roche).

14. Puromycin dihydrochloride, 10 mg/mL.

15. RIPA buffer: 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 % (v/v) NP40, 0.5 % (w/v) sodium deoxycholate, 0.1 % (w/v) SDS. Add protease inhibitors freshly before use.

16. Black 96-well polystyrene plate, non-treated, flat bottom.

17. Infinite F500 Plate Reader (Tecan).

18. BCA Protein Assay Reagent (Thermoscientific).

19. Flow cytometer.

3 Methods

3.1 Transcription Factor Library Construction

1. For the construction of the transcription factor (TF) library, a collection of KRAB-ZFPs is obtained by PCR amplification or cDNA synthesis and is then placed in the destination gateway-compatible vector (pSIN.TRE.Gw.3xHA.Puro). For the purposes of this protocol, cDNA synthesis should be done directly in a gateway-compatible Entry vector (L1-L2). When performing the synthesis, suppress the stop codon from the

cDNA, in order to obtain HA-fused proteins in the final vector. For PCR amplification of the open reading frame (ORF) of the TF of interest up to an Entry clone, please refer to [19] (*see* **Note 1**).

2. Once you have the purified Entry plasmids, proceed to the LR reaction. Mix together 100 ng of the Entry vector and 100 ng of the pTRE destination vector, and complete to a final volume of 4 µL with TE buffer. Add 1 µL of LR clonase and incubate from 1 h to overnight (ON) (for large fragments or when performing several reactions at once, we recommend the latter). Add 1 µL of proteinase K, incubate for 10 min at 37 °C, and transform everything into 50 µL of competent bacteria (we typically use HB101). Plate the transformed bacteria into LB agar plates containing 100 µg/mL ampicillin and incubate the plates overnight at 37 °C (*see* **Note 2**).

3. Check the cloned fragments by colony PCR. Pick 3–5 colonies from each plate. Prepare a PCR reaction containing 10 µL of GoTaq Green Master Mix 2× (Promega), 1 µL of 10 µM CMV1F primer, and 1 µL of 10 µM PGK3R primer, and complete to 20 µL final volume with nuclease-free water. Pick a colony by touching it with a sterile tip, then touch a new LB agar plate containing 100 µg/mL ampicillin, and finally transfer the tip to the tube containing the PCR mixture. Mix well and remove the tip after 5 min. Do the PCR reaction in a thermal cycler with a heated lid using the following program:

1	95 °C	2 min
2	95 °C	1 min
3	55 °C	1 min
4	72 °C	4 min
5	72 °C	2 min
6	4 °C	hold

* cycle between steps 2–4 and perform from 25 to 30 cycles.

Evaluate the PCR products on an agarose gel and select the colonies with the correct molecular weight.

4. Culture a positive bacterial colony per TF in 2 mL of LB media supplemented with 50 µg/mL ampicillin at 37 °C for 16–18 h (*see* **Note 3**). Use this culture to perform a glycerol stock for long-term storage and the remaining for plasmid DNA purification. For the glycerol stock, mix together 100 µL of bacterial culture and 100 µL of sterile glycerol 60 %. Store the glycerol stock at –80 °C (*see* **Note 4**). Spin the remaining culture at $3600 \times g$ for 15 min and use it to perform DNA extraction with the Nucleospin 96 Plasmid Kit (Macherey-Nagel). Quantify

the purified DNA (we typically use 8-channel Nanodrop) and dilute them to 20 ng/µL. Arrange the DNA stock in 96-well plates and store them at –20 °C until the day of the screen.

3.2 DNA Target Sequence Cloning

1. For the DNA sequence cloning, the retrotransposon is cloned using the pENTR/D-TOPO Cloning Kit (Life Technologies) and it is further transferred to the gateway-compatible destination vector (pRRL.R1R2.PGK.GFP) by an LR reaction. Design specific primers for your sequence of interest (we had good results with Primer 3 software in the past [20]) and add the sequence CACC on the 5′ end of the forward primers (*see* **Note 5**).

2. Use these primers to set up a PCR reaction containing 1 µL of dNTP mix (10 mM of each dNTP), 1 µL of each primer (10 µM), 3 µL of 10× polymerase buffer, 0.5 µL Pfu DNA polymerase, and the source of cDNA (typically we use 50 ng of genomic DNA), and complete to a final volume of 30 µL with nuclease-free water. Do the PCR reaction(s) in a thermal cycler with a heated lid using the following program:

1	95 °C	2 min
2	95 °C	1 min
3	Tm – 5 °C	1 min
4	72 °C	2 min/kb of ORF
5	72 °C	2 min
6	4 °C	hold

* cycle between steps 2–4 and perform from 25 to 30 cycles.

Evaluate the PCR products on an agarose gel and proceed to the TOPO reaction (*see* **Note 6**).

3. Set up a TOPO reaction containing 2 µL of the PCR product, 0.5 µL of salt solution, and 0.5 µL of pENTR/D-TOPO vector. Incubate for 5 min at RT and transform the entire reaction into 25 µL of One Shot TOP10 competent bacteria. Plate the transformed bacteria onto LB agar plate containing 50 µg/mL of kanamycin and incubate the plate ON at 37 °C.

4. Check the size and identity of the plasmids by restriction enzyme digestion (*see* **Note 7**). Pick a few colonies and start bacterial cultures in 2 mL of LB supplemented with 50 µg/mL of kanamycin (we typically screen 4–5 colonies per TOPO reaction). Perform DNA purification of the plasmids (we use the QIAprep Spin Miniprep Kit, Qiagen) and screen them using restriction enzyme digestion. Select the appropriate enzymes to verify your sequence (*see* **Note 8**) and assemble the following reaction: 2 µL of enzyme buffer, 2 µL of BSA (1 mg/

mL), 3 μL of purified plasmid, and 0.5 μL of each restriction enzyme and complete to 20 μL with distilled water. Incubate for the time and temperature specified by the enzyme manufacturer and analyze the resulting fragments on agarose gel. Select at least one positive plasmid from the restriction enzyme screen and verify the cloned sequence by Sanger sequencing using M13 primers.

5. Select the correct clones and use them for the LR reaction. Assemble an LR reaction as described in **step 2** of Subheading 3.1, using the pRRL.R1R2.PGK.GFP vector as destination plasmid. Select 3–5 colonies to screen by restriction enzyme digestion as described above. Choose one positive plasmid for each sequence and use it for the establishment of the stable cell lines.

3.3   Establishment of Stable Cell Lines Containing the DNA Target Sequence

1. We have established this screening assay using murine KRAB-ZFP proteins, and for that purpose we have chosen human 293T cells. Firstly, these cells are easily transfectable, and secondly, they are functional for KAP1-mediated repression of mouse KRABZFPs [21]. The advantage of this system is that, since KRAB-ZFPs have evolved to become species specific, human 293T cells can be used to assess the function of mouse KRAB-ZFPs without interference from endogenous human factors. For the establishment of stable cell lines, 293T cells are transduced at a low multiplicity of infection (MOI) with the DNA target pRRL.R1R2.PGK.GFP lentiviral vector in order to achieve approximately one provector copy per cell. For lentiviral vector production *see* [18].

2. Plate 293T cells at a confluence of 1×10^5 cells per well into a 12-well dish. After 6 h, transduce them with lentiviral vectors containing the DNA target pRRL.R1R2.PGK.GFP plasmid at MOIs of 0.05 and 0.1 (titers calculated in 293T) (*see* **Note 9**). After 72 h, verify the GFP expression of transduced cells by flow cytometry. Select the samples with a percentage of GFP-positive cells between 5 and 20 % and expand those cells into a 10 cm culture dish.

3. Perform flow cytometry sorting of the GFP-positive cells. For the sorting, we typically use the FACSAria II machine (BD Biosciences). Put the GFP-positive sorted cells back in culture, expand them, and keep a stock in the liquid nitrogen (*see* **Note 10**).

3.4   Screening Assay

1. The screen is done by transfecting the cell line—containing the *cis*-acting sequence of interest—against the library of TFs. Before starting, make sure that you have the cell lines in culture and enough plasmid DNA for the transfections. Also make sure that you have the control cell lines and control TFs of

Control plate:

	1	2	3	4	5	6	7	8	9	10	11	12
A	Pro Lac											-
B	Pro 809											-
C	Pro Mock											-
D					Empty							-
E												-
F												-
G	Test Lac											293T
H	Test Mock											293T GFP

Assay plate (tested cell line):

	1	2	3	4	5	6	7	8	9	10	11	12
A	E	Lac										E
B	m	Mock										m
C	p											p
D	t				Tested TFs							t
E	y											y
F								Lac				
G								Mock				
H												

Fig. 2 Control and assay plate design with overview of the cell lines and plasmids to be used in transfections. The control plate is done with the control cell line containing the PBS^{Pro} sequence. The PBS^{Pro} line is transfected with its known interactor, ZFP809, with a control plasmid (in this case, expressing the protein LacZ), or with mock transfection (control for puromycin selection). A control of the tested cell line (to be transfected with the control LacZ plasmid and mock transfection) can be included. On column 12, a negative 293T control and a transfection control with 293T cells transfected with an empty pRRL.R1R2.PGK.GFP are performed to check for transfection efficiency. Other controls can be added in the empty wells. The assay plates contain the tested KRAB-ZFPs that will be used to transfect the tested cell line, as well as the transfection and mock control (which should be present at least twice)

interest. Figure 2 depicts an example of control and assay plates (*see* **Note 11**).

Each assay plate should have a negative control (we typically use a plasmid with LacZ in place of the TF ORF) and a mock transfection (control for puromycin) that are present at least twice in each plate. Also, in the control plate, as positive control for the repression assay we use KRAB-ZFP809 and its previously identified DNA target, the PBS^{Pro} sequence (TGGGGGCTCGTCCGGGATCGGGAGACCCC) [22]. A 293T cell line containing the PBS sequence upstream of the PGK-GFP cassette is generated according to Subheadings 3.2 and 3.3. Finally, a GFP-negative control (293T cells) and a transfection control (293T cells transfected with empty pRRL. R1R2.PGK.GFP plasmid) are also added to the control plate.

2. Harvest the cell line of interest and the control cell lines (*see* **Note 12**). Rinse the cells with PBS, aspirate the PBS, and replace it with enough trypsin to cover the cell layer. After 1–3 min, harvest the cells with a 10× excess of DMEM supplemented with FCS 10 % and PS. Count the cells and put them in

suspension at 5×10^4 cells/mL in DMEM supplemented with FCS and PS (calculate the final volume of cells needed according to the number of transfections).

3. The library of TFs should be arranged in 96-well plates and each transfection is performed in triplicate. Each individual transfection is done with the following volumes: 7 µL of TF DNA plasmid stock (100–150 µg of DNA), 0.6 µL of Fugene 6 Transfection Reagent, and 2.4 µL Opti-MEM (Life Technologies) (*see* **Note 13**). First prepare a general Fugene and Opti-MEM mix for all the plates. Then prepare a 3.5× DNA mix for each 96-well TF DNA stock plate to be tested: in a 96-well V-bottom plate pipet 24.5 µL of each TF plasmid, and add 10.5 µL of the Fugene-Opti-MEM mix to each well. Incubate for 5 min, mix gently, and pipet 10 µL of the DNA mix to each 96-well culture plate triplicate. We typically perform this screen using the Sciclone ALH 3000 (Caliper Life Sciences) liquid-handling robot, but it can also be done with multichannel pipets.

4. Add 100 µL of the cell suspension directly to the 96-well culture plates containing the TF plasmid mix (*see* **Note 14**). To add the cells (and liquids throughout the screen) we typically use the Multidrop Combi dispenser (Thermoscientific) equipped with a standard tube dispensing cassette (*see* **Note 15**).

5. The next day (16–20 h after transfection), add a final concentration of 5 µg/mL of doxycycline to induce TF expression from the TRE promoter. Assemble a mix of DMEM (supplemented with FCS and PS) and 10 µg/mL doxycycline, and add 100 µL to each well. If using the Multidrop dispenser, use medium speed.

6. The following day, add a final concentration of 1 µg/mL of puromycin to each well to select for the presence of the TRE vector. Dilute puromycin in PBS to a concentration of 21 µg/mL and add 10 µL to each well.

7. On day 6 after transfection, harvest the cells for the readout. For the harvesting, we recommend to process one 96-well plate at a time. Using a multichannel pipet, aspirate the media from the wells, rinse the cells with 50 µL of PBS, aspirate the PBS, and add 120 µL of RIPA buffer (supplemented with protease inhibitors) to all the wells. Shake the plate for 15 s (using the Multidrop) and put it on ice for at least 10 min. Keep the plates on ice for the following procedures.

8. Once all the plates are processed mix the cell lysate a couple of times and transfer 90 µL of each well to a new black 96-well flat-bottom plate (*see* **Note 16**). Store the remaining cell lysates at −20 °C for further BCA analysis. Measure the fluorescence of the samples on a plate reader equipped with fluorescence

readout, with excitation set at 485 nm and emission at 520 nm (we typically use the Tecan Infinite F500) (*see* **Note 17**). Calculate the fluorescence value by subtracting the value of a well containing only buffer from all the other wells.

9. Thaw the frozen lysates and perform protein quantification by BCA in 96-well plates. Prepare the BSA standards through serial dilution in RIPA buffer to obtain the following concentrations: 2, 1, 0.500, 0.250, 0.125, 0.062, and 0.031 mg/mL. Place 10 μL of each standard and 10 μL of RIPA buffer supplemented with protease inhibitor (blank) in the first column of the 96-well plate and pipet 10 μL of the lysates into the remaining wells (exclude column 12 of the sample plate) (*see* **Note 18**). Prepare a master mix of BCA reagent mixture containing 200 μL of BCA reagent A and 4 μL of BCA reagent B per well. Add 200 μL of BCA reagent mix to each well and incubate the plates for 30 min at 37 °C in the dark. Immediately measure the absorbance at 570 nm and use the standards to calculate the protein concentration of the samples.

10. Calculate the normalized fluorescence by dividing the fluorescence by the protein concentration. Calculate the mean and standard deviation of the triplicates and plot them for visualization. We typically analyze the results using R. The candidate hits can be identified by first selecting the ten TFs with the lowest normalized fluorescence values per plate, and then we select only the ones that are among the ten lowest on all three replicates of the plate (*see* **Note 19**).

11. After selecting the candidates, they need to be tested by flow cytometry to identify the hits. Transfect the 293T cell line containing the retrotransposon upstream of the PGK-GFP cassette with the candidate TFs. Plate the 293T cells in 24-well plates at 1.5×10^4 cells/well. From 3 to 10 h later, transfect the cells with the TF plasmid from the stock. Mix 0.38 μL Fugene 6, 14.62 μL Opti-MEM, and 10 μL DNA from the stock (150–200 ng), per transfection, per well (*see* **Note 20**). Incubate the mix for 5 min at RT and add 25 μL to the cells. Transfect each TF in duplicate and also perform a transfection control in an unrelated 293T pRRL.R1R2.PGK.GFP cell line (e.g., the PBS^Pro cell line), to control for unspecific effects. Include LacZ as a negative transfection control for the test cell lines. Also, as a positive repression control, transfect the PBS^Pro cell line with ZFP809. On the next day (16–20 h after transfection), add doxycycline to a final concentration of 5 μg/mL to induce TF expression. The following day, add puromycin to a final concentration of 1 μg/mL. On day 6 after transfection, harvest the cells, resuspend them in DMEM, spin them, remove the media, and resuspend them in 300 μL of PBS supplemented with 2 % FCS. Analyze them by flow cytometry to evaluate

GFP expression. Calculate the median fluorescence intensity of the different TFs and compare them with the control transfection (LacZ) and with the control cell line. You have a hit when a TF specifically represses the DNA sequence of interest.

4 Notes

1. We recommend verification of the cloned/synthesized cDNA at the Entry plasmid level by DNA sequencing to avoid perpetuating errors or mutations to the next steps.

2. For a large number of TFs, LR reactions can be performed in 96-well plates and transformed into 96-well plates containing competent bacteria. In that case, instead of plating them onto agar plates, transformed bacteria are grown in 96-well 1 mL culture plates in LB medium supplemented with ampicillin (50 µg/mL). Hence, there is no selection of single clones and the positively transformed population is directly used for the bacterial culture, glycerol stock, and PCR check.

3. When dealing with repetitive sequences, as in the case of KRAB-ZFPs (and also for the DNA target sequence when cloning retrotransposons), we recommend performing the bacterial culture at 30 °C for 24 h instead of 37 °C ON. The lower temperature decreases the potential plasmid recombination.

4. Manipulate the glycerol stock very carefully, avoiding unnecessary freezing and thawing. Furthermore, besides the glycerol stock, we advise keeping a stock of purified DNA.

5. Make sure that the designed primers are unique. We typically use the UCSC genome browser PCR tool (http://genome.ucsc.edu/cgi-bin/hgPcr?command=start), NCBI primer blast (http://www.ncbi.nlm.nih.gov/tools/primer-blast/), and SIB Tag Scan (http://ccg.vital-it.ch/tagger/tagscan.html). Furthermore, verify that the reverse primer does not contain the complementary sequence (CACC) to the TOPO overhang at its 5′ end. Furthermore, when cloning several *cis*-acting sequences at once we recommend the insertion of a unique site for restriction enzyme on the 5′ end (remember to keep the CACC on the outer 5′) of at least one of the primers. This facilitates the bacterial colony screening.

6. The PCR reaction should be optimized to obtain as product a clear single band of the correct size. If this is not possible, the band of interest can be cut from the gel and purified using the MinElute Gel Extraction kit (Qiagen).

7. When cloning many target sequences, colony PCR screening can be used instead of restriction enzyme digestion. We prefer

the latter because it is more effective at detecting undesired sequences.

8. *Ssp*I can be used as a single cutter in the plasmid and another single cutter for the insert should be selected. If unique restriction enzyme sites were included on the primers sequences, they can be used at this step.

9. We have tested different MOIs, and an MOI of 0.1 or 0.05 usually yields a percentage of GFP-positive cells between 5 and 20 % for 293Ts, but other MOIs can be tested. This percentage is predicted to yield about one vector copy number per cell and, according to our copy number tests, that is normally true. If using cells other than 293Ts, the MOI needs to be adjusted to yield between 5 and 20 % of GFP-positive cells.

10. The established cell lines can be checked for vector copy number by qPCR in order to determine if they are in a range of 1 copy per cell.

11. We recommend separate assembling of the control and assay plates, for easier reuse and changes of the plates. The control plate can also be adapted to each screen and can include other positive and negative controls.

12. We recommend splitting the 293T cell lines the day before the screen, from a confluent plate to a new one at 1:3. We have observed that this can improve the transfection efficiency.

13. We have obtained good transfection efficiency with these amounts of reagents, but it can vary greatly according to the DNA quality. If necessary, different amounts and ratios of Fugene and DNA can be tested.

14. The cells have to be added to the plates containing the DNA-Fugene mix at most 30 min to 1 h after the DNA mix has been prepared. Waiting for longer periods can decrease the transfection efficiency.

15. When using the dispenser, make sure to keep it sterile by placing it inside a laminar flow and by decontaminating it with ethanol before use. Unless otherwise stated, use maximum speed for liquid dispensing. Also, in the case of the control cell lines that are plated only into few wells, manual handheld dispenser pipets can be used, instead of the dispenser.

16. Be careful not to take up any bubbles when transferring the lysate to the new plate. As the readout is done from the upper part of the plate, the presence of any bubbles can compromise the readout.

17. For the readout it can be necessary to adjust measurement parameters according to the cell line. Before each readout we typically optimize the gain and the Z-position.

18. Column 12 is excluded for practical reasons and it is dispensable at this step.

19. The candidate hits can also be identified by significant differences using paired *t*-test, when compared to LacZ control transfections. However, this method has proven to be too stringent in some cases and some true biological candidates are missed. For that reason, we have opted to use a less stringent criterion (ten lowest candidates). We have observed that the latter decreases the chances of identifying false negatives that could be discarded as candidates due to the noise. On the other hand, this method leads to the increase of false positives as candidate hits, which are included in the candidate test by flow cytometry. However, it should be noted that after the flow cytometry test we have observed no false positives, and all candidates we have identified so far are be *bona fide* interactors.

20. We typically use these amounts of DNA and Fugene for the transfection in 24-well plates in order to save reagents. However, if desired, an alternative is to use 2.5 times the amounts used in the 96-well plate transfections mentioned in **step 3** of Subheading 3.4.

Acknowledgements

References

1. Rowe HM, Trono D (2011) Dynamic control of endogenous retroviruses during development. Virology 411(2):273–287. doi:10.1016/j.virol.2010.12.007

2. Leung DC, Lorincz MC (2012) Silencing of endogenous retroviruses: when and why do histone marks predominate? Trends Biochem Sci 37(4):127–133. doi:10.1016/j.tibs.2011.11.006

3. Emerson RO, Thomas JH (2009) Adaptive evolution in zinc finger transcription factors. PLoS Genet 5(1):e1000325

4. Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM (2009) A census of human transcription factors: function, expression and evolution. Nat Rev Genet 10(4):252–263

5. Friedman JR, Fredericks WJ, Jensen DE, Speicher DW, Huang XP, Neilson EG, Rauscher Iii FJ (1996) KAP-1, a novel core- pressor for the highly conserved KRAB repression domain. Genes Dev 10(16):2067–2078

6. Schultz DC, Friedman JR, Rauscher Iii FJ (2001) Targeting histone deacetylase complexes via KRAB-zinc finger proteins: the PHD and bromodomains of KAP-1 form a cooperative unit that recruits a novel isoform of the Mi-2α subunit of NuRD. Genes Dev 15(4):428–443

7. Schultz DC, Ayyanathan K, Negorev D, Maul GG, Rauscher Iii FJ (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(8):919–932

8. Nielsen AL, Ortiz JA, You J, Oulad-Abdelghani M, Khechumian R, Gansmuller A, Chambon P, Losson R (1999) Interaction with members of

the heterochromatin protein 1 (HP1) family and histone deacetylation are differentially involved in transcriptional silencing by members of the TIF1 family. EMBO J 18(22):6385–6395

9. Sripathy SP, Stevens J, Schultz DC (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(22):8623–8638

10. Rowe HM, Jakobsson J, Mesnard D, Rougemont J, Reynard S, Aktas T, Maillard PV, Layard-Liesching H, Verp S, Marquis J, Spitz F, Constam DB, Trono D (2010) KAP1 controls endogenous retroviruses in embryonic stem cells. Nature 463(7278):237–240. doi:10.1038/nature08674

11. Turelli P, Castro-Diaz N, Marzetta F, Kapopoulou A, Raclot C, Duc J, Tieng V, Quenneville S, Trono D (2014) Interplay of TRIM28 and DNA methylation in controlling human endogenous retroelements. Genome Res 24(8):1260–1270. doi:10.1101/gr.172833.114

12. Matsui T, Leung D, Miyashita H, Maksakova IA, Miyachi H, Kimura H, Tachibana M, Lorincz MC, Shinkai Y (2010) Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. Nature 464(7290):927–931. doi:10.1038/nature08858

13. Liu S, Brind'Amour J, Karimi MM, Shirane K, Bogutz A, Lefebvre L, Sasaki H, Shinkai Y, Lorincz MC (2014) Setdb1 is required for germline development and silencing of H3K9me3-marked endogenous retroviruses in primordial germ cells. Genes Dev 28(18):2041–2055. doi:10.1101/gad.244848.114

14. Castro-Diaz N, Ecco G, Coluccio A, Kapopoulou A, Yazdanpanah B, Friedli M, Duc J, Jang SM, Turelli P, Trono D (2014) Evolutionarily dynamic L1 regulation in embryonic stem cells. Genes Dev 28(13):1397–1409. doi:10.1101/gad.241661.114

15. Jacobs FMJ, Greenberg D, Nguyen N, Haeussler M, Ewing AD, Katzman S, Paten B, Salama SR, Haussler D (2014) An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. Nature 516(7530):242–245. doi:10.1038/nature13760

16. Simicevic J, Deplancke B (2010) DNA-centered approaches to characterize regulatory protein-DNA interaction complexes. Mol Biosyst 6(3):462–468. doi:10.1039/b916137f

17. Dey B, Thukral S, Krishnan S, Chakrobarty M, Gupta S, Manghani C, Rani V (2012) DNA-protein interactions: methods for detection and analysis. Mol Cell Biochem 365(1-2):279–299. doi:10.1007/s11010-012-1269-z

18. Barde I, Salmon P, Trono D (2010) Production and titration of lentiviral vectors. In: Jacqueline N Crawley et al. (ed) Current protocols in neuroscience, Chapter 4: Unit 4.21. doi:10.1002/0471142301.ns0421s53

19. Hens K, Feuz JD, Deplancke B (2012) A high-throughput gateway-compatible yeast one-hybrid screen to detect protein-DNA interactions. Methods Mol Biol 786:335

20. Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JAM (2007) Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Res 35(Suppl 2):W71–W74

21. Rowe HM, Friedli M, Offner S, Verp S, Mesnard D, Marquis J, Aktas T, Trono D (2013) De novo DNA methylation of endogenous retroviruses is shaped by KRAB-ZFPs/KAP1 and ESET. Development 140(3):519–529. doi:10.1242/dev.087585

22. Wolf D, Goff SP (2009) Embryonic stem cells use ZFP809 to silence retroviral DNAs. Nature 458(7242):1201–1204

III. Results – Part II: Gm6871 and KAP1 regulate LINE-1

Manuscript title

Evolutionally dynamic L1 regulation in embryonic stem cells

Authors

Nathaly Castro-Diaz, Gabriela Ecco, Andrea Coluccio, Adamandia Kapopoulou, Benyamin Yazdanpanah, Marc Friedli, Julien Duc, Suk Min Jang, Priscilla Turelli, and Didier Trono

Summary of results and contribution

One of the candidates identified by our screen was Gm6871. We initially identified B2/SINE sequences as targets of this KRAB-ZFP, but upon ChIP-seq characterization we observed that it was largely recruited to LINE1s. That led us to characterize further its role in the context of another project in the lab, regarding the KRAB/KAP1 control of LINE-1s.

LINE-1s are the only autonomous TE still active in humans. While KAP1 role in regulating ERVs in ES cells was better elucidated, its contribution to the regulation of LINEs was unknown. On the contrary, it was believed that piRNAs played a more important role in regulating LINEs. In the present manuscript, we show for the first time that KRAB-ZFPs and KAP1 regulate LINE-1s, and provide compelling evidence that KRAB-ZFPs regulate endogenous retroviruses. Prior to

our work, other groups had observed KRAB-ZFP-mediated regulation of exogenous retroviruses only (such as ZFP809 and MLV), with one group suggesting with only little evidence, a role for ZFP819 in regulating IAPs and LINEs. Moreover, we found that the KRAB/KAP1 complex mainly targets more ancient LINE-1 families, while younger (such as human-specific) elements are regulated by DNA methylation – a result consistent with the discovery that piRNAs regulate the youngest L1Hs family in human ES cells –. This led us to propose an evolutionary model in which newly emerged LINE-1s are initially repressed by small RNA-induced DNA methylation, and, with time, KRAB-ZFPs repressors are selected to bind to these TEs.

My contribution to this work comprises nearly all the data generated for the mouse system. The key contribution was to show for the first time that the KRAB-ZFP family is implicated in the control of LINEs. We initially identified Gm6871 to bind to LINE1 elements through ChIP-seq experiments. LINE-1s of the L1MdF2 and L1MdF3 families were overrepresented in the detected ChIP-seq peaks, and the Gm6871 identified motif was present in 95% of bound elements. We further characterized Gm6871 by immunofluorescence and immunoprecipitation and we observed that it is mainly expressed in ES cells, it localizes to the nucleus, and it interacts with KAP1 to repress LINE-1s. In parallel, we generated ChIP-seq data for KAP1 in mouse ES cells and observed that, similarly to what is seen in humans, the youngest mouse LINE families are not bound by KAP1, but the co-repressor targets L1MdF and L1MdF2 families, estimated to be between 7.3 and 3.8 million years old. Finally, RNA-seq experiments further demonstrated that Gm6871/KAP1-bound LINE-1s were significantly more expressed in mouse ES cells depleted for Gm6871 when compared to control, confirming the role of Gm6871 and KAP1 in regulating these elements in mouse ES cells.

Evolutionally dynamic L1 regulation in embryonic stem cells

Nathaly Castro-Diaz, Gabriela Ecco, Andrea Coluccio, Adamandia Kapopoulou, Benyamin Yazdanpanah, Marc Friedli, Julien Duc, Suk Min Jang, Priscilla Turelli, and Didier Trono[1]

School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland

Mobile elements are important evolutionary forces that challenge genomic integrity. Long interspersed element-1 (L1, also known as LINE-1) is the only autonomous transposon still active in the human genome. It displays an unusual pattern of evolution, with, at any given time, a single active L1 lineage amplifying to thousands of copies before getting replaced by a new lineage, likely under pressure of host restriction factors, which act notably by silencing L1 expression during early embryogenesis. Here, we demonstrate that in human embryonic stem (hES) cells, KAP1 (KRAB [Krüppel-associated box domain]-associated protein 1), the master cofactor of KRAB-containing zinc finger proteins (KRAB-ZFPs) previously implicated in the restriction of endogenous retroviruses, represses a discrete subset of L1 lineages predicted to have entered the ancestral genome between 26.8 million and 7.6 million years ago. In mice, we documented a similar chronologically conditioned pattern, albeit with a much contracted time scale. We could further identify an L1-binding KRAB-ZFP, suggesting that this rapidly evolving protein family is more globally responsible for L1 recognition. *KAP1* knockdown in hES cells induced the expression of KAP1-bound L1 elements, but their younger, human-specific counterparts (L1Hs) were unaffected. Instead, they were stimulated by depleting DNA methyltransferases, consistent with recent evidence demonstrating that the PIWI–piRNA (PIWI-interacting RNA) pathway regulates L1Hs in hES cells. Altogether, these data indicate that the early embryonic control of L1 is an evolutionarily dynamic process and support a model in which newly emerged lineages are first suppressed by DNA methylation-inducing small RNA-based mechanisms before KAP1-recruiting protein repressors are selected.

[*Keywords*: DNA methylation; evolution; KAP1; KRAB-ZFPs; LINE1; embryonic stem cells]

Supplemental material is available for this article.

Received March 20, 2014; revised version accepted May 28, 2014.

More than half of the human genome is derived from mobile elements, most of which are retrotransposons spreading by reverse transcription of an RNA intermediate and integration of the resulting DNA product (Cordaux and Batzer 2009). These endogenous retroelements (EREs) represent essential evolutionary forces but also threats to genomic integrity and, as such, are subjected to transcriptional repression from the earliest stages of embryogenesis. Reciprocal selective pressures are exerted between EREs and host defenses engaged in their control, which can often be traced through phylogenetic studies (Furano and Boissinot 2008).

Long interspersed element-1 (L1, also known as LINE-1) is the only autonomous transposon still active in humans. About 500,000 copies of L1 are present in the human genome, amounting to some 20% of its DNA content. Many L1 integrants are 5′-truncated owing to

the abortive tendency of the target-primed reverse transcription mechanism used by this class of retroelements. Nevertheless, the human genome contains some 100 retrotransposition-competent L1 elements, >40 of which are highly active (Brouha et al. 2003; Beck et al. 2011). Furthermore, L1 provides the *trans*-acting functions required for the transposition of nonautonomous retroelements such as SINEs (short interspersed nuclear elements, which include Alu repeats in humans) and SVAs (SINE–VNTR–Alu, a composite hominoid-restricted ERE) (Dewannieux et al. 2003; Finnegan 2012). The 6- to 7-kb-long genome of a full-length L1 comprises a 5′ untranslated region (UTR) promoter region; two ORFs encoding, respectively, a nucleic acid-binding protein and a product endowed with endonuclease and reverse transcriptase activity; and a 3′ UTR ending with a poly(A) tail (Babushok

48

and Kazazian 2007; Rosser and An 2012). As other EREs, L1 shapes transcriptional networks, for instance, through L1-initiated cellular transcripts or L1-contained enhancers or insulators (Speek 2001; Nigumann et al. 2002; Matlik et al. 2006; Slotkin and Martienssen 2007; Faulkner et al. 2009). L1 elements present in the human or mouse genomes can be subdivided into subfamilies based on nucleotide substitutions, insertions, and/or deletions. Furthermore, phylogenetic studies interestingly indicate that this class of retroelements displays an unusual pattern of evolution in which a single L1 lineage at a time is generally active within the genome of a species and amplifies to thousands of copies before its replacement by a new lineage, likely under selective pressures exerted by host defense mechanisms (Cordaux and Batzer 2009).

EREs are silenced during early embryogenesis by histone methylation, histone deacetylation, and DNA methylation through sequence-specific mechanisms that counter the wave of epigenetic modifications—mainly DNA demethylation—required for the reprogramming typical of this developmental period (Rowe and Trono 2011). For endogenous retroviruses (ERVs), key mediators of this process are the DNA-binding Krüppel-associated box domain-containing zinc finger proteins (KRAB-ZFPs) and their cofactor, KAP1 (KRAB-associated protein 1), also known as TRIM28 (tripartite motif protein 28) (Wolf and Goff 2007, 2009; Matsui et al. 2010; Rowe et al. 2010). In human embryonic stem (hES) and mouse ES (mES) cells, the KRAB-ZFP-mediated docking of KAP1 at EREs triggers the formation of heterochromatin through the recruitment of the SETDB1 (also known as ESET) histone methyltransferase, responsible for trimethylating histone 3 at Lys9; histone deacetylases; and HP1 (heterochromatin protein 1), which collectively induce transcriptional repression (Schultz et al. 2002; Ivanov et al. 2007). The further recruitment of DNA methyltransferases (DNMTs) results in permanent silencing marks, which are subsequently maintained throughout development without need for persistent expression of sequence-specific ERE-recognizing repressors (Quenneville et al. 2012; Rowe et al. 2013a).

Previous studies detected a modest up-regulation of L1 in KAP1- or SETDB1-deleted mES cells (Matsui et al. 2010; Rowe et al. 2010), suggesting that this class of ERE is regulated by alternative pathways. In line with this hypothesis, recent data pointed to the importance of small RNA-based repression in the control of L1 expression in human pluripotent stem cells (Ciaudo et al. 2013; Fadloun et al. 2013; Heras et al. 2013; Marchetto et al. 2013). The present study reveals that the KRAB/KAP1 pathway and DNA methylation, the known output of small RNA-based mechanisms, are both engaged in restricting L1 in ES cells but act on evolutionarily distinct sets of elements, which provides a remarkable illustration of the reciprocal selective pressures exerted between EREs and the host mechanisms responsible for their control.

Results

KAP1 *associates with full-length L1 in hES cells*

In order to investigate a possible role for KAP1 in the control of L1, we performed chromatin immunoprecipitation (ChIP) followed by deep sequencing (ChIP-seq) in H1 hES cells. We found that ~8% of the total of L1-derived sequences annotated in the human genome somehow overlapped with KAP1 peaks in these cells. As most of the L1 sequences are 5′ truncated, we reasoned that only L1 copies endowed with a 5′ UTR would require transcriptional control; hence, we focused our analysis on L1 sequences >5 kb, assuming that they corresponded in their majority to full-length integrants. Fulfilling this prediction, 52% of these L1 sequences harbored a KAP1 peak, usually over their first 1000 base pairs (bp), contrasting with only 2% of elements <5 kb (Fig. 1A,B; Supplemental Fig. S1). Furthermore, while KAP1 was present on full-length L1s in hES cells, it was not significantly enriched at any L1-derived sequence in the differentiated human cell line HEK293 (Fig. 1A,B; Supplemental Fig. S1). A more detailed mapping of the ChIP-seq tags indicated that most of the KAP1 peaks targeted the middle region of the 5′ UTR, encompassing L1 nucleotides +300 to +600 (Fig. 1C). H3K9me3-specific ChIP-seq analyses confirmed a strong coincidence between KAP1 peaks and deposition of this repressive mark at the 5′ end of full-length L1 elements (Fig. 1D), with only a small minority of L1 bearing only KAP1 and with H3K9me3 seldom detected without the corepressor (Fig. 1E). Extending these findings, we could document the accumulation of KAP1 and H3K9me3 at the 5′ end of full-length L1 elements by performing the same type of analysis in mES cells (Supplemental Fig. S2).

KAP1-*bound (KB) L1 sequences can act as* cis-*repressors in hES cells*

In order to assess the functional consequences of KAP1 recruitment at L1 sequences, we cloned KB regions (as defined by ChIP-seq in hES cells) from an L1PA4 and an L1PA5 element upstream of a PGK-GFP reporter cassette within the context of lentiviral vectors using corresponding non-KB (NKB) L1 fragments as negative controls (Fig. 2A,B). We then monitored GFP fluorescence in hES and 293T cells transduced with these vectors (Fig. 2C; Supplemental Fig. S3A). Expression from vectors containing full-length KB L1 fragments was progressively repressed in hES cells but not in 293T cells. In contrast, GFP fluorescence induced by the empty vector or harboring NKB L1 fragments remained strong over time in both cell types. ChIP followed by quantitative PCR (ChIP-qPCR) with PGK-specific primers confirmed that repression correlated with KAP1 enrichment and deposition of H3K9me3 (Fig. 2D; Supplemental Fig. S3B). Furthermore, L1-mediated KAP1 recruitment strongly stimulated the CpG methylation of the adjacent PGK promoter in hES cells (Fig. 2E). In order to define further the L1PA4- and L1PA5-derived KAP1-recruiting elements identified through this assay, we cut these ~1-kb-long sequences
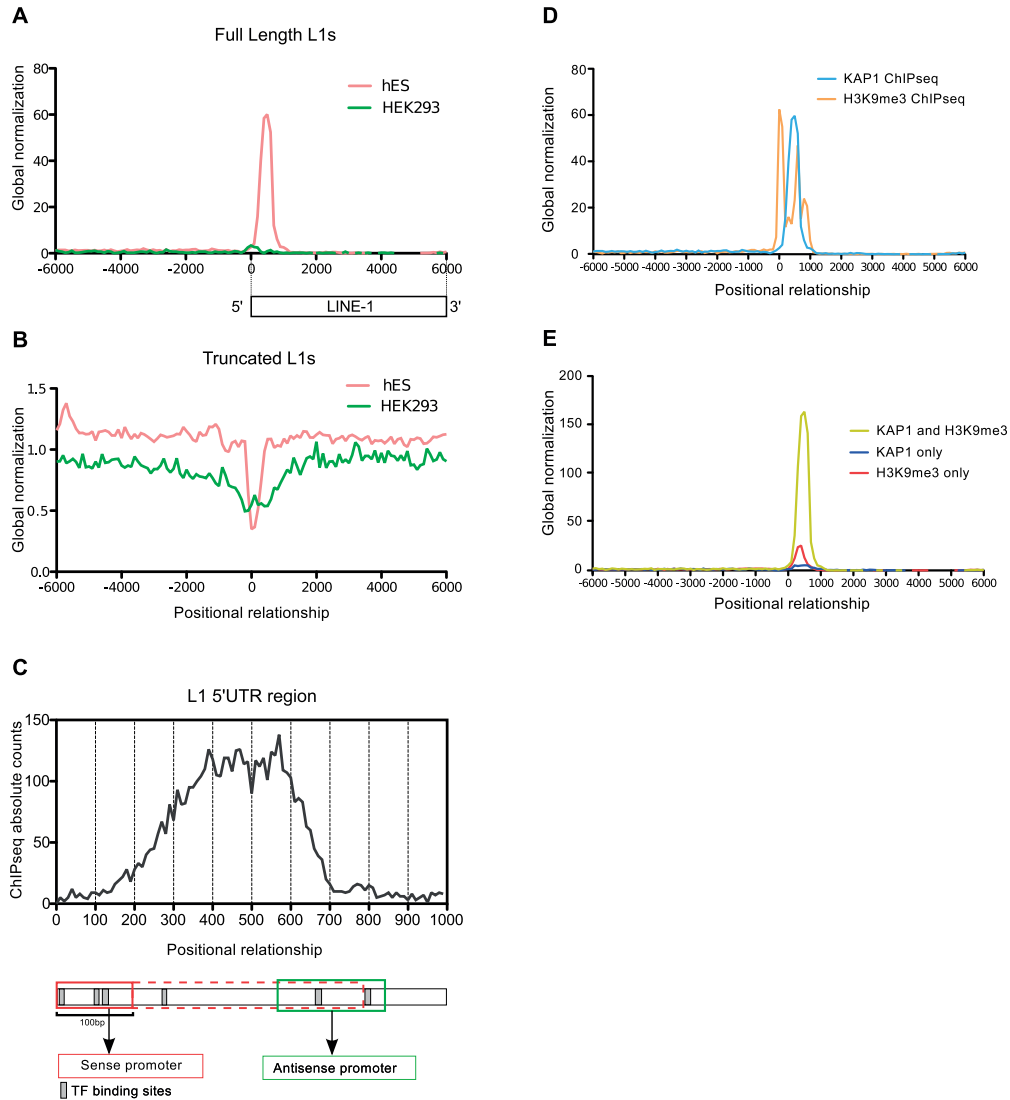
Figure 1. KAP1 coincides with H3K9me3 at the 5′ end of full-length L1 in hES cells. Distribution of ChIP-seq KAP1 peaks relative to the 5′ end of full-length elements (*A*) or the center of truncated L1 elements (*B*) in hES and HEK293 cells. The profiles were normalized to the total number of ChIP-seq peaks for each cell line. (*C*) KAP1 ChIP-seq peak distribution over the first kilobase of L1. The L1 5′ UTR is schematized *below*, with sense and antisense promoters as red and green boxes, respectively. Sense promoter is diversely depicted as mainly located in the first 100 bp or extending up to 700 bp. (*D*) Overlap of KAP1 and H3K9me3 ChIP-seq tags relative to the 5′ end of full-length L1 elements. (*E*) Relative frequency of KAP1+H3K9me3, KAP1-only, and H3K9me3-only peaks at this location.

into subfragments of ∼200 bp (Fig. 2B). This revealed that the *cis*-repressors contained in these retrotransposons coincided with the top of the corresponding KAP1 ChIP-seq peak (Fig. 2B; Supplemental Figs. 2FG, S3C,D). Of note, the KAP1-binding L1PA4 D subfragment induced faster and stronger repression than its full-length parent, suggesting that the latter contains elements with conflicting influences. In addition, while the tested L1-PA4 leader contained one KAP1-responsive *cis*-repressor, its

L1PA5 counterpart harbored two such elements. Collectively, these data support a model in which the KAP1 corepressor is tethered to the 5′ end of subfamilies of L1 elements in hES cells, triggering their epigenetic silencing. Of note, our attempt to abrogate L1-induced KAP1-mediated repression of the PGK promoter by shRNA-mediated KAP1 depletion failed, probably because sufficient levels of *KAP1* knockdown could not be maintained over time (data not shown).
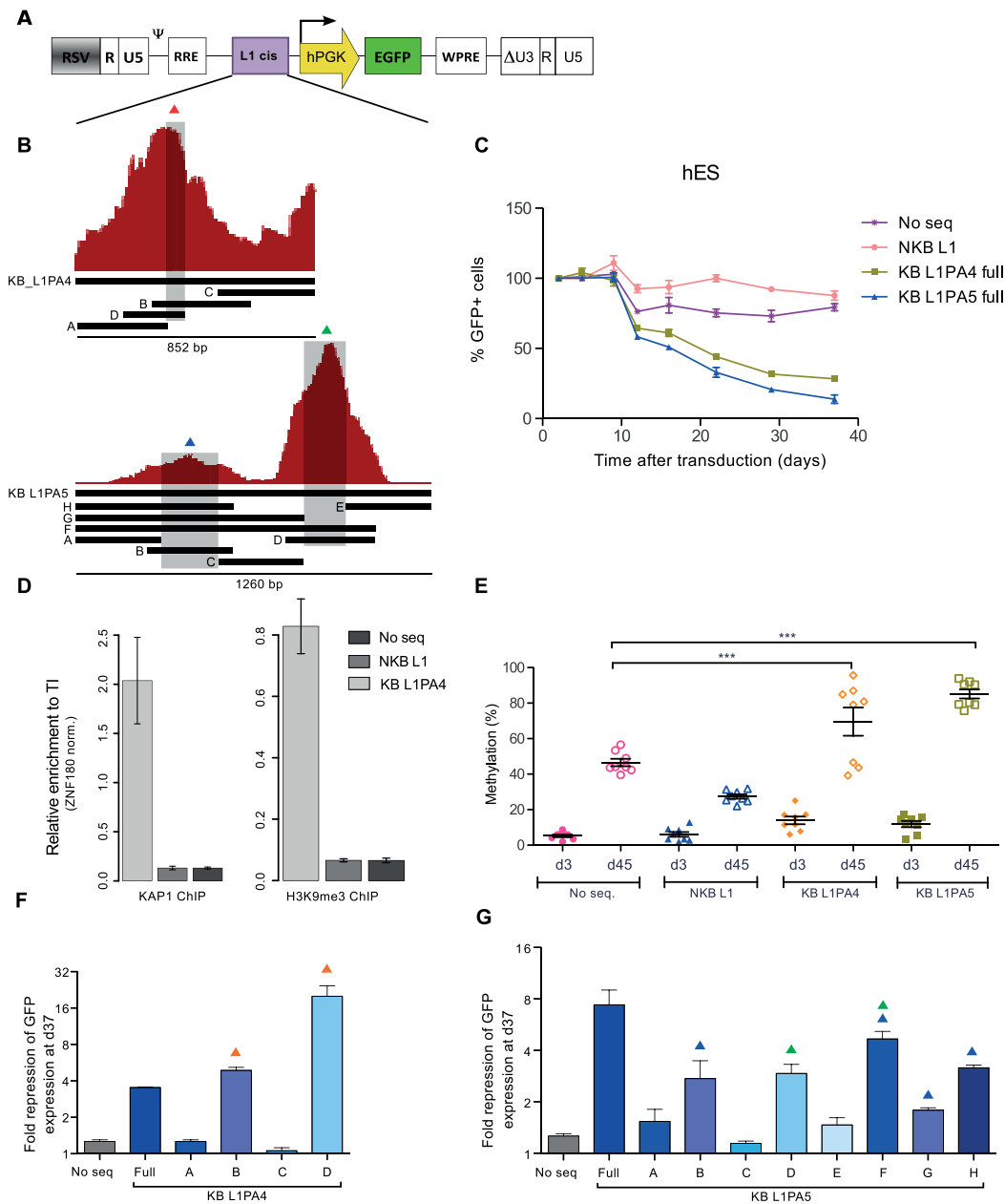
Figure 2. KAP1-binding L1 fragments can induce repression and DNA methylation of a heterologous promoter in hES. (*A*) KB (KB L1PA4 and KB L1PA5) and NKB (NKB L1PA4) L1 sequences were cloned in depicted lentiviral vector upstream of a PGK-EGFP expression cassette. The resulting vectors were transduced in hES, and EGFP expression was monitored over time by FACS. (*B*) Schematic representation of the KAP1 ChIP mapped on the L1PA4 and L1PA5 5′ end, with indication of derived fragments and subfragments cloned in the vector depicted in *A*. (*C*) Monitoring of GFP expression in hES cells transduced with the indicated vectors. (No seq) Lentiviral vector with no ERE-derived fragment upstream of the expression cassette. The figure shows the mean and SD of two biological replicates. (*D*) KAP1 and H3K9me3 recruitment to indicated lentiviral vectors in hES, assessed 35 d after transduction by ChIP-qPCR using PGK-specific primers. The figure illustrates the mean and SD of technical replicates. This experiment was performed twice with similar results (see Supplemental Fig. S3). Relative enrichment was determined by normalizing to a known positive (ZNF180 3′ UTR) control. (*E*) Influence of the L1 *cis*-acting sequences on the methylation of the nearby PGK promoter. Methylation of eight CpG positions was evaluated by pyrosequencing at days 4 and 35 after transduction of hES cells with the PGK-GFP lentiviral vectors. Mean and standard error mean (SEM) of two biological replicates is shown. Statistical differences were determined by one-way ANOVA test using the Bonferroni multiple test adjustment. (***) $P \leq 0.001$. (*F,G*) Fold repression of the indicated vectors containing L1 subfragments described in *B*, assessed 37 d after transduction (respect to day 5). Overtime fold repression is presented in Supplemental Figure S3. Colored triangles indicate the presence of L1 sequences overlapping with the summits of the respective KAP1 ChIP-seq peaks as depicted in *B*.

51

The KRAB–KAP1 *system recognizes evolutionarily discrete subfamilies of human and mouse L1*

L1 displays an unusual pattern of evolution in mammals, with a single active lineage at any given time (Smit et al. 1995; Khan 2005; Sookdeo et al. 2013). This allows the approximate aging of L1 integrants in the genome of higher species and their sequence-based grouping in chronologically ordered subfamilies. By exploiting this feature, we could determine that KAP1 associated with only a small percentage of full-length human L1 belonging to lineages older than 26.8 million years (L1MA4 to L1PA7 subfamilies) and was practically absent from L1Hs; that is, L1 elements that invaded the human ancestral genome after the human–chimpanzee divergence some 7.6 million years ago. In contrast, KAP1 was recruited to a high fraction of L1PA6 to L1PA3 elements, peaking at >80% for the L1PA5, L1PA4, and L1PA3 subfamilies (Fig. 3A). Furthermore, H3K9me3 enrichment over full-length L1s from different subfamilies matched their KAP1-binding pattern, with this histone mark absent from very old L1s, highly enriched on the KAP1-recruiting L1PA5/PA4 and the rare KB L1HS, and present, albeit at much lower levels, on KAP1-devoid L1Hs (Supplemental Fig. S4). Remarkably, a similarly chronological pattern of KAP1 recruitment was recorded in mES cells, with KAP1 enrichment the highest on L1MdF and L1MdF2, estimated to be between 7.3 million and 3.8 million years old, and much lower on both older and younger L1 integrants (Fig. 3B).

In the context of a screen based on ChIP-seq of mES cells with HA-tagged KRAB-ZFPs, we identified Gm6871 as a L1 ligand with 104 full-length elements bound by both KAP1 and this mouse-specific KRAB-ZFP (Fig. 3C). The majority of them belonged to the L1MdF2 (64%) and L1MdF3 (13%) subfamilies, and a search performed on all Gm6871-recruiting sequences identified a putative Gm6871 DNA-binding motif (Fig. 3D) present in 95% of these L1 elements, contrasting with only 0.2% of elements from the younger L1MdA and L1MdT subfamilies. *Gm6871* was so far only a predicted gene, but we could detect its expression in both pluripotent and differentiated cells, albeit with higher levels in mES cells compared with fibroblasts (Supplemental Fig. S5), as previously described for many KRAB-ZFPs (Corsinotti et al. 2013). We also could document the nuclear localization of a HA-tagged derivative of Gm6871 expressed in mES cells by lentivector-mediated transduction (Supplemental Fig. S6A) and demonstrate an interaction between Gm6871 and KAP1 by coimmunoprecipitation of extracts from Gm6871-HA-expressing mES and 293T cells and by KAP1-GST pull-down assay (Supplemental Fig. S6B,C). To ascertain the functional relevance of this interaction, we depleted endogenous Gm6871 in mES cells by lentivector-mediated RNAi and evaluated L1 mRNA expression by RNA deep sequencing (RNA-seq). Consistent with the ChIP-seq results, upon *Gm6871* knockdown, we observed a significant increase in the levels of L1 sequences identified in control cells as binding either KAP1+Gm6871 or Gm6871 alone but not of L1s bound only by KAP1 or associated with neither protein (Fig. 3C,E). In order to

demonstrate further the implication of Gm6871 in the control of specific L1s, we performed a KAP1 ChIP in mES cells transduced with a control or *Gm6871*-directed shRNA-expressing lentiviral vector followed by qPCR with primers specific for three KAP1- and Gm6871-associated L1 elements. As controls, we included ICR (imprinting control region) sequences, known to recruit KAP1 independently of Gm6871 (Quenneville et al. 2011), and another genomic locus highly enriched for KAP1 and Gm6871 in our ChIP-seqs. The results revealed a mild but reproducible reduction of KAP1 enrichment at the tested Gm6871-recruiting loci upon *Gm6871* knockdown, while association of the corepressor with ICRs was unaffected (Supplemental Fig. S7). Finally, RT-qPCR performed in mES cells confirmed that these L1 elements were up-regulated upon removal of SETDB1 (Supplemental Fig. S8), the histone methyltransferase responsible for H3K9me3 induction by the KRAB–KAP1 complex (Schultz et al. 2002; Iyengar and Farnham 2011). Altogether, these results demonstrate that Gm6871 tethers KAP1 and associated chromatin modifiers to a specific subset of murine L1s and strongly suggest that the KRAB-ZFP family at large is involved in the sequence-specific repression of LINEs in higher vertebrates.

The KB subset of L1 is activated by KAP1 *depletion in hES cells*

To probe the impact of KAP1 on the transcriptional control of L1, we used lentivector-mediated RNAi coupled with RNA-seq in hES cells. Global expression of full-length L1 was increased in KAP1-depleted compared with control ES cells (Fig. 4A; Supplemental Fig. S9), but this difference came only from KB elements (Fig. 4B). Analyzing levels of L1 transcripts for the various subfamilies (Fig. 4C) further revealed that ancient, infrequently KB elements were lowly expressed at baseline and were not or were only moderately affected by knocking down the corepressor. In comparison, members of the highly KAP1-enriched L1PA4 and L1PA5 subfamilies were more strongly expressed in control cells and were significantly up-regulated in KAP1-depleted cells. Of note, the fold change in L1PA4 and L1PA5 expression levels between control and KAP1-depleted cells was not only statistically significant but also the strongest among all evaluated subfamilies. Finally, expression of youngest elements (L1PA2 and L1Hs) was highest at baseline and unchanged upon *KAP1* knockdown.

Youngest human L1 are up-regulated upon depletion of DNMTs

DNA methylation is involved in the long-term transcriptional control of EREs, including L1. Correspondingly, our analysis of MeDIP-seq data from the Epigenomics Mapping Consortium (Bernstein et al. 2010) indicated that DNA methylation is enriched at the 5′ region of mappable full-length L1 integrants in human H1 ES cells (Fig. 5A). It also revealed L1PA4 and L1PA5 as the most methylated and L1Hs as the least methylated subfamilies

Castro-Diaz et al.
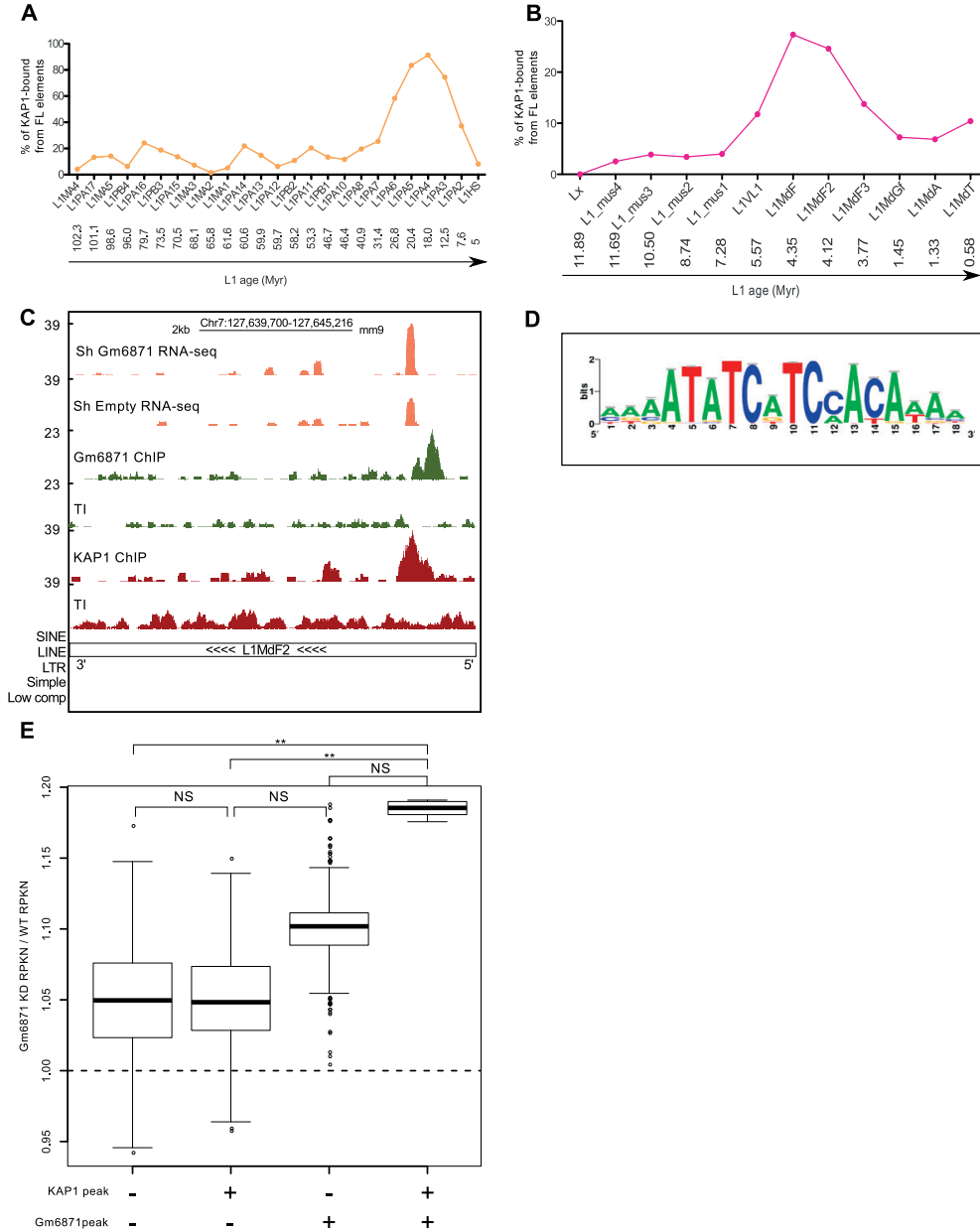
Figure 3. Evolutionarily dynamic and KRAB-ZFP-mediated KAP1–L1 interaction. Percentage of KB full-length (FL) L1 elements per subfamily in hES (*A*) and mES (*B*) cells, arranged from the oldest to the youngest subfamily using ages obtained from previously published divergence analysis studies (Khan 2005; Sookdeo et al. 2013). (Myr) Million years. (*C*) Screenshot of a representative L1MdF2 element, illustrating RNA-seq coverage plots from control (shEmpty) and Gm6871 knockdown mES cells as well as *gm6871* and *Kap1* ChIP-seq tracks. (*D*) Putative *gm6871* DNA-binding motif identified by computing *gm6871* ChIP-seq peaks with the RSAT software (Thomas-Chollier et al. 2012). (*E*) Relative change in the expression (RPKN [normalized reads per kilobase]) of murine full-length L1s bound or not bound by KAP1 and/or Gm6871 between Gm6871 knockdown and wild-type mES cells. The raw data were bootstrapped 1000 times with a resampling size of 100 for the plot design. The statistical analyses were calculated on the entire raw data by Wilcoxon nonparametric test. (NS) $P > 0.05$; (**) $P \leq 0.01$.

Figure 4. KAP1 depletion leads to up-regulation of KB L1 in hES cells. (*A*) Comparative expression of full-length L1 elements in hES cells transduced with control (shE) or *Kap1* knockdown (shKAP1) lentiviral vectors. (*B*) Relative change in the expression of full-length L1 elements bound or not by KAP1, comparing *Kap1* knockdown and control hES cells. (*C*) Comparative expression of full-length L1 in control versus Kap1 knock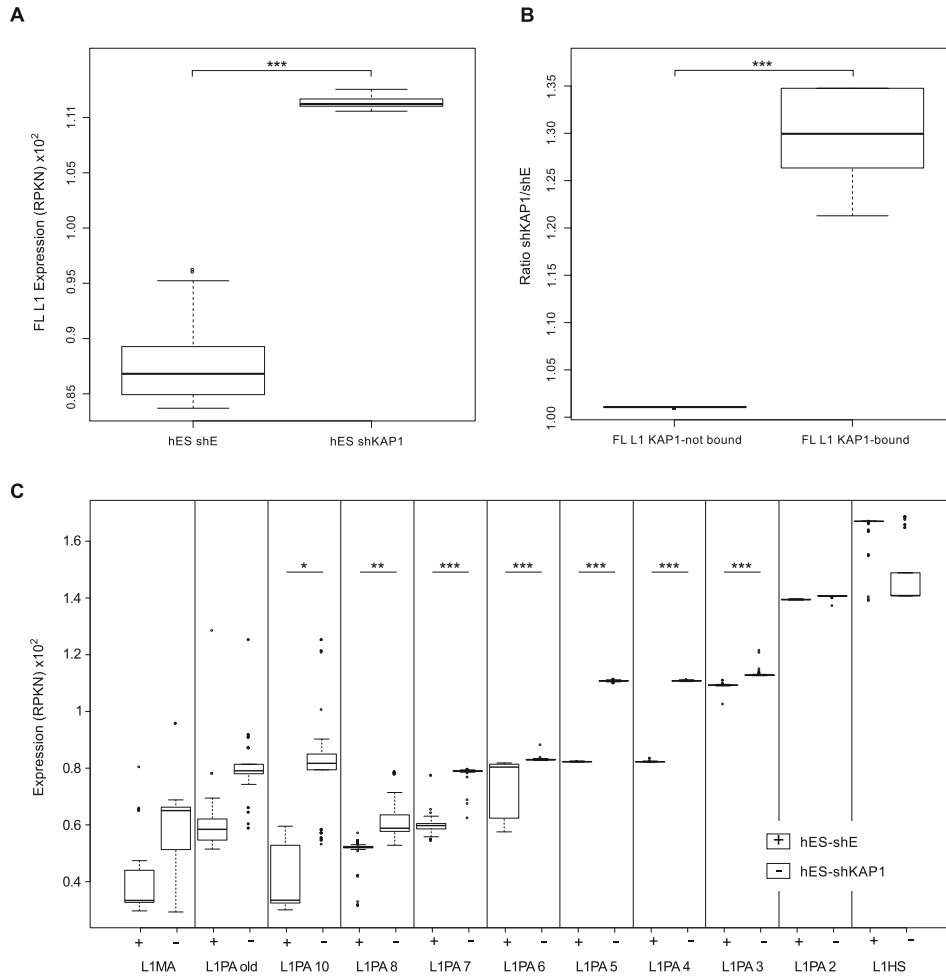down hES cells, examining each L1 subfamily separately, arranged from the oldest to the youngest one. The "L1MA" category corresponds to the families L1MA1 to L1MA9. The "L1PA old" category corresponds to the families L1PA11 to L1PA17. The raw data have been bootstrapped 1000 times for the plot design. Expression corresponds to RPKN values (see the Material and Methods), with *P*-values ([*] $P \leq 0.05$; [**] $P \leq 0.01$; [***] $P \leq 0.001$) calculated on the raw data by Wilcoxon nonparametric test.

(Fig. 5B). Therefore, we investigated the relative impact of KAP1-mediated and DNA methylation-mediated mechanisms in the control of L1. For this, we generated hES cell populations depleted for the de novo (DNMT3A and DNMT3B) and maintenance (DNMT1) DNMTs by lentivector-mediated RNAi. DNMT3A and DNMT3B could be stably knocked down for >22 d, whereas DNMT1 expression was partially recovered at that point, suggesting a growth disadvantage in the absence of this enzyme (data not shown). Still, cells in which all three DNMTs were strongly depleted (Supplemental Fig. S9A,B) could be readily obtained and kept in culture for the time of our

study, as reflected by their complete loss of DNA methylation at the GRB10 ICR after 5 or 9 d of triple knockdown (Supplemental Fig. S9C). Most interestingly, comparing the expression of individual L1 elements revealed that, in DNMT-depleted cells, it was the members of the youngest, KAP1-unbound L1 subfamilies (L1PA2 and L1Hs) that were the most up-regulated at that point, whereas older elements, whether KAP1-controlled or not, were not or were very modestly affected (Fig. 5C). In an attempt to explore further the interplay between KAP1-mediated and DNA methylation-mediated repression of L1, we separated L1PA4 and L1Hs family members in KB and KAP1-devoid
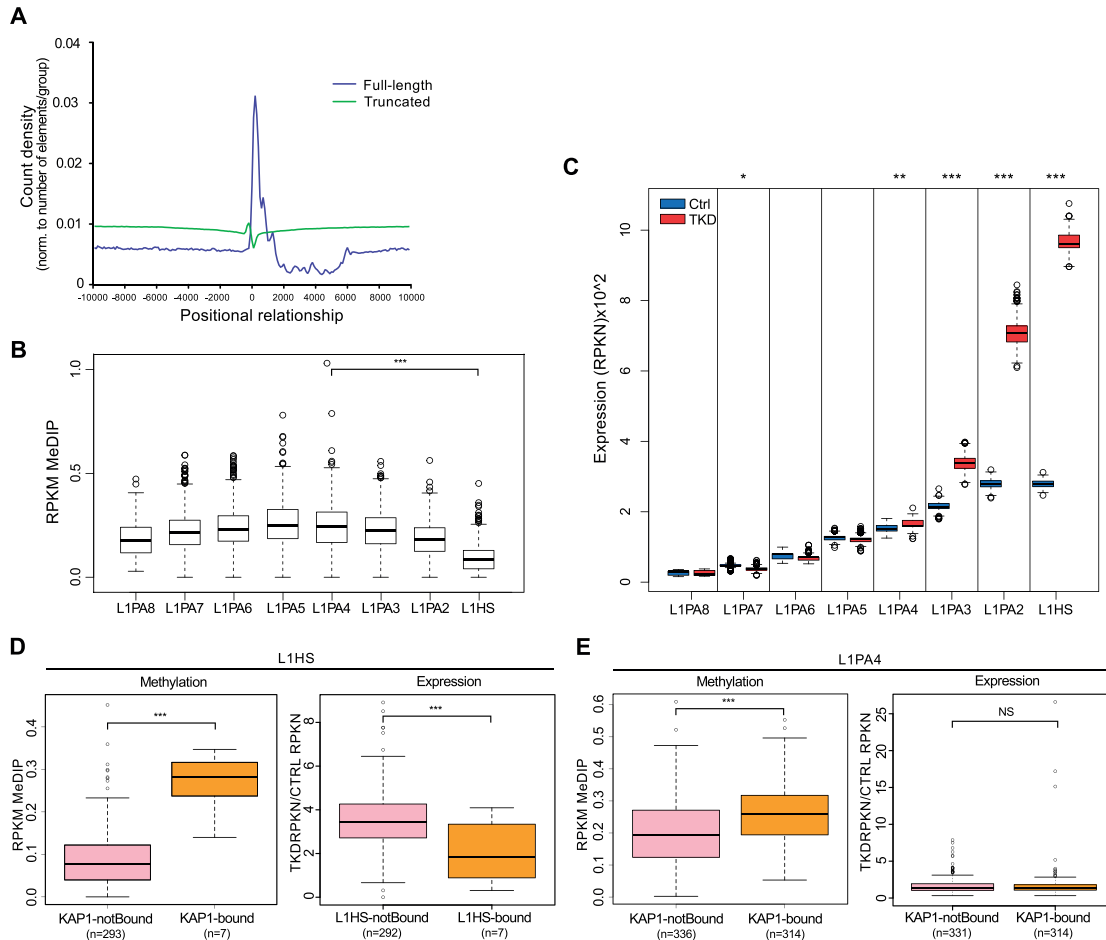
54

Castro-Diaz et al.

Figure 5. DNMT depletion induces up-regulation of younger L1 subfamilies. (*A*) Comparative expression of full-length L1 in control and DNMT knockdown (triple knockdown [TKD]) hES cells, analyzing each subfamily separately as described in Figure 4C. (*B*) Distribution of MeDIP-seq reads relative to the 5′ end of full-length (pink line) or truncated (yellow line) L1, normalizing profiles to the total number of elements per group. (*C*) DNA methylation levels on full-length L1 elements separated by subfamilies. (*D*) DNA methylation levels in full-length L1PA4 and L1Hs bound or not by KAP1 in hES cells, based on the numbers of MeDIP-seq reads per million base pairs per kilobase of L1 (RPKM). (*E*) Relative change in the expression of the same L1 elements, comparing triple knockdown and control hES cells. *P*-values ([NS] *P* > 0.05; [*] *P* ≤ 0.05; [**] *P* ≤ 0.01; [***] *P* ≤ 0.001) were calculated with Wilcoxon nonparametric test.

elements. Next, we looked at their methylation levels at baseline and at their expression upon *DNMT* triple knockdown. For L1Hs, we found that the KAP1-devoid elements, which were the overwhelming majority within this group, were significantly less methylated at baseline than their rare KB counterparts and that they alone were induced upon DNMT knockdown (Fig. 5D). Within the L1PA4 subfamily, baseline DNA methylation was globally higher and more homogeneous, with only slightly lower levels for KAP1-free members. Furthermore, expression of all L1PA4 elements was comparable in control and triple knockdown cells. Of note, depleting KAP1 in *DNMT* knockdown cells was highly toxic, precluding the

further exploration of potential synergies between the two L1 repression pathways.

Discussion

The transcriptional silencing of EREs is essential to protect genomic integrity, particularly during the vulnerable phases of developmental reprograming that occurs in ES and germ cells. Previous studies have revealed the roles of KRAB-ZFPs and their cofactor, KAP1, in the early embryonic repression of ERVs (Wolf and Goff 2009; Rowe et al. 2010; Tan et al. 2013), whereas small RNA-based mechanisms have been thought to prevail for the silencing of

55

L1 elements, as initially discovered in germ cells (Yang and Kazazian 2006; Aravin et al. 2007; Carmell et al. 2007; Beck et al. 2011). The present study actually establishes that L1 expression is also controlled by the KRAB–KAP1 system. Furthermore, our data, coupled with the recent demonstration that PIWI partakes in the regulation of L1Hs elements in human pluripotent cells (Marchetto et al. 2013), strongly support an evolutionary model in which the transcription of newly emerged L1 lineages is first repressed by small RNA-induced DNA methylation before KAP1-mediated silencing takes over through the selection of KRAB-ZFPs capable of tethering the master corepressor to their sequence (Fig. 6).

In both hES and mES cells, we found that KAP1 regulates L1 but that this control is restricted to lineages that have entered the corresponding ancestral genomes during the periods 31 million to 7.6 million years ago and 5.6 million to 3.8 million years ago, respectively. We identified a novel KRAB-ZFP responsible for tethering KAP1 to and controlling the expression of a subset of murine L1, strongly suggesting that these DNA-binding proteins are collectively involved in recognizing this class of retroelements, as previously observed for other EREs (Tan et al. 2013), and that in return, L1 has contributed to the species-specific diversification of the KRAB-ZFP gene family. However, we also determined that younger L1 lineages are generally not subjected to KRAB/KAP1-mediated regulation, whether in humans or mice. We found that the human-specific L1Hs, most of which neither recruit KAP1 nor are activated by KAP1 depletion, were instead induced upon depletion of DNMTs in hES cells. This observation fits well with the recent discovery that the PIWI-interacting RNAs (piRNAs)–PIWI system partakes in the early embryonic control of youngest L1 lineages in humans and apes (Marchetto et al. 2013). PIWI-mediated control, which was initially thought to be relevant only in germ cells, is indeed triggered by the recognition of L1-proximal sequences by a complex encompassing a member of the PIWI subclade of Argonaute proteins and L1-derived piRNAs, which leads to L1 transcriptional inhibition via DNA methylation (Aravin et al. 2007; Carmell et al. 2007; De Fazio et al. 2011). Whether other small RNA-based mechanisms reported to partake in the early embryonic control of L1 (Ciaudo et al. 2013; Fadloun et al. 2013; Heras et al. 2013) also act in a lineage-specific fashion remains to be determined.

Our finding that KAP1 binds a significant subset of L1s in ES cells but only exceptionally in HEK293T cells fits with the establishment of permanent silencing marks on EREs, including LINEs, during the early embryonic period. However, that it still is found on some L1 integrants in the differentiated cells suggests that particular L1s and their control mechanisms have been coopted to fulfil some roles in adult somatic tissues.

The presence of two KAP1-repressed DNA elements in a L1PA5-derived sequence (Fig. 2) and the weak effect of Gm6871 knockdown on L1 transcription raise the possibility of some redundancy in the KRAB-ZFP-mediated
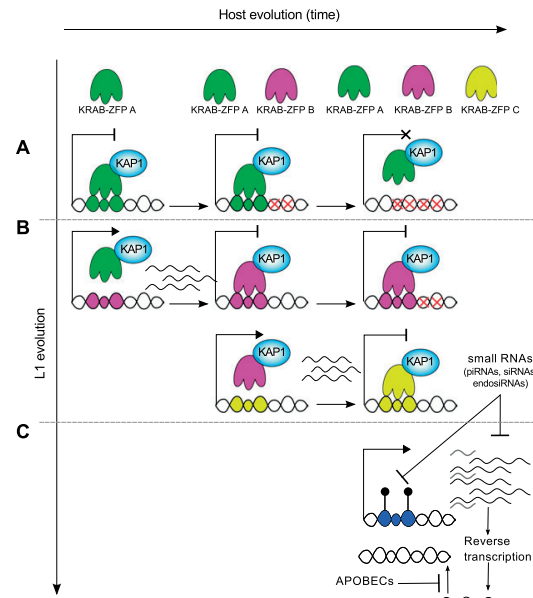

Figure 6. Model for the evolutionarily dynamic control of L1. (*A*) Very ancient L1s (shown in the *top* row) may have been once recognized by the KRAB/KAP1 system but have since then accumulated mutations (red crosses) abrogating binding by cognate KRAB-ZFPs but also transcription ability. (*B*) More recent subfamilies recruit KAP1 through sequence-specific KRAB-ZFPs but also may have some mutations taming their baseline expression. (*C*) The youngest L1 elements are highly transcribed and are not yet recognized by any KRAB-ZFP but produce small RNAs such as piRNAs, which in turn down-regulate their expression via DNA methylation and see their retrotransposition further blocked by proteins such as APOBEC family members.

control of L1s. Furthermore, although several mechanisms of L1 restriction have been described, their inactivation never results in spectacular up-regulation of these elements (nothing comparable, for instance, with the several hundred-fold induction undergone by some ERVs when KAP1 is deleted in mES cells) (Rowe et al. 2010). While this suggests that L1s are subjected to several layers of control, KAP1-restricted L1s belong to subfamilies more ancient and less active than human L1Hs and may have accumulated, over time, mutations that attenuate their transcriptional potential, dampening their up-regulation upon KAP1 removal. As for more ancient L1 lineages, their lack of KAP1 binding, coupled with their low level of baseline expression and inertia upon either KAP1 or DNMT depletion, is likely explained by the accumulation of inactivating mutations, alleviating the need for any sort of transcriptional control.

That KAP1-regulated elements are unaffected by depleting DNMTs, whereas the KAP1 recruitment at ERVs ultimately leads to their DNA methylation, is not surprising. The KRAB/KAP1 system indeed represses transcription of EREs primarily via histone deacetylation,

56

H3K9 trimethylation, and HP1 recruitment, with DNA methylation occurring only secondarily to ensure the permanence of the silencing process (Quenneville et al. 2012; Rowe et al. 2013a). Our observation that KB murine L1 elements are up-regulated upon *Setdb1* knockout in mES cells confirms the primary importance of histone methylation-based mechanisms in their control. Some rare L1Hs were found to be KB and accordingly were barely induced upon *DNMT* knockdown, in contrast to their far more prevalent KAP1-devoid counterparts. For L1PA4 subfamily members, we did not see any induction in *DNMT* triple knockdown cells whether they bore KAP1 or not. However, these elements were globally highly methylated, which may explain their resistance to the *DNMT* knockdown. It could also be that other epigenetic modifications, some of which may be lasting consequences of earlier KAP1 recruitment, partake in their repression.

Retrotransposons are mutagenic yet harbor *cis*-acting activities, many of which contribute to shaping transcriptional networks, including in ES cells (Bourque et al. 2008; Kunarso et al. 2010; Jacques et al. 2013; Rowe et al. 2013b; Ward et al. 2013). They thus have both a detrimental and an evolutionarily beneficial potential, which requires that they be very delicately controlled. We propose that this is accomplished, at least for the youngest, most active L1 elements, via autoregulation of piRNA production, a repression mechanism that is in part self-imposed. For elements that escape this process, additional restrictions are exerted at the post-transcriptional level, for instance, through lethal editing of reverse transcripts by the APOBEC3B cytidine deaminase (Bogerd et al. 2006; Chiu and Greene 2008; Wissing et al. 2011; Marchetto et al. 2013). After some time, KAP1-induced restriction, which appears more stringent, takes over through the selection of L1-recognizing KRAB-ZFPs. It will be interesting to ask whether a similar level of complexity prevails to the control of this class of retroelements in germ cells, where the reprogramming of epigenetic marks opens another window for their activation.

Materials and methods

Plasmids and lentiviral vectors

pLKO.1.puro shRNA vectors were used for *KAP1*, *Gm6871*, and *DNMT1* knockdown. shRNAs against *DNMT3A* and *DNMT3B* were cloned into the pLVTHM vector, which was further modified to express neomycin, hygromycin, or blasticidine resistance genes instead of GFP. For each shRNA vector, an empty version (without shRNA) was cloned as a control. The shRNA targeting sequences were obtained through the RNAi Consortium (http://www.broadinstitute.org/rnai/public) and are listed in the Supplemental Material (Supplemental Table 2). L1 *cis*-acting sequences (see "Genomic Coordinates" in the Supplemental Material) were cloned into the pENTR/D/TOPO vector an then into an in-house cloned gateway destination vector by LR recombination (pRRL.R1-R2.PGK.GFP). Codon-optimized *Gm6871* was synthesized and introduced by Gateway cloning in a puromycin selectable lentivector under a tetracyclin-inducible TRE promoter to obtain an HA-tagged protein (pSIN-TRE-Gm6871-3xHA, Addgene). LV production protocols are detailed at http:// tronolab.epfl.ch. LV backbones are available at Addgene (http://www.addgene.org).

ES cell culture and transduction

The H1 hES cell line (WA01, WiCell) was cultured in mTeSR1 medium (Stem Cell Technologies) on hES-qualified Matrigel (BD Biosciences) and in the presence of ROCK inhibitor (Y-27632). mES (ES3 and J1) cell lines were cultured as previously described (Rowe et al. 2013b). J1 cells culture was further supplemented with 1 μM PD0325901 and 3 μM CHIR99021. mES cells were grown on 0.1% gelatin-coated (48723-500G-F, Sigma) plates. Transductions were done at a multiplicity of infection (MOI; determined in HCT116 or 3T3 cells) of 0.25–50. Whenever required, cells were selected with 100 μg/mL hygromycin, 10 μg/mL blasticidin, 0.25 μg/mL or 1.0 μg/mL puromycin, or 200 μg/mL neomycin. Pluripotency was monitored by FACS using a human pluripotent stem cell transcription factor analysis kit (BD Biosciences) or mouse anti SSEA-1 PE-conjugated antibody (560142, BD Pharmingen).

RT-qPCR and RNA-seq

Total RNA was extracted and DNAse I-treated using a spin column-based RNA purification kit (Macherey Nagel). cDNA was synthesized starting from 500 ng of RNA and using random hexamers and SuperScript II (Invitrogen). Primers (Supplemental Table 1) were used for SYBR Green qPCR (Applied Biosystems), and their specificity was confirmed with dissociation curves. RT-qPCR reactions were performed in triplicate for each cDNA sample. hES RNA-seq was generated with RNA extracted 14 d after KAP1 depletion (Turelli et al. 2014) or 9 d after triple *DNMT* depletion. mES RNA-seq was done in ES3 cells 4 d after the sh-*Gm6871* knockdown vector transduction (MOI50), in duplicate (independent transductions). Knockdown levels were of 0.87 and 0.88 by qPCR. The 76- or 100-bp single-end reads from the Illumina HiSeq sequencing instrument were mapped using the Bowtie short read aligner (Langmead et al. 2009) to the annotated sequence of individual full-length L1 (minimum 5 kb in length) (lists provided in the Supplemental Material). The annotation and genomic coordinates of full-length L1 elements were obtained from the University of California at Santa Cruz genome browser. Reads mapping to multiple locations were evenly distributed across those locations, and a maximum of three mismatches was allowed. The RPKN (normalized reads per kilobase) values were calculated using an in-house R program and correspond to the read counts normalized to the length of the repeated element and to the total number of reads mapped to the transcriptome.

ChIP-seq and ChIP-qPCR

Chromatin was prepared from 1×10^7 H1 hES or J1 mES cells (for KAP1 ChIPs) and from 2×10^7 ES3 cells (for Gm6871 ChIP) as previously described (Barde et al. 2013; Rowe et al. 2013b; Turelli et al. 2014) with KAP1-specific (Tronolab, SY326768 or ab10483, Abcam), H3K9me3-specific (Diagenode), or HA-specific (Covance; MMS-101P) antibodies. For sequencing, total input (TI) and ChIP library preparation was performed as described in Santoni de Sio et al. (2012) using between 2 and 10 ng of chromatin. Sequencing was performed on an Illumina genome analyzer IIx, with each library sequenced in 80-base single-read or 100-bp reads run. The 80- to 100-bp single-end or paired-end reads generated were mapped to the human genome assembly hg19 or mouse genome assembly mm9 using the Bowtie short read

57

aligner (Langmead et al. 2009), allowing up to two to three mismatches, and all multiple matches were discarded. The peaks were called using the MACS program (Zhang et al. 2008) and were normalized to the TI. When defining KB and NKB L1 sequences, only KAP1 peaks with a MACS score [Log$_{10}$(pval)] >100 were considered. ChIP-seq data in HEK293 cells (Iyengar et al. 2011) were obtained from the ENCODE database (https://genome.ucsc.edu/ENCODE). H3K9me3 ChIP-seq data in mES (ES3) cells was previously published in Rowe et al. (2013b). Motif search was performed with RSAT (Thomas-Chollier et al. 2012) using Gm6871 called peaks as input and unbound repeated regions as background control. Correlation analysis between ChIP-seq peaks, MeDIP-seq tags, and L1 elements was done using the ChIP-cor analysis module (http://ccg.vital-it.ch/chipseq/chip_cor.php).

FACS analysis

Cells were analyzed on a FACScan machine (Becton Dickinson). Analysis was performed with FlowJo software (version 8-1.8.6, Treestar, Inc.).

DNA methylation

For quantitative bisulfite pyrosequencing, genomic DNA was converted (1–2 µg per sample) using an Epitect bisulfite kit (591014, Qiagen) and used for PCR (primers were designed on the converted antisense and sense strand, respectively, using PyroMark Assay Design 2.0 software). Purity of PCR products was verified on agarose gels for each experiment before immobilizing on 96-well plates using a vacuum prep workstation and pyrosequencing using PyroMark gold reagents (972804, Qiagen; Center for Integrative Genomics, University of Lausanne, Switzerland). Results were analyzed using Pyro Q-CpG software. Primer sequences are in Supplemental Table 1. MeDIP-seq data sets (Hs1376 and Hs1303) were downloaded from http://www.genboree.org/epigenomeatlas. COBRA methylation analysis was performed using primers for the GRB10 human ICR (see the Supplemental Material) and as previously described (Xiong and Laird 1997).

Immunoblotting

Cells were washed with ice-cold PBS and resuspended in radio-immunoprecipitation (RIPA) buffer to prepare total cell extracts. Protein amount was quantified by BCA protein assay reagents (Pierce) and normalized for loading on a 10% denaturing SDS–polyacrylamide gel. Wet transfer was performed, and the primary antibodies used were anti-DNMT1 (rabbit pAb; ab87654, Abcam), anti-DNMT3A (mouse mAb; ab13888, Abcam), anti-DNMT3B (rabbit pAb; ab2851, Abcam), and β-tubulin (rabbit pAb; ab21058, Abcam).

Immunofluorescence

mES cells were transduced with Gm6871-HA, ZFP809-HA, or LacZ-HA and cultured with 5 µg/mL doxycycline. Cells were fixed in methanol for 10 min and labeled with anti-HA antibody (MMS-101P, Covance) followed by Alexa488-conjugated anti-mouse antibody. Nuclei were stained with DAPI. Images were acquired using a 63× lens on a Zeiss Axiovert 200M microscope.

Coimmunoprecipitation and GST pull-down

Gm6871-HA, ZFP809-HA, KRAB-deleted ZFP809, or LacZ-HA plasmids were used to transduce mES cells or transfect 293T cells. Cells were cultured with 5 µg/mL doxycycline for at least 48 h, harvested, and lysed with lysis buffer (50 mM Tris HCl at pH 8, 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycolate) supplemented with protease inhibitors under constant agitation for 30 min. Lysate was sonicated twice for 10 sec at 30% duty cycle. Immunoprecipitation was performed overnight with HA antibody (MMS-101P, Covance) in immunoprecipitation buffer (50 mM Tris HCl at pH 8, 150 mM NaCl, 5 mM EDTA, 0.1% NP-40) supplemented with protease inhibitors. All steps were performed at 4°C. Immunoblotting was performed with either anti-KAP1 antibody (ab10483, Abcam) followed by HRP-conjugated anti-rabbit antibody or HRP-conjugated anti-HA antibody (12013819001, Roche). Ex vivo GST pull-down assay was performed as previously described (Yahi et al. 2008).

Accession numbers

RNA-seq and ChIP-seq data were deposited in the Gene Expression Omnibus database at the NCBI under the accession numbers GSE57989 (Turelli et al. 2014) and GSE58323.

References

Aravin AA, Hannon GJ, Brennecke J. 2007. The Piwi–piRNA pathway provides an adaptive defense in the transposon arms race. *Science* **318:** 761–764.

Babushok DV, Kazazian HH. 2007. Progress in understanding the biology of the human mutagen LINE-1. *Hum Mutat* **28:** 527–539.

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

Beck CR, Garcia-Perez JL, Badge RM, Moran JV. 2011. LINE-1 elements in structural variation and disease. *Annu Rev Genomics Hum Genet* **12:** 187–215.

Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, Kellis M, Marra MA, Beaudet AL, Ecker JR, et al. 2010. The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol* **28:** 1045–1048.

Bogerd HP, Wiegand HL, Hulme AE, Garcia-Perez JL, O'Shea KS, Moran JV, Cullen BR. 2006. Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. *Proc Natl Acad Sci* **103:** 8780–8785.

Bourque G, Leong B, Vega VB, Chen X, Lee YL, Srinivasan KG, Chew JL, Ruan Y, Wei CL, Ng HH, et al. 2008. Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Res* **18:** 1752–1762.

58

Brouha B, Schustak J, Badge RM, Lutz-Prigge S, Farley AH, Moran JV, Kazazian HH Jr. 2003. Hot L1s account for the bulk of retrotransposition in the human population. *Proc Natl Acad Sci* **100:** 5280–5285.

Carmell MA, Girard A, van de Kant HJ, Bourc'his D, Bestor TH, de Rooij DG, Hannon GJ. 2007. MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Dev Cell* **12:** 503–514.

Chiu YL, Greene WC. 2008. The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements. *Annu Rev Immunol* **26:** 317–353.

Ciaudo C, Jay F, Okamoto I, Chen CJ, Sarazin A, Servant N, Barillot E, Heard E, Voinnet O. 2013. RNAi-dependent and independent control of LINE1 accumulation and mobility in mouse embryonic stem cells. *PLoS Genet* **9:** e1003791.

Cordaux R, Batzer MA. 2009. The impact of retrotransposons on human genome evolution. *Nat Rev Genet* **10:** 691–703.

Corsinotti A, Kapopoulou A, Gubelmann C, Imbeault M, Santoni de Sio FR, Rowe HM, Mouscaz Y, Deplancke B, Trono D. 2013. Global and stage specific patterns of Kruppel-associated-box zinc finger protein gene expression in murine early embryonic cells. *PLoS ONE* **8:** e56721.

De Fazio S, Bartonicek N, Di Giacomo M, Abreu-Goodger C, Sankar A, Funaya C, Antony C, Moreira PN, Enright AJ, O'Carroll D. 2011. The endonuclease activity of Mili fuels piRNA amplification that silences LINE1 elements. *Nature* **480:** 259–263.

Dewannieux M, Esnault C, Heidmann T. 2003. LINE-mediated retrotransposition of marked Alu sequences. *Nat Genet* **35:** 41–48.

Fadloun A, Le Gras S, Jost B, Ziegler-Birling C, Takahashi H, Gorab E, Carninci P, Torres-Padilla M-E. 2013. Chromatin signatures and retrotransposon profiling in mouse embryos reveal regulation of LINE-1 by RNA. *Nat Struct Mol Biol* **20:** 332–338.

Faulkner GJ, Kimura Y, Daub CO, Wani S, Plessy C, Irvine KM, Schroder K, Cloonan N, Steptoe AL, Lassmann T, et al. 2009. The regulated retrotransposon transcriptome of mammalian cells. *Nat Genet* **41:** 563–571.

Finnegan DJ. 2012. Retrotransposons. *Curr Biol* **22:** R432–R437.

Furano AV, Boissinot S. 2008. Long interspersed nuclear elements (LINEs): evolution. *eLS* doi: 10.1002/9780470015902.a0005304.pub2.

Heras SR, Macias S, Plass M, Fernandez N, Cano D, Eyras E, Garcia-Perez JL, Cáceres JF. 2013. The Microprocessor controls the activity of mammalian retrotransposons. *Nat Struct Mol Biol* **20:** 1173–1181.

Ivanov AV, Peng H, Yurchenko V, Yap KL, Negorev DG, Schultz DC, Psulkowski E, Fredericks WJ, White DE, Maul GG, 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.

Iyengar S, Farnham PJ. 2011. KAP1 protein: an enigmatic master regulator of the genome. *J Biol Chem* **286:** 26267–26276.

Iyengar S, Ivanov AV, Jin VX, Rauscher FJ 3rd, Farnham PJ. 2011. Functional analysis of KAP1 genomic recruitment. *Mol Cell Biol* **31:** 1833–1847.

Jacques P-É, Jeyakani J, Bourque G. 2013. The majority of primate-specific regulatory sequences are derived from transposable elements. *PLoS Genet* **9:** e1003504.

Khan H. 2005. Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. *Genome Res* **16:** 78–87.

Kunarso G, Chia NY, Jeyakani J, Hwang C, Lu X, Chan YS, Ng HH, Bourque G. 2010. Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat Genet* **42:** 631–634.

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

Marchetto MCN, Narvaiza I, Denli AM, Benner C, Lazzarini TA, Nathanson JL, Paquola ACM, Desai KN, Herai RH, Weitzman MD, et al. 2013. Differential L1 regulation in pluripotent stem cells of humans and apes. *Nature* **503:** 525–529.

Matlik K, Redik K, Speek M. 2006. L1 antisense promoter drives tissue-specific transcription of human genes. *J Biomed Biotechnol* **2006:** 71753.

Matsui T, Leung D, Miyashita H, Maksakova IA, Miyachi H, Kimura H, Tachibana M, Lorincz MC, Shinkai Y. 2010. Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature* **464:** 927–931.

Nigumann P, Redik K, Matlik K, Speek M. 2002. Many human genes are transcribed from the antisense promoter of L1 retrotransposon. *Genomics* **79:** 628–634.

Quenneville S, Verde G, Corsinotti A, Kapopoulou A, Jakobsson J, Offner S, Baglivo I, Pedone PV, Grimaldi G, Riccio A, et al. 2011. In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. *Mol Cell* **44:** 361–372.

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

Rosser JM, An W. 2012. L1 expression and regulation in humans and rodents. in *Front Biosci (Elite Ed)* **4:** 2203–2225.

Rowe HM, Trono D. 2011. Dynamic control of endogenous retroviruses during development. *Virology* **411:** 273–287.

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

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

Rowe HM, Kapopoulou A, Corsinotti A, Fasching L, Macfarlan TS, Tarabay Y, Viville S, Jakobsson J, Pfaff SL, Trono D. 2013b. TRIM28 repression of retrotransposon-based enhancers is necessary to preserve transcriptional dynamics in embryonic stem cells. *Genome Res* **23:** 452–461.

Santoni de Sio FR, Barde I, Offner S, Kapopoulou A, Corsinotti A, Bojkowska K, Genolet R, Thomas JH, Luescher IF, Pinschewer D et al. 2012. KAP1 regulates gene networks controlling T-cell development and responsiveness. *FASEB J* **26:** 4561–4575.

Schultz DC, Ayyanathan K, Negorev D, Maul GG, Rauscher FJ. 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.

Slotkin RK, Martienssen R. 2007. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet* **8:** 272–285.

Smit AF, Tóth G, Riggs AD, Jurka J. 1995. Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences. *J Mol Biol* **246:** 401–417.

59

Sookdeo A, Hepp CM, McClure MA, Boissinot S. 2013. Revisiting the evolution of mouse LINE-1 in the genomic era. *Mob DNA* **4:** 3.

Speek M. 2001. Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. *Mol Cell Biol* **21:** 1973–1985.

Tan X, Xu X, Elkenani M, Smorag L, Zechner U, Nolte J, Engel W, Pantakani DV. 2013. Zfp819, a novel KRAB-zinc finger protein, interacts with KAP1 and functions in genomic integrity maintenance of mouse embryonic stem cells. *Stem Cell Res* **11:** 1045–1059.

Thomas-Chollier M, Darbo E, Herrmann C, Defrance M, Thieffry D, van Helden J. 2012. A complete workflow for the analysis of full-size ChIP-seq (and similar) data sets using peak-motifs. *Nat Protoc* **7:** 1551–1568.

Turelli P, Castro-Diaz N, Marzetta F, Kapopoulou A, Raclot C, Duc J, Tieng V, Quenneville S, Trono D. 2014. Interplay of TRIM28 and DNA methylation in controlling human endogenous retroelements. *Genome Res* doi: 10.1101/gr.172833.114.

Ward MC, Wilson MD, Barbosa-Morais NL, Schmidt D, Stark R, Pan Q, Schwalie PC, Menon S, Lukk M, Watt S, et al. 2013. Latent regulatory potential of human-specific repetitive elements. *Mol Cell* **49:** 262–272.

Wissing S, Montano M, Garcia-Perez JL, Moran JV, Greene WC. 2011. Endogenous APOBEC3B restricts LINE-1 retrotransposition in transformed cells and human embryonic stem cells. *J Biol Chem* **286:** 36427–36437.

Wolf D, Goff SP. 2007. TRIM28 mediates primer binding site-targeted silencing of murine leukemia virus in embryonic cells. *Cell* **131:** 46–57.

Wolf D, Goff SP. 2009. Embryonic stem cells use ZFP809 to silence retroviral DNAs. *Nature* **458:** 1201–1204.

Xiong Z, Laird PW. 1997. COBRA: a sensitive and quantitative DNA methylation assay. *Nucleic Acids Res* **25:** 2532–2534.

Yahi H, Fritsch L, Philipot O, Guasconi V, Souidi M, Robin P, Polesskaya A, Losson R, Harel-Bellan A, Ait-Si-Ali S. 2008. Differential cooperation between heterochromatin protein HP1 isoforms and MyoD in myoblasts. *J Biol Chem* **283:** 23692–23700.

Yang N, Kazazian HH Jr. 2006. L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. *Nat Struct Mol Biol* **13:** 763–771.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. 2008. Model-based analysis of ChIP-seq (MACS). *Genome Biol* **9:** R137.

60

Evolutionally dynamic L1 regulation in embryonic stem cells

Nathaly Castro-Diaz, Gabriela Ecco, Andrea Coluccio, et al.

Supplemental Material	http://genesdev.cshlp.org/content/suppl/2014/06/11/gad.241661.114.DC1.html
References	This article cites 53 articles, 15 of which can be accessed free at: http://genesdev.cshlp.org/content/28/13/1397.full.html#ref-list-1
	Articles cited in: http://genesdev.cshlp.org/content/28/13/1397.full.html#related-urls
Related Content	**LINEing germ and embryonic stem cells silencing of retrotransposons** Takashi Ishiuchi and Maria-Elena Torres-Padilla Genes Dev. July 1, 2014 28: 1381-1383 **piRNA pathway targets active LINE1 elements to establish the repressive H3K9me3 mark in germ cells** Dubravka Pezic, Sergei A. Manakov, Ravi Sachidanandam, et al. Genes Dev. July 1, 2014 28: 1410-1428
Creative Commons License	This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see http://genesdev.cshlp.org/site/misc/terms.xhtml). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/.
Email Alerting Service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**
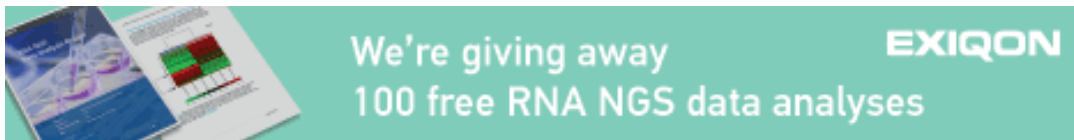
To subscribe to *Genes & Development* go to:
http://genesdev.cshlp.org/subscriptions

IV. Results – Part III: ZFP932/Gm15446 control ERVKs and neighboring genes expression in pluripotent and differentiated cells

Manuscript title

Transposable elements and their KRAB-ZFP controllers regulate gene expression in adult tissues

Authors

Gabriela Ecco, Marco Cassano, Annamaria Kauzlaric, Julien Duc, Andrea Coluccio, Sandra Offner, Michaël Imbeault, Helen M. Rowe, Priscilla Turelli, and Didier Trono

Summary of results and contribution

Another candidate that we identified with our screen was ZFP932, which was binding to two DNA sequences, both containing ERVK elements. We also found that this factor had a paralog, Gm15446. Combining genome-wide expression and KRAB-ZFPs binding profiles, we showed that ZFP932/Gm15446 regulate ERVKs and neighboring genes in murine ES cells. Interestingly, these KRAB-ZFP paralogs bind partially overlapping but still distinct subsets of ERVKs, at their

3'end. Since these and many other KRAB-ZFPs are not only expressed in ES but also in a wide range of tissues, we then asked if the KRAB/KAP1 complex could repress TEs in differentiated cells. In KAP1- and KRAB-ZFPs-depleted cells, binding profiles and chromatin marks analysis confirmed that these repressors regulate TEs and genes in differentiated cells and *in vivo*. Our results, thus, establish that the KRAB/KAP1 complex regulates TEs and neighboring genes in differentiated cells. Additionally, they strongly suggest that TEs and KRAB-ZFPs participate in transcriptional networks that likely regulate not only development but also many physiological events in vertebrates.

I was the leading scientist in this project, and my contribution to this manuscript comprises its development from the elaboration of the idea, to the execution/data analysis, and the writing of the final text, with supervision of Dr. Priscilla Turelli and Prof. Didier Trono. Over its course, I had significant help from the other co-authors at different steps (especially regarding *in vivo* experiments, generation of high-throughput data in differentiated cells, and data analysis).

Transposable elements and their KRAB-ZFP controllers regulate gene expression in adult tissues

Manuscript accepted for publication in Developmental Cell

Gabriela Ecco, Marco Cassano, Annamaria Kauzlaric, Julien Duc, Andrea Coluccio, Sandra Offner, Michaël Imbeault, Helen M. Rowe[1], Priscilla Turelli, and Didier Trono[*]

School of Life Sciences, École Polytechnique Fédérale de Lausanne (EPFL), Station19, 1015 Lausanne, Switzerland

[*] To whom correspondence should be addressed: e-mail: didier.trono@epfl.ch

[1]Present address: Centre for Medical Molecular Virology, Division of Infection and Immunity, University College London, 90 Gower Street, London WC1E 6BT, UK.
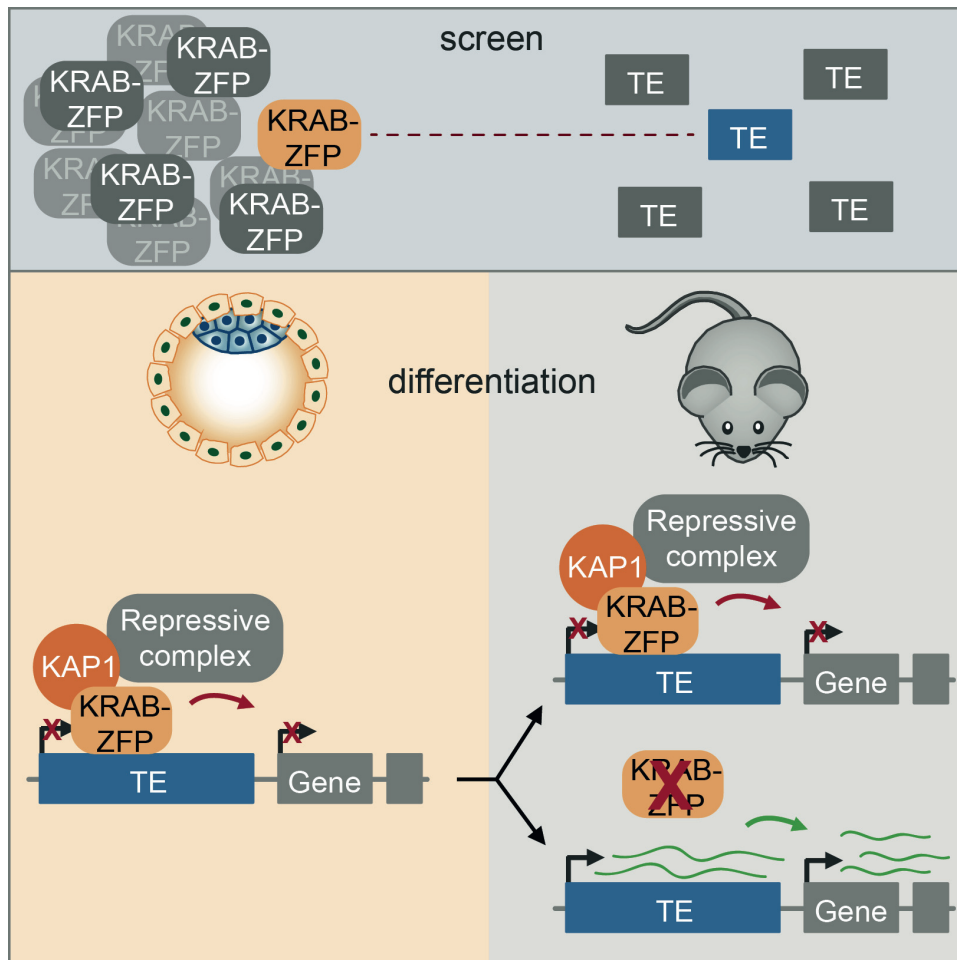
KRAB-ZFPs and KAP1 are embryonic controllers of transposable elements (TE), and were thought to be dispensable for this process in adult tissues. Ecco et al. demonstrate instead that these modulators control TE expression in somatic cells as well, including *in vivo*, where they partner up to regulate cellular genes.

HIGHLIGHTS

- Development of a large-scale functional screen matching KRAB-ZFPs to DNA targets
- ZFP932 and its paralog Gm15446 regulate different subsets of ERVKs, by 3'end binding
- KRAB-ZFPs/KAP1 regulate TEs in somatic cells via histone modifications
- KRAB-ZFPs/KAP1 use TE-residing platforms to regulate gene expression in adult tissues

SUMMARY

KRAB-containing zinc finger proteins (KRAB-ZFPs) are early embryonic controllers of transposable elements (TEs), which they repress with their cofactor KAP1 through histone and DNA methylation, a process thought to result in irreversible silencing. Using a target-centered functional screen, we matched murine TEs with their cognate KRAB-ZFP. We found the paralogs ZFP932 and Gm15446 to bind overlapping but distinguishable subsets of ERVK (endogenous retrovirus K), repress these elements in embryonic stem cells, and regulate secondarily the expression of neighboring genes. Most importantly, we uncovered that these KRAB-ZFPs and KAP1 control TEs in adult tissues, in cell culture and *in vivo*, where they partner up to modulate cellular genes. Therefore, TEs and KRAB-ZFPs establish transcriptional networks that likely regulate not only development but also many physiological events. Given the high degree of species-specificity of TEs and KRAB-ZFPs, these results have important implications for understanding the biology of higher vertebrates, including humans.

INTRODUCTION

Transposable elements (TEs) account for more than half of the human and murine genomes (Lander et al., 2001; Waterston et al., 2002). Long considered as purely parasitic, they are now recognized as important motors of evolution, yet they also represent genomic threats requiring control from the earliest stages of development. Whether they are DNA transposons or retrotransposons – endogenous retroviruses (ERVS), LINEs SINEs and SVAs, reviewed in Friedli and Trono, 2015a)–, TEs can disrupt genes, alter their transcription, or serve as ground for recombination, and have been implicated in diseases such as cancer and diabetes (Hancks and Kazazian, 2012; Jern and Coffin, 2008). However, growing evidence indicates that TEs can be co-opted for the benefit of the host, with for instance expression of zygotic activation genes driven from the LTR (long terminal repeat) of MERVL (murine endogenous retrovirus L) in the mouse, and many binding sites for pluripotency factors residing within mobile DNA elements in the human genome (Bourque et al., 2008; Chuong, 2013; Dupressoir et al., 2012; Fort et al., 2014; Macfarlan et al., 2012).

TEs are repressed through RNA- and protein-based epigenetic mechanisms instated during the first days of embryogenesis. KRAB-containing zinc finger proteins (KRAB-ZFPs) constitute a large family of transcription factors implicated in this process. KRAB-ZFPs bind to specific DNA sequences through an array of zinc fingers, and recruit their cofactor KAP1, which serves as a scaffold for a heterochromatin-inducing complex encompassing histone methyltransferase, histone deacetylase, nucleosome remodelling and DNA methyltransferase activities (reviewed in Rowe and Trono, 2011). Depletion of KAP1 or its partner histone methyltransferase SETDB1 in murine or human embryonic stem (ES) cells activates the expression of endogenous retroelements (EREs) (Matsui et al., 2010; Rowe et al., 2010; Turelli et al., 2014). This impacts expression of nearby genes, as KAP1 and associated effectors control TE-originating promoter or enhancer effects (Rebollo et al., 2012b; Rowe et al., 2013b; Wolf et al., 2015). Furthermore, a few individual KRAB-ZFPs have been confirmed to repress retroelements in pluripotent cells, such as ZFP809 for murine leukemia virus (MLV) and its endogenous relatives (Wolf and Goff, 2007,

2009 ; Wolf et al., 2015), or Gm6871 and ZNF93 for mouse and human LINEs, respectively (Castro-Diaz et al., 2014; Jacobs et al., 2014). Although recent findings indicate that many KRAB-ZFPs have EREs as their preferential genomic targets (Najafabadi et al., 2015) (and our unpublished results), detailed functional data are missing about most members of the family.

Encoded in the hundreds by the genomes of higher vertebrates, KRAB-ZFPs first emerged in early tetrapods some 350 million years ago, and the expansion of this gene family subsequently mirrored waves of retroviral invasion into the genomes of these species (Thomas and Schneider, 2011). During this process, *Krab-zfp* genes underwent strong positive selection at positions encoding for amino acids predicted to determine DNA binding specificity, consistent with a role in countering rapidly mutating genetic invaders (Emerson and Thomas, 2009). Furthermore, the study of a couple of KRAB-ZFP/TE target pairs suggests the parallel evolution of restriction factors and TE mutants escaping their inhibition (Jacobs et al., 2014). This led to the suggestion of a host-invader arms race, where KRAB-ZFPs were primarily selected to silence TEs.

In ES cells, the KRAB-ZFP-mediated docking of KAP1 and associated epigenetic modifiers at TEs triggers the deposition of repressive chromatin marks such as trimethylation of histone 3 on lysine 9 (H3K9me3), and methylation of CpG dinucleotides by *de novo* DNA methyltransferases (Rowe and Trono, 2011). Once established, DNA methylation is perpetuated across cell divisions, and its establishment at TEs during early embryogenesis is thought to result in permanent silencing, without need for persistent expression of their cognate KRAB-ZFP repressors (Quenneville et al., 2012; Wolf et al., 2015). Cumulated evidence indicates that DNA methylation is indeed an important mechanism of TE control in somatic tissues (Hutnick et al., 2010; Jackson-Grusby et al., 2001; Rowe et al., 2013a; Walsh et al., 1998). However, we previously observed that a significant fraction of TEs bound by KAP1 in human ES cells still carries the corepressor in mature T lymphocytes (Turelli et al., 2014), and that *Kap1* deletion in neuronal progenitors activates some EREs (Fasching et al., 2015). A few ERVs are similarly induced in murine B-lymphocytes and fibroblasts deleted for *Setdb1*, the histone methyltransferase associated with KAP1 (Collins et al.,

2015; Wolf et al., 2015). Moreover, many KRAB-ZFPs are expressed not only in ES cells but also in a variety of tissues (Barde et al., 2013; Corsinotti et al., 2013; Lizio et al., 2015).

In order to investigate KRAB-ZFPs/TE interactions, we developed a functional screen to identify KRAB-ZFPs responsible for recognizing specific DNA sequences. This led us to characterize two members of the family, ZFP932 and its paralog Gm15446, which we found to regulate distinct subsets of endogenous retrovirus-K (ERVK) in murine ES cells. Invalidating current models (Maksakova et al., 2008; Mikkelsen et al., 2007; Rowe and Trono, 2011; Wolf et al., 2015), we further determined that these two KRAB-ZFPs also regulate their TE targets in differentiated tissues, through histone-based mechanisms not always correlated with the DNA methylation status of these loci. Furthermore, the dynamic control of these TEs by their KRAB-ZFP repressors modulated the expression of cellular genes in several adult tissues examined, both in cell culture and *in vivo*. We conclude that TEs and their KRAB-ZFP controllers are broad regulators of cellular gene expression, likely engaged in influencing multiple aspects of the biology of higher species.

RESULTS

A functional screen identifies KRAB-ZFPs repressing specific DNA targets

In order to identify KRAB-ZFPs responsible for regulating specific TEs, we developed a large-scale functional screen based on the repression of a reporter cassette through the intermediate of nearby DNA baits (Figure 1A). Putative KRAB-ZFP-binding sequences were selected based on the coincidence of KAP1, SETDB1 and H3K9me3 peaks in chromatin immunoprecipitation / deep sequencing (ChIP-seq) studies of murine ES cells (Bilodeau et al., 2009; Rowe et al., 2013b). Sequences contained within TEs or zinc finger proteins were favored, and imprinting control regions (known for their recruitment of ZFP57) were excluded. We selected 19 such targets (Table S1), and added the PBS[Lys1,2] sequence – the primer binding site (PBS) sequence of many retroviruses, demonstrated to drive KAP1-mediated repression, but through an unknown

KRAB-ZFP intermediate (Wolf et al., 2008). Lentivectors containing these baits upstream of a PGK-GFP cassette were first introduced into murine ES cells, which revealed that 10 out of 20 of these TE-derived fragments induced KAP1-dependent repression of GFP expression (Figure S1). The corresponding vectors were used to engineer stable 293T cell lines, which were transfected in triplicates with individual members of a library of 211 murine KRAB-
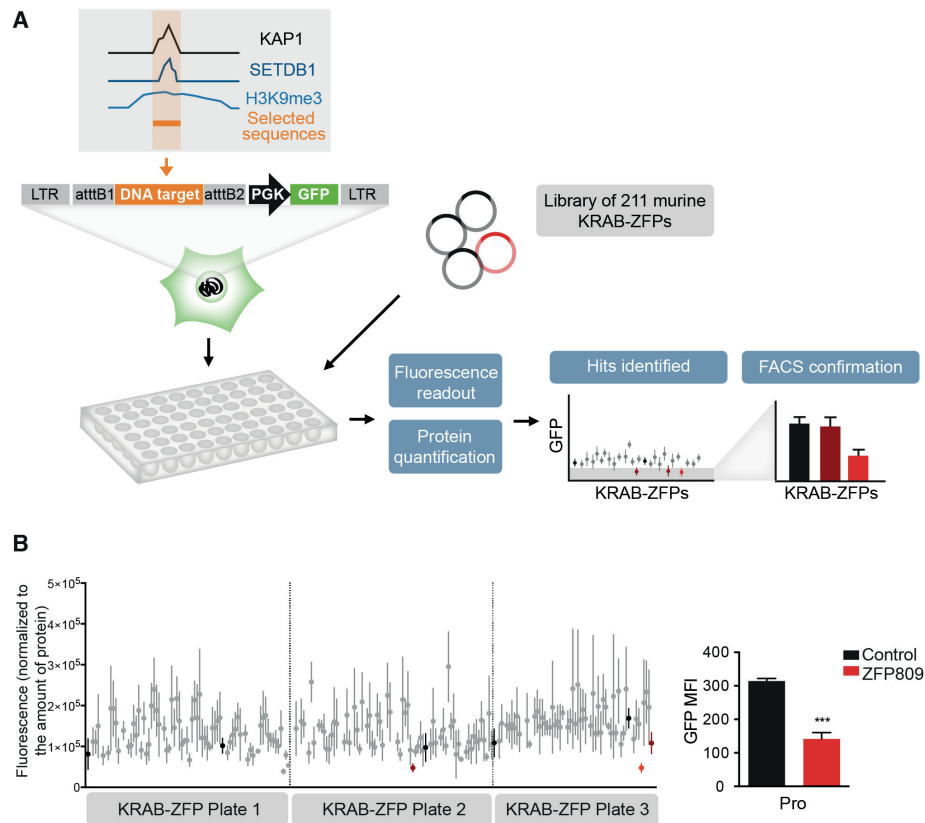
Figure 1. A functional screen to match KRAB-ZFPs repressors with their genomic targets. (**A**) Methodological outline. A library of murine KRAB-ZFPs-expressing plasmids was transfected in 293T stable cell lines containing the DNA sequence of interest upstream a PGK-GFP cassette. Fluorescence readout in 96-well plates was normalized for protein content, and hits were tested by FACS for confirmation. (**B**) Validation of the screen using the previously characterized ZFP809/PBS[Pro] (Pro) pair. Fluorescence readout was performed (left) and identified hits were tested by FACS. The only hit confirmed by FACS was ZFP809 (right panel). Black: transfection control; red: hits identified by plate reader; light red: hits confirmed by FACS. Error bars represent SD, ***p < 0.001, Student's t test. See also Figure S1, Table S1, and Table S2.

ZFPs (Table S2) and examined 6 days later for GFP expression with a fluorescence plate reader; the hits were then confirmed by FACS analysis. A control cell line, transduced with a vector containing the MLV PBS^Pro sequence upstream of the PGK-GFP cassette properly singled out ZFP809 as its cognate repressor, giving us confidence in our approach (Figure 1B). Out of the 10 tested DNA sequences, 3 allowed the identification of a specific KRAB-ZFP ligand (Figures 2A and S2A-S2C). The P9 bait, which comprises a SINE-B2 element, was repressed by Gm6871. The P5 and P8 baits, both derived from ERVKs (RLTR44-int for P5; RLTR9A3, MMERVK10D3_I-int, and RLTR6-int for P8), were repressed by ZFP932. To confirm the matches thereby identified, we introduced the DNA bait-PGK-GFP lentiviral vectors in mouse ES cells depleted for the individual KRAB-ZFPs by RNA interference. For P5, P8 or P9, depleting the corresponding KRAB-ZFP released GFP expression, while it had no effect on a vector containing the PBS^Pro control sequence (Figures 2B and S2D). Furthermore, repression could be restored by overexpressing an shRNA-resistant form of the specific KRAB-ZFP for P9 and P8 (significantly), and to a lesser extent for P5 (not significantly, possibly due to initial lower values). Finally, we could document the bindings of Gm6871 to the P9 sequence, and of ZFP932 to both of its genomic targets by chromatin immunoprecipitation followed by quantitative PCR (ChIP-qPCR) (Figures 2C and S2E). Interestingly, Gm6871 also binds some LINE-1 elements, which it can repress in murine ES cells (Castro-Diaz et al., 2014). In Gm6871-specific ChIP-seq analyses, not only LINEs, but also the SINE-containing P9 locus was detected (Figure S2B), indicating that a KRAB-ZFP can recognize TEs from different subgroups.
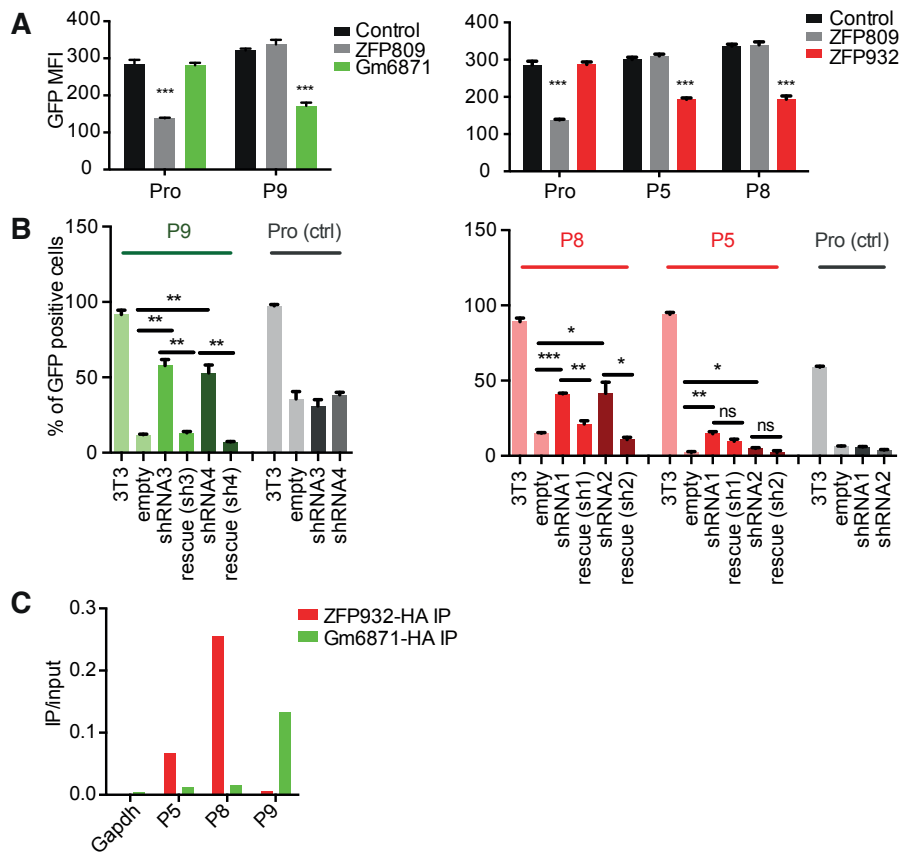
Figure 2. Identification of Gm6871 and ZFP932 as ligands of KAP1-repressed TE sequences. (**A**) FACS confirmation of hits identified through screening of selected target sequences. (**B**) GFP repression assay in control, ZFP932- or Gm6871-KD, or ZFP-complemented murine ESC. (**C**) HA ChIP PCR of Gm6871 and ZFP932 in mES overexpressing the corresponding tagged protein (results representative of 3 independent experiments). Error bars represent SD, * $p < 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, ns= not significant, Student's t test. See also Figure S2.

ZFP932 and its paralog, Gm15446, bind to distinct subsets of ERVK endogenous retroelements

We then turned to the biological characterization of *Zfp932*. We realized that this mouse-specific KRAB-ZFP gene has a paralog, *Gm15446*, located in close proximity on chromosome 5 (Figure 3A), and also targeted by our *Zfp932*-directed siRNAs. When examining the predicted DNA-contacting amino acids within the zinc finger arrays (Liu et al., 2014) of ZFP932 and Gm15446, we saw that the two proteins differed by only two point mutations and the presence
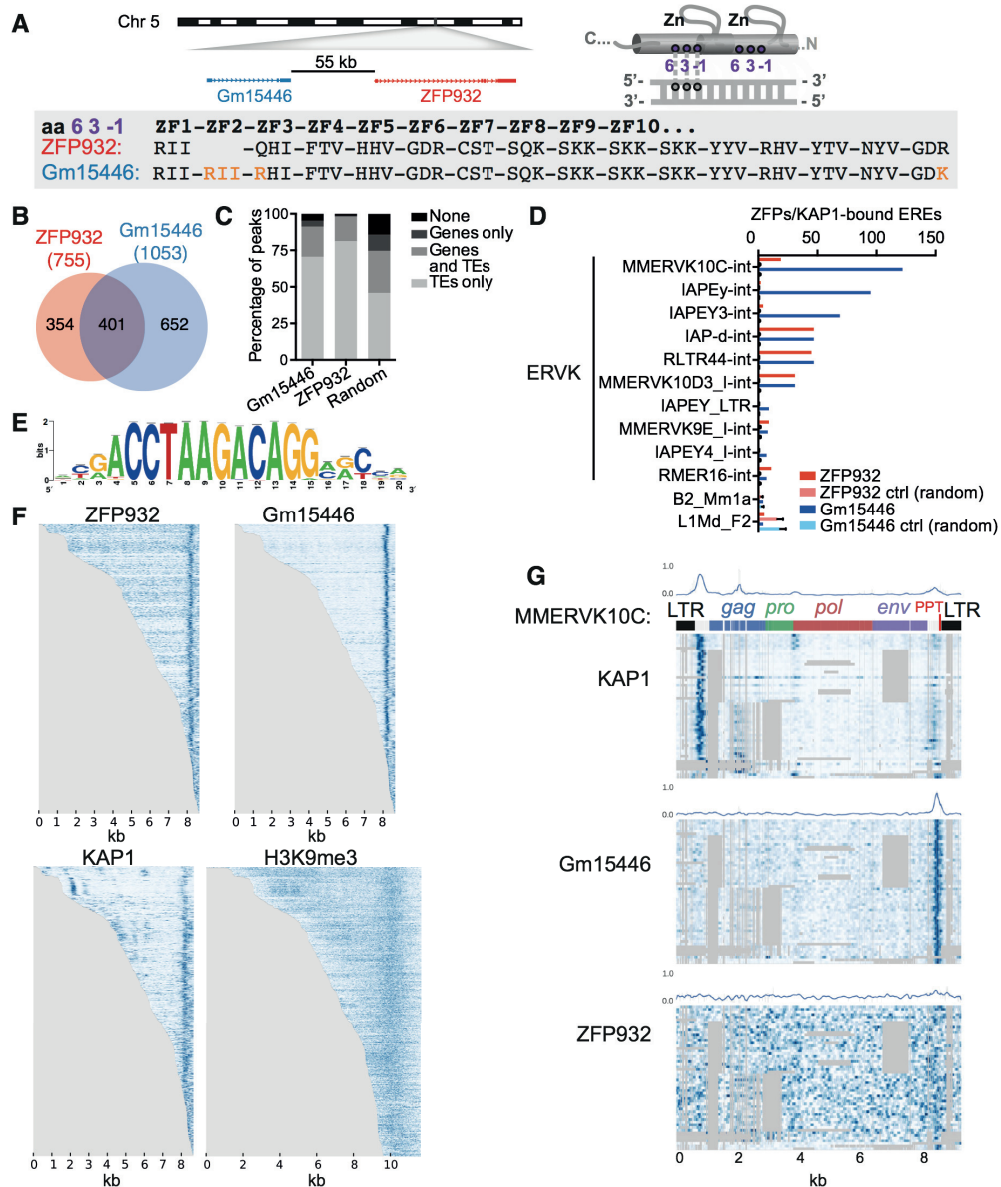
Figure 3. ZFP932 and its paralog Gm15446 bind with KAP1 to the 3' end of distinct subsets of ERVK. (**A**) Representation of genomic location of *Zfp932* and *Gm15446* (left), and scheme of the 3 amino acids important for DNA recognition on a zinc finger domain (right). Lower panel depicts the comparison of the 3 amino acids predicted to contact DNA of each zinc finger of ZFP932 and Gm15446, highlighting differences in orange. (**B**) Venn diagram of ZFP932 and Gm15446 binding sites determined by ChIP-seq in mESC. (**C**) Percentage of ZFP932, Gm15446, or random control peaks on genic and repeated regions. Random control is based on the mean of 100 random shuffling of the peaks of Gm15446 (KRAB-ZFP ChIP with more peaks). (**D**) Top 10 ERE families bound by ZFP932 or Gm15446 together with KAP1 in mESC. SINEs and LINEs (bottom) served as negative controls. A control with 100 random shuffling of the binding sites is also shown. (**E**) Predicted DNA-binding motif for ZFP932 and Gm15446. (**F**) ChIP-seq coverage plots in mESC on all EREs bound by ZFP932 or Gm15446, ranked by size. Each row is independently normalized, with enrichment proportional to darkness of the blue color. For H3K9me3, representation extends 1 kb up and downstream of the ERE boundaries. (**G**) Coverage plot on multiple alignment of different ChIP-seqs in ES cells on "full-length" (>5kb) MMERVK10C-int bound by ZFP932 or Gm15446. Repbase MMERVK10C-int consensus is represented on top. Mean of binding coverage is illustrated in each plot. Each row is independently normalized, with enrichment proportional to darkness of the blue color. Gray areas correspond to gaps in multiple alignments. Error bars represent SD. See also Figure S3.

of one extra zinc finger in Gm15446. In 293T cells, overexpression of either protein led to repression of the same targets, albeit with some differences in efficiency (Figure S3A).

We next investigated by ChIP-seq the genomic targets of ZFP932 and Gm15446 in mouse ES cells overexpressing HA-tagged derivatives of these proteins (Figure S3B). The intersection of duplicate ChIP-seq experiments yielded a total of 755 peaks for ZFP932 and 1053 for Gm15446, amongst which 401 were shared (Figure 3B). Overlapping these data with genic regions and using our updated census of repeats in the mouse genome (see Supplemental Experimental Procedures for details), we observed that most of the KRAB-ZFPs peaks were on TEs (Figure 3C) – either in or outside genes –, the majority of which were ERVKs (68.7% of peaks for ZFP932 and 91.5% for Gm15446). To decrease the possibility of unspecific targets due to overexpression of the proteins, we overlapped the results of these ChIP-seqs with that of similar analysis performed on endogenous KAP1 in murine ES cells. In total, 226 peaks for ZFP932 and 448 for Gm15446 overlapped with KAP1 peaks, mostly at ERVKs, confirming that these KRAB-ZFPs recognize this ERE. Interestingly, some differences were noted between the two paralogs (Figure 3D). For instance, ZFP932 and Gm15446 were similarly enriched at RLTR44-int, IAP-d-int, and MMERVK10D3_I-int elements, but Gm15446 was more frequently found at MMERVK10C-int, IAPEy-int, and IAPEY3-int. Using ChIP-seq data, we identified a motif present in 80% of the ZFP932- and Gm15446-binding sites and 82.7% of ZFPs-enriched ERVKs, but absent from members of the ERV1 family, not targeted by these KRAB-ZFPs (Figure 3E). The underlying sequence is different from a ZFP932-recruiting motif identified the *Ptch1* gene promoter in a limb mesenchymal cell line (He et al., 2011). However, we did not detect ZFP932/Gm15446 at this site. It could be that *Ptch1* binds ZFP932 only in particular cellular contexts, and via another protein.

We then mapped more precisely the ZFP932/Gm15446/KAP1 binding sites within the genome of their ERV targets. MLV and some IAPs (intracisternal A article, a subtype of ERV) were previously found to recruit KAP1 via the proximal part of their provirus, consistent with the observed repression of their nearby 5'LTR (long terminal repeat) (Rowe et al., 2010; Wolf and Goff, 2009a; Wolf et al.,

2015). Surprisingly, both ZFP932 and Gm15446 were instead enriched near the 3'end of their target retroelements, together with KAP1 and H3K9me3 (Figure 3F). An analysis using multiple alignments of MMERVK10C-int elements revealed that the two paralogs and KAP1 bound a region situated just after *env*, upstream of the 3'LTR, partly overlapping with the 3' polypurine tract (PPT), an element important for reverse transcription (Figure 3G). Interestingly, KAP1 was also enriched at two 5' sites in these retroelements, where it was likely recruited by other KRAB-ZFPs since neither of the two paralogs was found there. Finally we asked more broadly if KAP1 binds to multiple regions of retroelements other than ZFP932/Gm15446 targets, and we identified several KAP1 peaks on IAPEz, another subtype of ERVs (Figures S3C-S3D). These results demonstrate that KAP1 can bind to multiple locations of the same TE, including its 3'end, via different KRAB-ZFPs.

ZFP932 and Gm15446 regulate ERVK and nearby gene transcription in mouse embryonic stem cells

To gain insight on the transcriptional impact of these two KRAB-ZFPs, we next deleted *Zfp932* and *Gm15446* from murine ES cells by CRISPR/Cas9-mediated editing of the locus and defined the transcriptome of the resulting cells by RNA-seq (Figure 4A). Mainly members of the LTR/ERV class of transposable elements were upregulated, consistent with the ChIP-seq data. Overlapping these results with those of RNA-seq performed in *Kap1* knockout (KO) ES cells (Figure S4A) unveiled 141 ERV, 17 LINE and 27 SINE integrants, the expression of which was significantly increased in both settings. ERVs that had been identified as ZFP932/Gm15446 binding targets were more upregulated in cells depleted for either protein or KAP1, compared with ERVs on which these KRAB-ZFPs and KAP1 had not been detected (Figure 4B). Of the bound ERVs with increased expression upon ZFPs depletion, many had only the 3'PPT KAP1 peak, or had a 3' peak markedly more pronounced than its 5' counterparts.
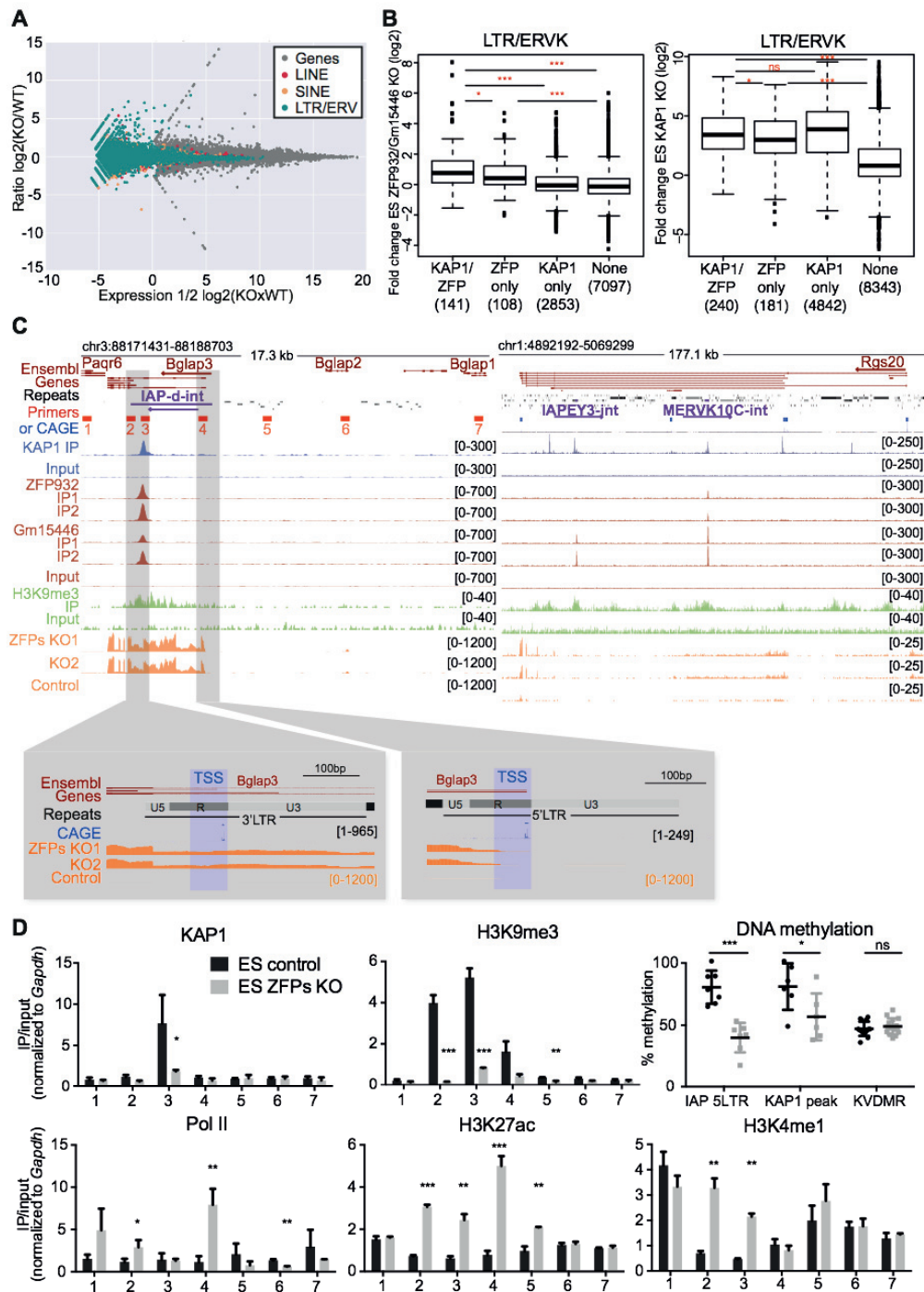
Figure 4. ZFP932/Gm15446/KAP1 binding regulates expression of ERVs and nearby genes in mESC. (**A**) MA-plot of RNA-seq from *Zfp932/Gm15446* KO vs. control ES cells. Expression of genes and TEs is shown. (**B**) Boxplots of ERVK fold expression in *Zfp932/Gm15446* KO (left) or *Kap1* KO (right) ES cells. (**C**) Scheme of *Bglap3* (left) and *Rgs20* (right) genomic loci. ChIP-seq and RNA-seq signal are depicted, along with primers used in panel D (numbered 1-7), and FANTOM CAGE data corresponding to TSSs. Orientation of the genes and TEs is indicated with an arrow. Repeats of interest are highlighted in purple. Lower panels correspond to two regions of zoom in the *Bglap3* locus, indicating the TSSs of the gene. CAGE data correspond to mapped TSS peaks and to the max signal of CAGE reads obtained in one of the tissues analysed by the FANTOM5 consortium. (**D**) ChIP-PCR analysis of different epigenetic marks and pyrosequencing determination of DNA methylation at the *Bglap3* locus in *Zfp932/Gm15446* KO vs. control ES cells. Location of primers is depicted in panel C above. Error bars represent SD, * p < 0.05, ** p

76

≤0.01, *** p ≤ 0.001, ns= not significant, Student's t test was used in panel D and Wilcoxon test in panel B. See also Figure S4.

It was previously demonstrated that the KRAB/KAP1-mediated control of endogenous retroelements preserves the transcriptional dynamics of ES cells by preventing TE-originating enhancer and promoter effects on nearby genes (Rowe et al., 2013b; Wolf et al., 2015). Seventy-one genes were significantly upregulated upon *Zfp932/Gm15446* KO, and 29 of those were significantly more expressed also in *Kap1* KO mES cells, 6 of which were within 50kb of an ERV bound by both the KRAB-ZFPs and their cofactor. Among these genes, *Bglap3* stood out as one of the most upregulated genes in the *Krab-zfps* KO cells (fold-change of 270.9, *P* value of 0.00000003). *Bglap3* has three short and two long transcripts, the latter interestingly containing a ZFP932/Gm15446-bound IAP-d element (Figure 4C). By comparing FANTOM CAGE transcriptional start site (TSS) data with RNA-seq, we observed that both the 5' and the 3'LTRs of the IAP function as promoters of the *Bglap3* gene, with transcripts starting exactly at the beginning of the LTR R region. Both *Bglap3* and its IAP were more expressed in murine ES cells deleted for either *Zfp932/Gm15446* or *Kap1*, demonstrating that the retroelement partakes in the controlling unit of this gene (Figures 4C and S4B-S4C). This effect was confirmed through ZFP932/Gm15446 knockdown (KD) experiments, properly controlled by restoration of the repression upon complementation with interfering RNAs-resistant derivatives of the two KRAB-ZFPs (Figure S4D). The *Rgs20* locus provided another example of gene-TE pair regulated by ZFP932/Gm15446, in which no TSS overlapped with the TEs, suggesting that this gene is regulated through enhancer effects (Figure 4C and S4B).

We deepened our study of this phenomenon by examining the epigenetic status of the *Bglap3* locus (Figure 4D). Upon *Zfp932/Gm15446* deletion, ChIP-PCR analyses revealed loss of KAP1 and H3K9me3 and gain of the activation marks H3K27ac and H3K4me1 at the IAP, together with the accumulation of PolII at the predicted TSSs of *Bglap3*. Furthermore, pyrosequencing unveiled a marked drop in DNA methylation both underneath the KAP1 peak and at the 5' end of the IAP.

These data indicate that, in murine ES cells, ZFP932/Gm15446 epigenetically regulate not only the expression of their target ERVKs but also the transcriptional influence of these TEs on nearby genes, via both promoter and enhancer effects.

The KRAB/KAP1 system regulates TEs in differentiated cells

To gain further insight into *Zfp932/Gm15446* functions we assessed their expression in representative murine tissues, using CAGE datasets generated by the FANTOM consortium (Lizio et al., 2015) (Figure 5A). Both paralogs were broadly expressed in adult murine cells, as was *Zfp809*. In contrast, *Zfp459* and *Zfp819*, which we previously found to be highly transcribed in murine pluripotent stem cells (Corsinotti et al., 2013), were largely restricted to undifferentiated cells. Because ZFP932 and Gm15446 displayed a strong preference for ERVK family members as their genomic targets, we suspected that they might control these TEs in adult tissues as well. To probe this hypothesis, we first examined by RNA-seq the transcriptomes of hepatocytes harvested from control and liver-specific *Kap1* KO mice, wild type and *Kap1*-deleted mouse embryonic fibroblasts (MEFs) (Figure S5A), and control and KAP1 KD C2C12 mouse myoblast cells. In all these settings, expression of ERVs was increased upon KAP1 depletion, even though induction was less pronounced than in *Kap1* KO ES cells (Figure 5B). We ranked the significantly upregulated families in each dataset and found that MERVK10C-int had the highest number of induced integrants in KAP1-depleted differentiated cells (Figure S5B). General MERVK10C-int primers confirmed by RT-qPCR the increased expression of these elements in liver and C2C12 depleted for KAP1, in spite of the absence of detectable changes in the DNA methylation status of these loci as measured by bisulfite sequencing (Figures S5C-S5D). Interestingly, there were differences in the subsets of ERVs deregulated in the various tissues following KAP1 removal (Figure 5B). For example, particular MMERGLN and ORR1A0 integrants located on chromosome 12 were upregulated in KAP1-depleted liver and ES cells, but not in their C2C12 counterparts (Figure S5C).
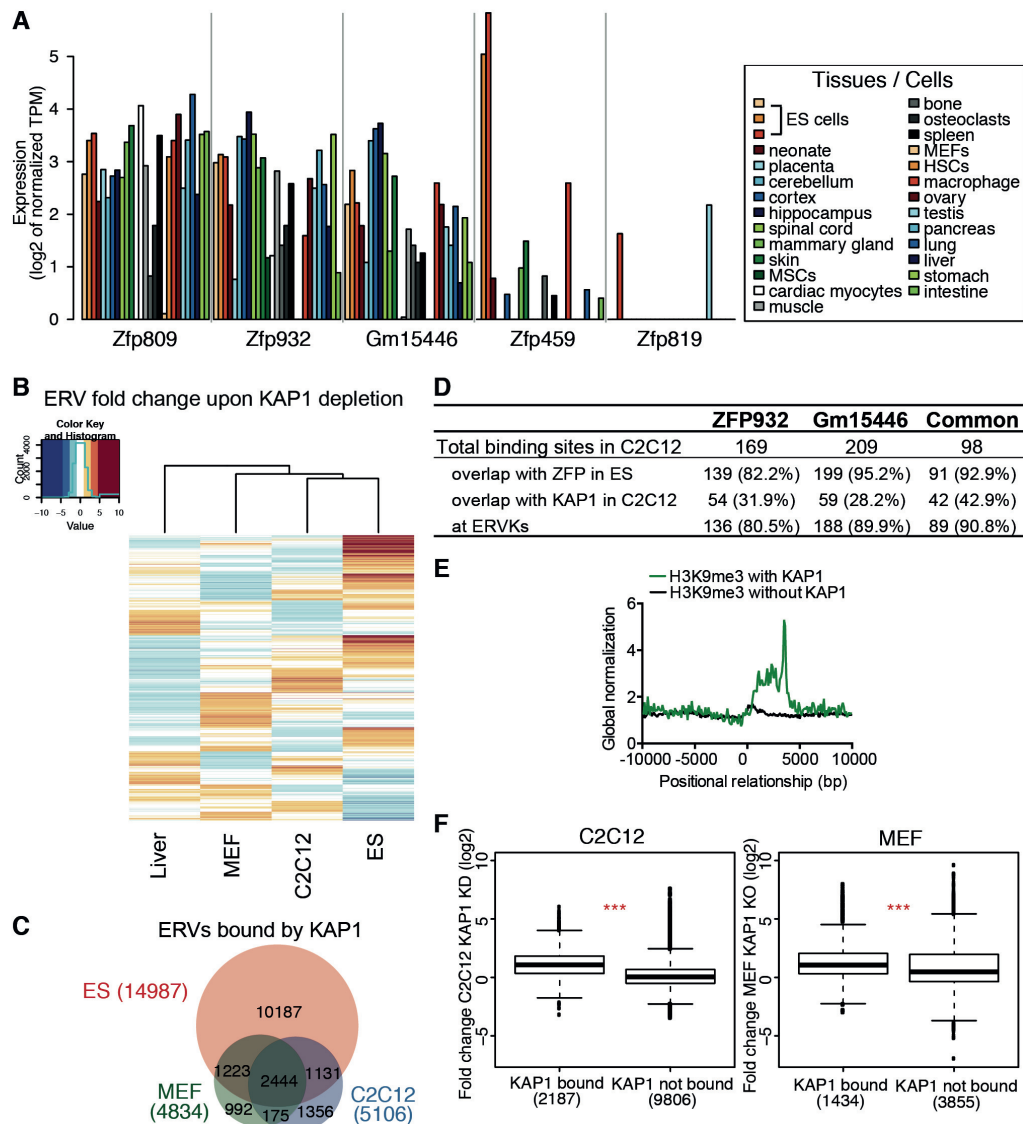
Figure 5. The KRAB/KAP1 system controls TE expression in differentiated tissues. (**A**) *Krab-zfp* genes mRNA expression according to FANTOM 5 CAGE data. MSCs, mesenchymal stem cells; HSCs, hematopoietic stem cells. (**B**) Heatmap of ERVs average fold-change expression (linear scale) upon KAP1 depletion in different tissues. Liver = *Kap1* KO versus control, C2C12 = KAP1 KD versus control, MEF = *Kap1* KO versus control, ES = *Kap1* KO versus control. (**C**) Venn diagram of ERVs bound by KAP1 in ES, C2C12, or MEF cells. (**D**) Comparative table of ZFP932, Gm15446, or common ZFP932 and Gm15446 binding sites in C2C12, determined by ChIP-seq. Absolute values and percentage relative to total binding sites are given. (**E**) Distribution of H3K9me3 ChIP-seq with or without presence of KAP1, in C2C12, relative to the 5' end of ERVK elements (5' end corresponds to 0 bp). (**F**) Boxplot of ERVKs fold-change expression in C2C12 KAP1 KD or MEFs *Kap1* KO versus control. *** p ≤ 0.001, Wilcoxon test. See also Figure S5.

Confirming that the KRAB/KAP1 system is involved in the somatic control of TEs, about 24% of ERVs bound by KAP1 in ES cells still bore the corepressor in C2C12 cells and MEFs (Figure 5C). Interestingly, for 30% of the ERV loci bound by KAP1 in C2C12 and 24% in MEFs no significant enrichment was found in ES cells. We also performed ZFP932- and Gm15446-specific ChIP-seq analyses in C2C12 cells expressing HA-tagged forms of these proteins (Figures 5D and S5E). We detected 169 ZFP932 and 209 Gm15446 binding sites, 98 of which were common, and the large majority of which were also bound by either one or both of these KRAB-ZFPs in ES cells. Most resided within TEs, namely ERVKs, and KAP1 was also detected at about a third of them, supporting a role for these ZFPs in regulating such elements in differentiated cells. We then asked if H3K9me3 marks were present at KAP1-bound ERVKs in C2C12. Using H3K9me3 ChIP-seq data in these cells, we observed a higher correlation of H3K9me3 at ERVKs with than without KAP1 (Figure 5E). General ERVK primers revealed upregulation of MERVK10C, IAP-d, and IAPEz in C2C12 cells depleted for SETDB1, suggesting that this histone methyltransferase partners with KAP1 also in differentiated cells to control TEs (Figure S5F). Finally, ERVKs that were bound by KAP1 in C2C12 or MEFs were significantly more expressed than their unbound counterparts when the corepressor was depleted (Figure 5F). Taken together, these results demonstrate a role for the KRAB/KAP1 system in the somatic control of TEs.

TEs and their KRAB-ZFP controllers regulate gene expression in adult tissues

We then asked if the KRAB/KAP1-mediated control of TEs also prevents the illegitimate transcription of nearby genes in adult cells. For this, we first analyzed our RNA-seq of KAP1-depleted cells, separating the genes in stable, up- or downregulated (deregulation criteria was 2-fold change cut-off and P value adjusted ≤ 0.05), and calculated their distance to closest KAP1-bound ERV. As expected, in ES cells upregulated genes were significantly closer to a KAP1-bound ERV than their downregulated or stable counterparts (Figure 6A). In the case of C2C12 KAP1 KD, although the same trend was observed, upregulated
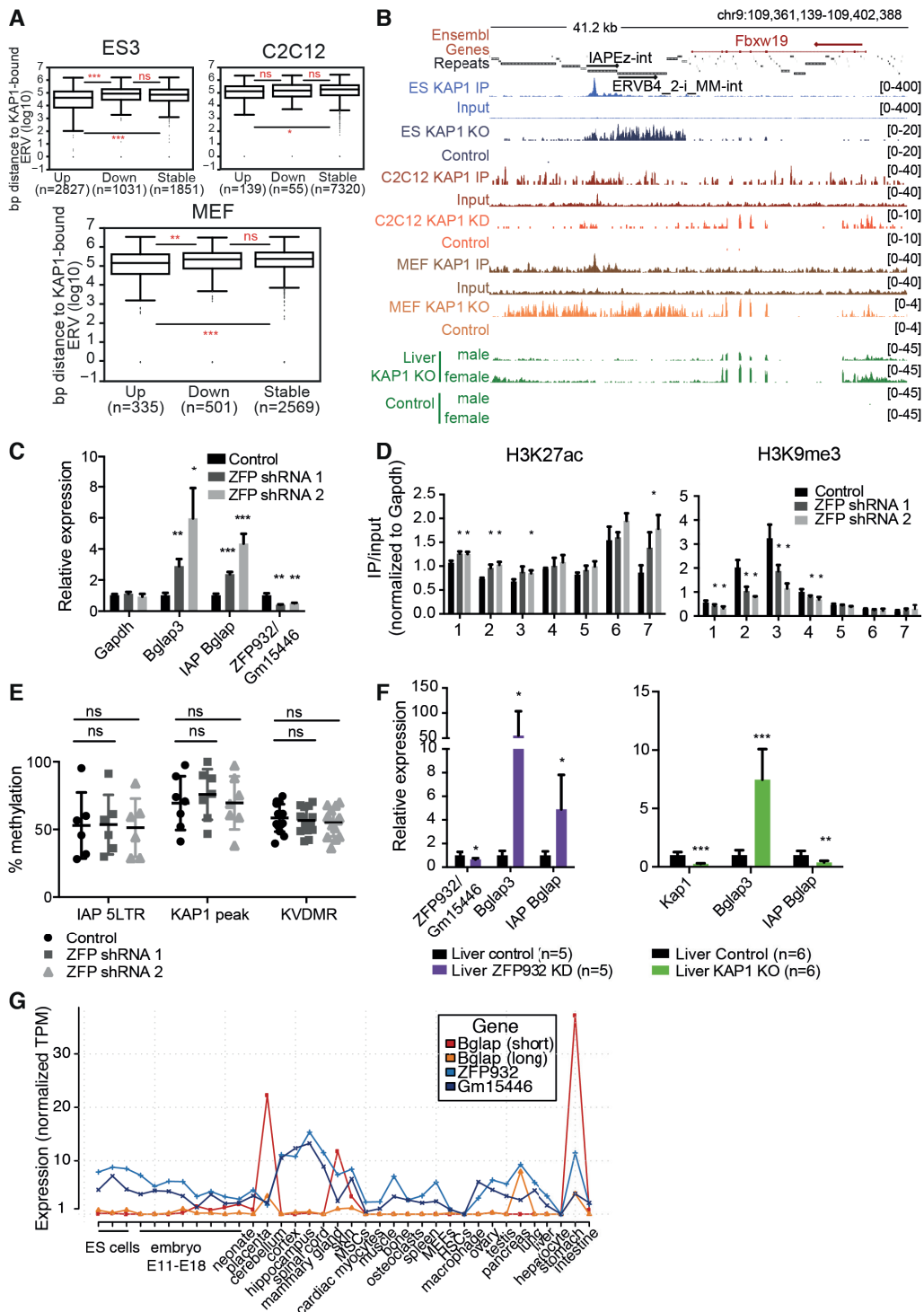
Figure 6. TEs and their KRAB/KAP1 controllers influence gene expression in adult tissues.
(**A**) Distance to nearest KAP1-bound ERV of upregulated (fold-change ≥ 2, padj ≤0.05), stable, or downregulated (fold-change ≤ 2, padj ≤ 0.05) genes in ES cells *Kap1* KO, C2C12 KAP1 KD, or MEFs *Kap1* KO. (**B**) Representative genomic region with ERV and *Fbxw19* gene regulated by KAP1 in differentiated cells. KAP1-depleted and control RNA-seq densities of different tissues/cells are depicted along with KAP1 ChIP-seq data. (**C**) mRNA expression of *Bglap3* gene and IAP Bglap in ZFP932/Gm15446 KD in C2C12 cells versus control (normalized to *ActinG* and

Tbp). (**D**) ChIP-PCR of chromatin marks on *Bglap3* locus in ZFP932/Gm15446 KD vs. control C2C12 cells. Numbers correspond to primers in Figure 4C. (**E**) DNA methylation analysis by pyrosequencing on IAP Bglap in same cells. (**F**) mRNA expression of *Bglap3* gene and IAP Bglap in ZFP932/Gm15446 KD or *Kap1* KO mice liver compared to control (normalized to *Gapdh* and *ActinB*). (**G**) *Zfp932*, *Gm15446*, and *Bglap3* gene expression in different murine tissues/cells according to FANTOM 5 CAGE data. MSCs, mesenchymal stem cells; HSCs, hematopoietic stem cells. Error bars represent SD, * p < 0.05, ** p ≤0.01, *** p ≤ 0.001, ns= not significant, Wilcoxon test was used in panel A and Student's t test in panels C-F. See also Figure S6.

genes were only significantly closer to KAP1-bound ERVs when compared to stable but not downregulated genes. Interestingly, for MEFs *Kap1* KO we observed the same pattern as in ES cells, suggesting that, even if to a smaller extent, KAP1-regulated TEs can impact neighbouring genes transcription also in differentiated cells.

We confirmed our hypothesis further by examining in more details a few upregulated genes situated within 100kb of a KAP1-bound ERV. At some loci, such as *Fbxw19*, KAP1 depletion resulted in activating both the ERV and its nearby gene in all cells examined (Figures 6B and S6A-B). At others, such as *Gm13251*, removal of the regulator induced expression in liver and in ES, but not in C2C12 cells (Figure S6B). We then examined the *Bglap3* locus, which we found enriched for ZFP932/Gm15446, KAP1, and H3K9me3 also in C2C12 cells (Figure S6C). Upon RNA interference-mediated KD of these two KRAB-ZFPs, both the IAP and *Bglap3* were activated (Figure 6C), coincident with loss of H3K9me3 and KAP1, and mild but significant gain of H3K27ac at the IAP (Figures 6D and S6D). However, in spite of its transcriptional activation, this ERV remained DNA methylated (Figure 6E).

To expand these findings *in vivo*, we generated ZFP932/Gm15446 KD mice by transduction of fertilized oocytes with shRNA-expressing lentiviral vectors. We could not assess major phenotypic abnormalities in these animals, but a marked upregulation of *Bglap3* gene and the upstream IAP was measured in their liver, albeit again with no detectable loss of DNA methylation at this locus (Figures 6F and S6E). ChIP-PCR experiments in murine wild type liver confirmed the enrichment of KAP1 and H3K9me3 at the IAP (Figure S6F). Interestingly, in the liver of conditional *Alb-Cre Kap1* KO mice, in which the regulator is deleted not

during early embryogenesis but some three weeks after birth, there was increased expression of *Bglap3* but not of the IAP, suggesting an uncoupling of the regulation of the two LTRs of the retroelement (Figure 6F). Also in this setting, we observed that the IAP Bglap remained highly methylated (Figure S6G). Finally, we asked if the *Bglap3* gene is expressed in physiological conditions and how does it correlate with the ZFPs transcription. By analysing CAGE expression data we observed that in tissues like placenta, mammary gland, and pancreas *Bglap3* transcripts are expressed and correlate with lower *Zfp932/Gm15446* levels (the long transcript driven by the IAP 5'LTR, and the short transcript by the 3'LTR) (Figure 6G). Collectively, these data establish that the KRAB/KAP1 system uses TE-residing platforms to regulate gene expression in adult cells.

DISCUSSION

This work establishes that TEs and their KRAB-ZFP controllers can regulate gene expression in adult tissues. Our results thus invalidate a generally accepted model, which assumes that the KRAB/KAP1 system irreversibly silences TEs during early embryonic development (Maksakova et al., 2008; Mikkelsen et al., 2007; Rowe and Trono, 2011; Walsh et al., 1998; Wolf et al., 2015). Furthermore, our study strongly suggests that *Krab-Zfp* genes are not selected simply to inactivate TEs, but also to domesticate their transcriptional potential for the benefit of the host.

Here, through a target-centric functional screen, we identified several KRAB-ZFPs responsible for the sequence-specific repression of TEs in mouse ES cells, including two KRAB-ZFP paralogs that recognized partly overlapping members of the ERVK family. Not only could we verify that these two KRAB-ZFPs controlled these ERVs and neighboring genes in ES cells, but we further discovered that they were expressed in a broad range of somatic murine cells. We could also demonstrate that, in several differentiated cell types, including *in vivo* in the liver, KAP1 and KRAB-ZFPs were still bound to and controlled TEs, and modulated the expression of nearby genes via promoter and enhancer effects, even if with a smaller global impact than the observed in ES cells. Remarkably, this process occurred without apparent alteration of the DNA methylation status of these loci, but primarily involved histone-based changes.

Although DNA methylation is traditionally considered as a silencing mechanism, emerging evidence indicates that its impact on gene expression is far more diversified (Jones, 2012; Schubeler, 2012). We had previously observed that docking of the KRAB/KAP1 complex at genomic loci leads to their DNA methylation when it occurs in early embryonic cells but not in their differentiated counterparts, where it induces repression solely by histone-based, hence fully reversible, changes (Groner et al., 2012; Quenneville et al., 2012; Wiznerowicz et al., 2007). The present data further indicate that protein-restricted chromatin modifications can reignite transcription at these sites in adult cells, in spite of persistent DNA methylation. This is consistent with recent

observations that some ERVs are activated without changes in their DNA methylation status in B cells depleted for the histone methyltransferase SETDB1 (Collins et al., 2015), and that, in human ES cells, transcription can occur from some highly methylated KAP1-controlled promoters (Turelli et al., 2014).

Whether in the context of murine or human ES cells, we had previously noted that KAP1 depletion does not systematically trigger the activation of all KAP1-bound TEs (Rowe et al., 2010; Turelli et al., 2014). Here, we observed that, in adult cells as well, the range of ERVs activated upon KAP1 or ZFP932/Gm16446 depletion differed depending on the cellular environment. In C2C12, for example, many IAPEz elements bound by KAP1 were not significantly de-repressed upon its removal (not illustrated). At the *Gm13251* locus, KAP1 depletion activated two ERVs in liver and in ES cells, but not in MEFs or C2C12 cells. Correspondingly, in *Setdb1* KO B-lymphocytes, the upregulation of selected ERVs correlated with the presence within their promoters of binding sites for B cell-specific transcription factors (Collins et al., 2015). It is likely that this type of restriction broadly applies to KRAB-ZFP-controlled loci, the expression of which is conditioned not only by removal or biochemical inactivation of their epigenetic repressors, but also by the presence of a proper set of activators, likely tissue-specific. In addition, specific loci may be subjected to dominant influences imposed by the local chromatin configuration, and other control mechanisms, such as small RNAs-based, may be at play in some environments (Bierhoff et al., 2014; Heras et al., 2013; Marchetto et al., 2013; Pezic et al., 2014). Finally, KAP1 can undergo locus-specific post-translational modifications switching its function from corepressor to coactivator (Singh et al., 2015), and KRAB-ZFPs could also be subjected to this type of regulation, leading to the recruitment of different sets of effector complexes.

Our demonstration that KRAB-ZFPs are involved in controlling TEs not only in embryonic but also differentiated cells is consistent with the widespread expression of these proteins in adult organisms, whether mouse or human (Lizio et al., 2015), and with the finding that a great majority of these transcription factors have TEs as their preferred genomic targets (Najafabadi et al., 2015) (and our unpublished results). It also contributes to explain the broad range of

phenotypes induced by the conditional KO of *Kap1* in the mouse, even though the master regulator carries out some KRAB-ZFP- and TE-independent functions (Iyengar and Farnham, 2011; Iyengar et al., 2011; McNamara et al., 2016; Singh et al., 2015).

Our study unveils several other unsuspected aspects of the KRAB/KAP1-mediated control of transposable elements. First, we found that a given KRAB-ZFP can recognize different subgroups of TEs, e.g. Gm6871 repressing members of both the LINE and SINE families. Second, we determined that the two ERVK-targeting KRAB-ZFPs identified here tether KAP1 near the 3' end of their retroviral targets, not at the PBS region or close to their promoter as previously documented for MLV, several other IAPs, and most LINEs (Castro-Diaz et al., 2014; Jacobs et al., 2014; Rowe et al., 2010; Wolf and Goff, 2009a; Wolf et al., 2015). Interestingly, this 3'end peak overlaps with the 3'PPT, a sequence important for reverse transcription. While it readily explained how ZFP932 and Gm15446 can control the impact of the IAP Bglap 3'LTR on the *Bglap3* gene, it was more surprising to observe that this distal location also served to repress the 5'LTR of this element, as demonstrated by the spreading of chromatin marks and its upregulation in ES cells or mice where these two KRAB-ZFPs were depleted. Third, we observed the differential regulation of the two LTRs of the same ERV, and consequently of the gene placed under their influence, via one KRAB/KAP1-binding site. Depletion of ZFP932/Gm15446 from the earliest times of embryonic development activated the expression of both *Bglap3* and its controlling IAP in the liver of adult mice, whereas the deletion of *Kap1* in this organ after birth only de-repressed the *Bglap3* gene, but not its IAP. It thus seems that the regulation of the 5' and 3' LTR of this endogenous retrovirus can be uncoupled. In the absence of obvious differences in the DNA methylation status of these two transcription units, the molecular basis of their distinctive behavior remains to be identified. We could even speculate that, as previously reported (Wolf et al., 2015), the depletion of KRAB/KAP1 regulation at different moments of development could lead to differential TE control due to deposition of other epigenetic marks. Finally, we found that KAP1 can be recruited to several regions of a same ERV, likely via distinct KRAB-ZFPs. It strongly suggests

that a TE can be regulated in temporally and functionally differential fashions, hence that the complexity of KRAB/KAP1-mediated regulation of TEs and their gene targets is much greater than envisioned so far.

Our results support a model in which the evolutionary selection of KRAB-ZFP genes is not only an arms race aimed at silencing TEs but also the instrument of their domestication. It is likely that TEs and their KRAB-ZFPs regulators modulate multiple aspects of the biology of tetrapods, superimposing a species-specific layer of control over canonical, conserved regulatory pathways. Indeed, a large fraction of recognizable mobile elements in the genome are unique to the corresponding species and its close relatives, both in sequence and location; accordingly, levels of orthology between KRAB-ZFPs of different organisms are limited. Therefore, KRAB-ZFPs and their targets must play major roles in the speciation of higher vertebrates, including humans.

EXPERIMENTAL PROCEDURES

Cell culture and mouse work. Murine ES cells and MEFs wild type and KO for *Kap1* were cultured and generated as previously described (Rowe et al., 2013a; Rowe et al., 2013b). Mouse C2C12 myoblasts were cultured as described (Singh et al., 2015). Two clonal *Krab-zfps* KO cell lines were generated using an integrase defective lentiviral vector containing the pLentiCRISPR with an sgRNA for *Zfp932* and *Gm15446* or a control luciferase sgRNA. KD experiments were performed with specific shRNA in pLKO lentiviral vectors. Hepatocyte-specific *Kap1* KO mice were generated and genotyped according to (Bojkowska et al., 2012). ZFP932/Gm15446 KD mice were generated by lentiviral transgenesis as described (Rowe et al., 2013a). All shRNAs and sgRNAs used in this study are listed in Supplemental Experimental Procedures. All animal experiments were approved by the local veterinary office and carried out in accordance with the EU Directive (2010/63/EU) for care and use of laboratory animals.

Functional screen. DNA target sequences to be tested were chosen from the overlap of publicly available murine ESC ChIP-seq data for KAP1 (GEO: GSE41903), SETDB1 (GEO: GSE18371), H3K9me3 (GEO: GSE41903), and absence of ZFP57 (GEO: GSE31183). PBS^{Lys1,2} sequence plus 19 sequences corresponding to KAP1 peaks were selected based on presence of TEs or of interesting KAP1 targets (such as 3'end of ZFP genes) (see Table S1). These sequences were cloned upstream a PGK-GFP cassette, tested for repression in murine ES, and only repressed sequences were tested in the screen. Lentiviral vectors with these sequences were used to transduce 293T cell lines, which were sorted for GFP presence, generating stable cell lines. The screen was performed by reverse transfection of a library of 211 murine KRAB-ZFPs (Table S2), in 96-well plates, in an automated fashion, in triplicates. Cells were harvested on day 6 after transfection, lysed, and GFP fluorescence was measured by plate reader (excitation at 485 nm and emission at 520 nm). Total protein content was quantified using BCA (BCA Protein Assay Reagent, Thermoscientific), and used to calculate normalized GFP fluorescence. Candidate hits were identified by selecting the 10 KRAB-ZFPs with the lowest normalized fluorescence values per 96-well plate, and only the ones that were present in all 3 replicates were

considered. Hits were identified by transfection of the 293T cell line of interest with the candidate KRAB-ZFPs, in 24-well plates, with FACS readout after 6 days. The detailed protocol is given in Supplemental Experimental Procedures.

RT-PCR and RNA-seq. Total RNA was extracted and DNAse-I treated using a spin column-based RNA purification kit (Macherey-Nagel). cDNA was prepared with SuperScript II reverse transcriptase (Invitrogen). Primers (listed in Supplemental Experimental Procedures) were used for SYBR Green qPCR (Applied Biosystems), and specificity was confirmed with dissociation curves. For mRNA sequencing, 100-bp single-end RNA-seq libraries were prepared using the Illumina TruSeq Stranded mRNA reagents (Illumina). Cluster generation was performed with the resulting libraries using the Illumina TruSeq SR Cluster Kit v4 reagents. Sequencing was performed with Illumina HiSeq 2500 in 100-bp reads run. The RNA-seq reads were mapped to the mm9 genome using TopHat (Kim et al., 2013), allowing multimapped reads to be randomly assigned once among the mapped loci. Gene counts were generated with HTseq-count program using default parameters, and TE counts were computed using BEDtools (multicov). Sequencing depth normalization and differential expression analyses were performed using the voom function of Bioconductor package LIMMA (Law et al., 2014).

ChIP-PCR and ChIP-seq. Chromatin immunoprecipitation and library preparation were done according to (Rowe et al., 2013b), with modifications as described in Supplemental Experimental Procedures. Sequencing was performed with Illumina HiSeq 2500 in 100-bp reads run. Reads were mapped to the mouse genome assembly mm9 using Bowtie 2 (Langmead and Salzberg, 2012), using the sensitive-local mode. Peaks were called using MACS (Zhang et al., 2008) or SICER software (for histone modification marks) (Zang et al., 2009), with total input as control.

DNA methylation analysis. Genomic DNA was extracted, converted using an Epitect Bisulfite kit (Qiagen), and used in two rounds of PCR followed by PCR product purification. Pyrosequencing and bisulfite sequencing were performed as previously described (Fasching et al., 2015; Rowe et al., 2013a).

Bioinformatics analyses and statistics. All genome-wide TE analyses were performed using a merged repeats track generated in-house (by using RepeatMasker 3.2.8 and merging homonymous ERV-int integrants or attributed LTRs within 400bp or less). Genomic region analyses were done with BEDTools (Quinlan and Hall, 2010). Motif search was performed using with RSAT (Medina-Rivera et al., 2015), using unbound ERVK elements as control. For coverage plots, ChIP-seq signal on each feature of interest were extracted from the bigwig file, scaled between 0 and 1; in some cases, multiple alignment was performed with MAFFT. Statistical difference was assessed by Student's t test, except for Figures 4B, 5F, and 6A for which Wilcoxon test was applied. Error bars represent ± 1SD. R version 3.1.2 (http://www.R-project.org) or GraphPad Prism version 4.0 (http://www.graphpad.com) were used for statistical analyses. Detailed bioinformatics analyses are provided in Supplemental Experimental Procedures.

Data Access. All next-generation sequencing data have been submitted to the NCBI Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/) database under the accession number GEO: GSE74278.

AUTHOR CONTRIBUITIONS

ACKNOWLEDGMENTS

REFERENCES

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

Bierhoff, H., Dammert, M.A., Brocks, D., Dambacher, S., Schotta, G., and Grummt, I. (2014). Quiescence-induced LncRNAs trigger H4K20 trimethylation and transcriptional silencing. Molecular cell *54*, 675-682.

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

Bojkowska, K., Aloisio, F., Cassano, M., Kapopoulou, A., Santoni de Sio, F., Zangger, N., Offner, S., Cartoni, C., Thomas, C., Quenneville, S., *et al.* (2012). Liver-specific ablation of Kruppel-associated box-associated protein 1 in mice leads to male-predominant hepatosteatosis and development of liver adenoma. Hepatology (Baltimore, Md.) *56*, 1279-1290.

Bourque, G., Leong, B., Vega, V.B., Chen, X., Lee, Y.L., Srinivasan, K.G., Chew, J.L., Ruan, Y., Wei, C.L., Ng, H.H., *et al.* (2008). Evolution of the mammalian transcription factor binding repertoire via transposable elements. Genome research *18*, 1752-1762.

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

Chuong, E.B. (2013). Retroviruses facilitate the rapid evolution of the mammalian placenta. BioEssays : news and reviews in molecular, cellular and developmental biology *35*, 853-861.

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. Proceedings of the National Academy of Sciences of the United States of America *112*, 8367-8372.

Corsinotti, A., Kapopoulou, A., Gubelmann, C., Imbeault, M., Santoni de Sio, F.R., Rowe, H.M., Mouscaz, Y., Deplancke, B., and Trono, D. (2013). Global and stage

specific patterns of Kruppel-associated-box zinc finger protein gene expression in murine early embryonic cells. PloS one *8*, e56721.

Dupressoir, A., Lavialle, C., and Heidmann, T. (2012). From ancestral infectious retroviruses to bona fide cellular genes: role of the captured syncytins in placentation. Placenta *33*, 663-671.

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

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

Fort, A., Hashimoto, K., Yamada, D., Salimullah, M., Keya, C.A., Saxena, A., Bonetti, A., Voineagu, I., Bertin, N., Kratz, A*., et al.* (2014). Deep transcriptome profiling of mammalian stem cells supports a regulatory role for retrotransposons in pluripotency maintenance. Nature genetics *46*, 558-566.

Friedli, M., and Trono, D. (2015). The developmental control of transposable elements and the evolution of higher species. Annual Review of Cellular and Developmental Biology *31*, 13.11-13.23.

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

Hancks, D.C., and Kazazian, H.H., Jr. (2012). Active human retrotransposons: variation and disease. Current opinion in genetics & development *22*, 191-203.

He, Z., Cai, J., Lim, J.W., Kroll, K., and Ma, L. (2011). A novel KRAB domain-containing zinc finger transcription factor ZNF431 directly represses Patched1 transcription. The Journal of biological chemistry *286*, 7279-7289.

Heras, S.R., Macias, S., Plass, M., Fernandez, N., Cano, D., Eyras, E., Garcia-Perez, J.L., and Caceres, J.F. (2013). The Microprocessor controls the activity of mammalian retrotransposons. Nature structural & molecular biology *20*, 1173-1181.

Hutnick, L.K., Huang, X., Loo, T.C., Ma, Z., and Fan, G. (2010). Repression of retrotransposal elements in mouse embryonic stem cells is primarily mediated

by a DNA methylation-independent mechanism. The Journal of biological chemistry *285*, 21082-21091.

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

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

Jackson-Grusby, L., Beard, C., Possemato, R., Tudor, M., Fambrough, D., Csankovszki, G., Dausman, J., Lee, P., Wilson, C., Lander, E., *et al.* (2001). Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. Nature genetics *27*, 31-39.

Jacobs, F.M., Greenberg, D., Nguyen, N., Haeussler, M., Ewing, A.D., Katzman, S., Paten, B., Salama, S.R., and Haussler, D. (2014). An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. Nature *516*, 242-245.

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

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

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome biology *14*, R36.

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

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

Law, C.W., Chen, Y., Shi, W., and Smyth, G.K. (2014). voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome biology *15*, R29.

Liu, H., Chang, L.H., Sun, Y., Lu, X., and Stubbs, L. (2014). Deep vertebrate roots for mammalian zinc finger transcription factor subfamilies. Genome biology and evolution *6*, 510-525.

Lizio, M., Harshbarger, J., Shimoji, H., Severin, J., Kasukawa, T., Sahin, S., Abugessaisa, I., Fukuda, S., Hori, F., Ishikawa-Kato, S.*, et al.* (2015). Gateways to the FANTOM5 promoter level mammalian expression atlas. Genome biology *16*, 22.

Macfarlan, T.S., Gifford, W.D., Driscoll, S., Lettieri, K., Rowe, H.M., Bonanomi, D., Firth, A., Singer, O., Trono, D., and Pfaff, S.L. (2012). Embryonic stem cell potency fluctuates with endogenous retrovirus activity. Nature *487*, 57-63.

Maksakova, I.A., Mager, D.L., and Reiss, D. (2008). Keeping active endogenous retroviral-like elements in check: the epigenetic perspective. Cellular and molecular life sciences : CMLS *65*, 3329-3347.

Marchetto, M.C., Narvaiza, I., Denli, A.M., Benner, C., Lazzarini, T.A., Nathanson, J.L., Paquola, A.C., Desai, K.N., Herai, R.H., Weitzman, M.D.*, et al.* (2013). Differential L1 regulation in pluripotent stem cells of humans and apes. Nature *503*, 525-529.

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

McNamara, R.P., Reeder, J.E., McMillan, E.A., Bacon, C.W., McCann, J.L., and D'Orso, I. (2016). KAP1 Recruitment of the 7SK snRNP Complex to Promoters Enables Transcription Elongation by RNA Polymerase II. Molecular cell *61*, 39-53.

Medina-Rivera, A., Defrance, M., Sand, O., Herrmann, C., Castro-Mondragon, J.A., Delerce, J., Jaeger, S., Blanchet, C., Vincens, P., Caron, C.*, et al.* (2015). RSAT 2015: Regulatory Sequence Analysis Tools. Nucleic acids research *43*, W50-56.

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

Najafabadi, H.S., Mnaimneh, S., Schmitges, F.W., Garton, M., Lam, K.N., Yang, A., Albu, M., Weirauch, M.T., Radovani, E., Kim, P.M.*, et al.* (2015). C2H2 zinc finger proteins greatly expand the human regulatory lexicon. Nature biotechnology *33*, 555-562.

Pezic, D., Manakov, S.A., Sachidanandam, R., and Aravin, A.A. (2014). piRNA pathway targets active LINE1 elements to establish the repressive H3K9me3 mark in germ cells. Genes & development *28*, 1410-1428.

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

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics (Oxford, England) *26*, 841-842.

Rebollo, R., Miceli-Royer, K., Zhang, Y., Farivar, S., Gagnier, L., and Mager, D.L. (2012). Epigenetic interplay between mouse endogenous retroviruses and host genes. Genome biology *13*, R89.

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

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

Rowe, H.M., Kapopoulou, A., Corsinotti, A., Fasching, L., Macfarlan, T.S., Tarabay, Y., Viville, S., Jakobsson, J., Pfaff, S.L., and Trono, D. (2013b). TRIM28 repression of retrotransposon-based enhancers is necessary to preserve transcriptional dynamics in embryonic stem cells. Genome research *23*, 452-461.

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

Schubeler, D. (2012). Molecular biology. Epigenetic islands in a genetic ocean. Science (New York, N.Y.) *338*, 756-757.

Singh, K., Cassano, M., Planet, E., Sebastian, S., Jang, S.M., Sohi, G., Faralli, H., Choi, J., Youn, H.D., Dilworth, F.J*., et al.* (2015). A KAP1 phosphorylation switch controls MyoD function during skeletal muscle differentiation. Genes & development *29*, 513-525.

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

Turelli, P., Castro-Diaz, N., Marzetta, F., Kapopoulou, A., Raclot, C., Duc, J., Tieng, V., Quenneville, S., and Trono, D. (2014). Interplay of TRIM28 and DNA methylation in controlling human endogenous retroelements. Genome research *24*, 1260-1270.

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

Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P.*, et al.* (2002). Initial sequencing and comparative analysis of the mouse genome. Nature *420*, 520-562.

Wiznerowicz, M., Jakobsson, J., Szulc, J., Liao, S., Quazzola, A., Beermann, F., Aebischer, P., and Trono, D. (2007). The Kruppel-associated box repressor domain can trigger de novo promoter methylation during mouse early embryogenesis. The Journal of biological chemistry *282*, 34535-34541.

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

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

Wolf, D., Hug, K., and Goff, S.P. (2008). TRIM28 mediates primer binding site-targeted silencing of Lys1,2 tRNA-utilizing retroviruses in embryonic cells. Proceedings of the National Academy of Sciences of the United States of America *105*, 12521-12526.

Wolf, G., Yang, P., Fuchtbauer, A.C., Fuchtbauer, E.M., Silva, A.M., Park, C., Wu, W., Nielsen, A.L., Pedersen, F.S., and Macfarlan, T.S. (2015). The KRAB zinc finger protein ZFP809 is required to initiate epigenetic silencing of endogenous retroviruses. Genes & development *29*, 538-554.

Zang, C., Schones, D.E., Zeng, C., Cui, K., Zhao, K., and Peng, W. (2009). A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. Bioinformatics (Oxford, England) *25*, 1952-1958.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W.*, et al.* (2008). Model-based analysis of ChIP-Seq (MACS). Genome biology *9*, R137.

SUPPLEMENTAL INFORMATION

A

B
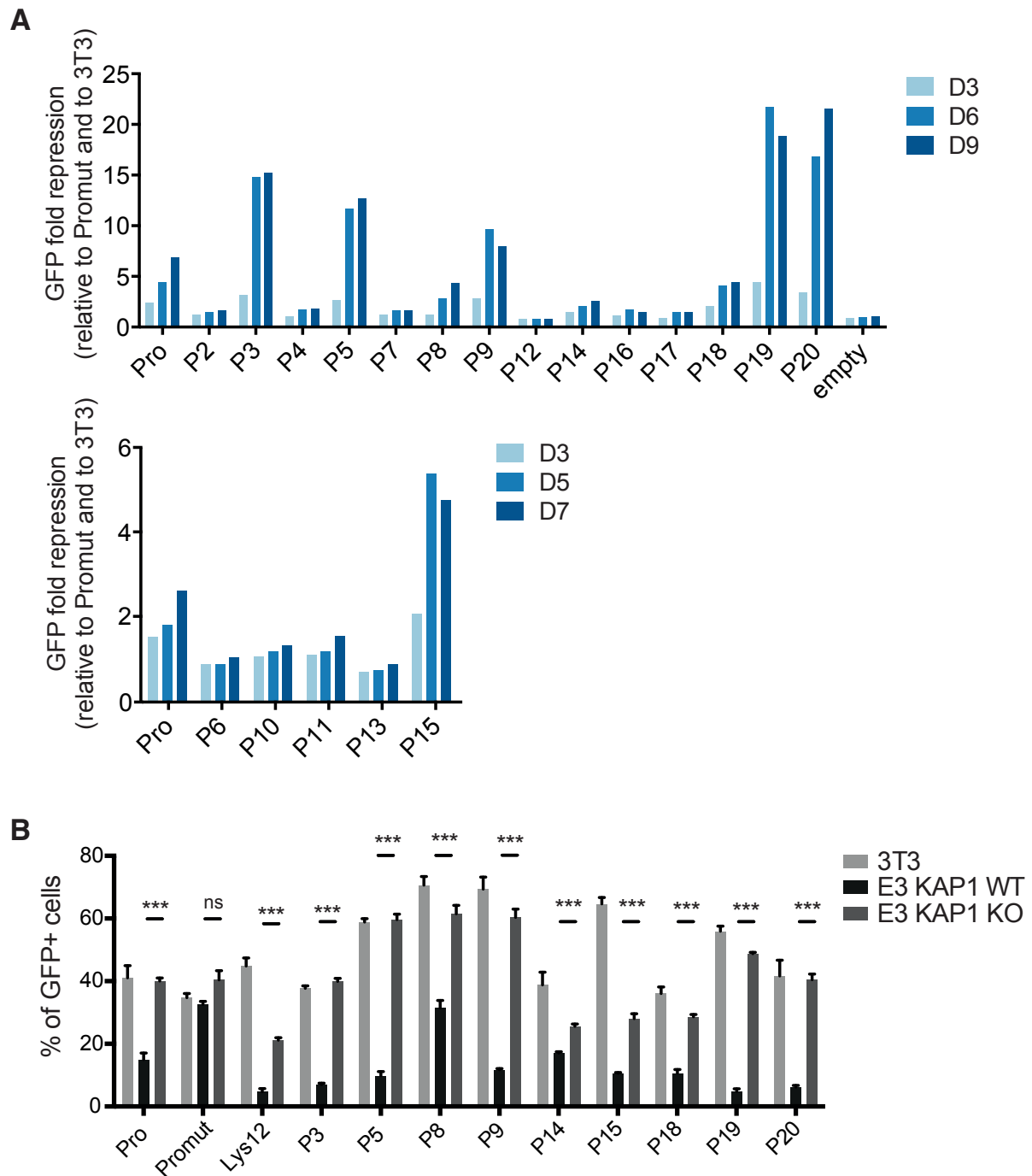
Figure S1, related to Figure 1. Ten of the DNA target sequences selected for the screen are repressed in murine ES in a KAP1-dependent manner.

(**A**) Repression assays in murine ES cells of selected DNA target sequences. ES or 3T3 cells were transduced with vectors containing the sequence of interest upstream a PGK-GFP cassette and percentage of GFP positive cells was measured by FACS at different time-points after transduction. Data are represented as GFP repression normalized to the negative control Promut and the corresponding 3T3 control. (**B**) GFP repression assay with strongly repressed target sequences in *Kap1* WT or KO mESC, using 3T3 cells as control (day 4 after transduction). Promut is a negative control, corresponding to a point mutant of PBS^Pro (also known as B2), which is not repressed in mESC. Error bars represent SD, *** p ≤ 0.001, ns= not significant, Student's t test.

A

B

C

D

E
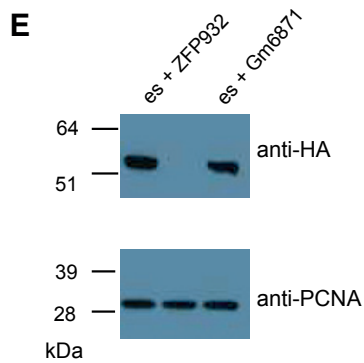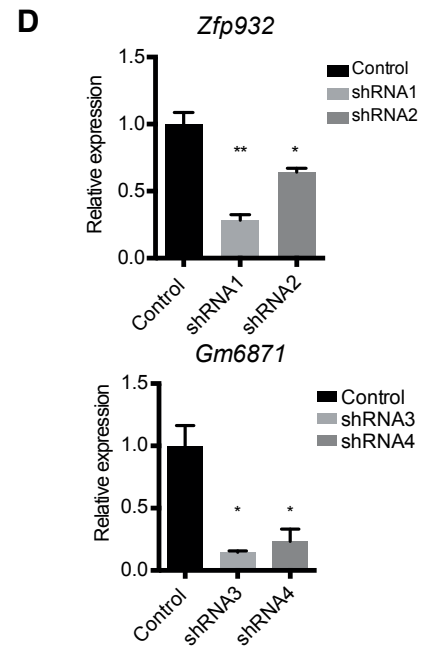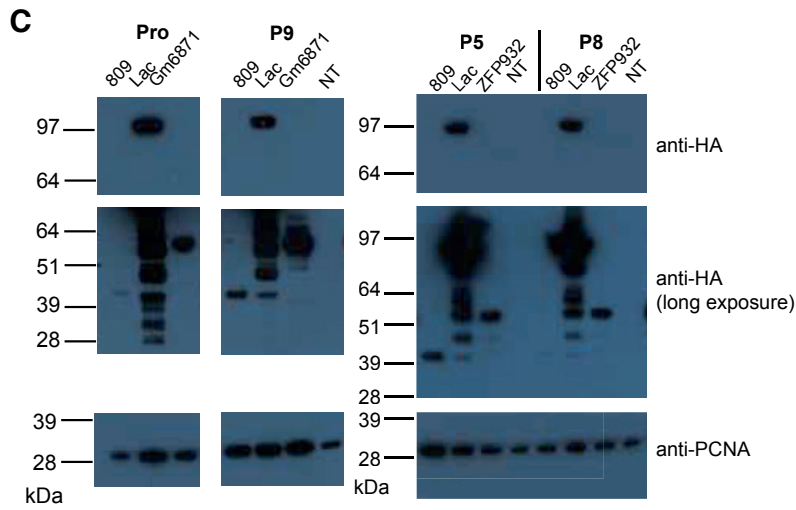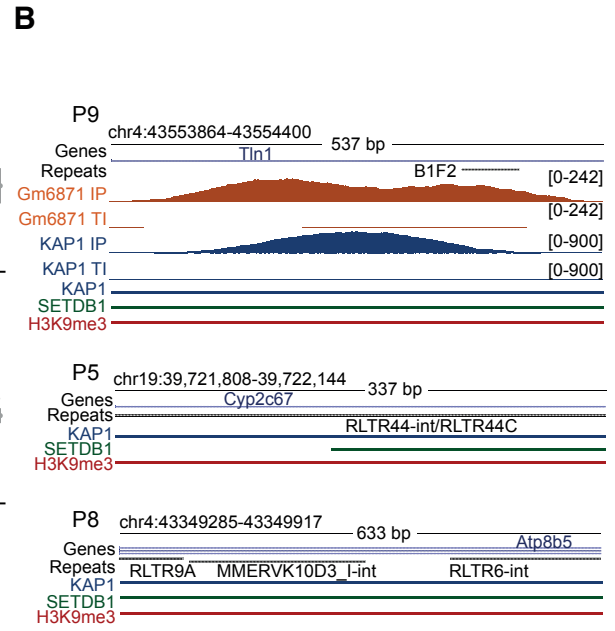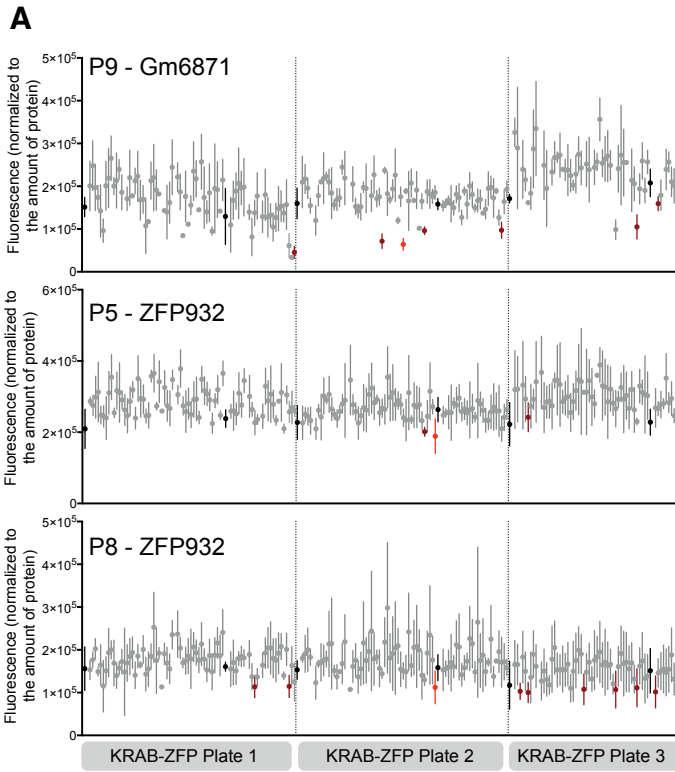
Figure S2, related to Figure 2. Hits identified by subjecting selected target sequences to functional screen.

(**A**) Normalized fluorescence high-throughput readout for P9 and Gm6871, and P5/P8 and ZFP932. Black: transfection control; red: hits identified by plate reader; light red: hits confirmed by FACS. (**B**) Genomic region of P9, P5 and P8. Different ChIP-seq peaks or reads densities are shown for each locus. (**C**) Western blot control of 293T cells overexpressing HA-tagged proteins, referent to experiment in Figure 2A. (**D**) Levels of *Zfp932* and *Gm15446* mRNA upon KRAB-ZFP depletion through shRNAs, corresponding to experiment depicted in Figure 2B. (**E**) Western blot of ES cell lines overexpressing HA-tagged ZFP932 or Gm6871, used for ChIP-PCR in Figure 2C. Error bars represent SD, * $p < 0.05$, ** $p \leq 0.01$, Student's t test.

A

GFP MFI (y-axis, 0–500) vs cis-acting sequence (Pro, P5, P8)

Legend:
- LacZ
- ZFP809
- ZFP932
- Gm15446

B

anti-HA

anti-PCNA

kDa

C

KAP1-bound IAPEz

LTR gag pro pol LTR

bp

D

chr8:123927320-123938011 10.7 kb IAPEz

Repeats LTRI -int ILTR

KAP1 IP [0-900]

KAP1 TI [0-900]

chr13:64978658-64987422 8.8 kb Cntnap3

Genes

Repeats IAPEz

LTRI -int ILTR

KAP1 IP [0-900]

KAP1 TI [0-900]
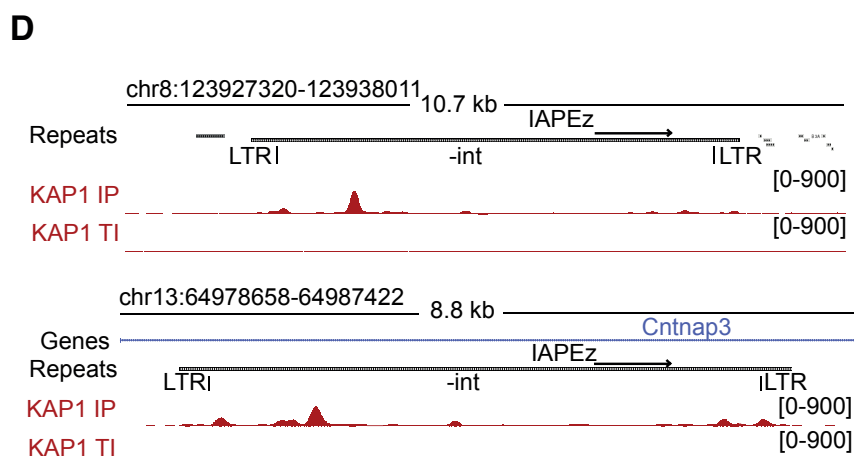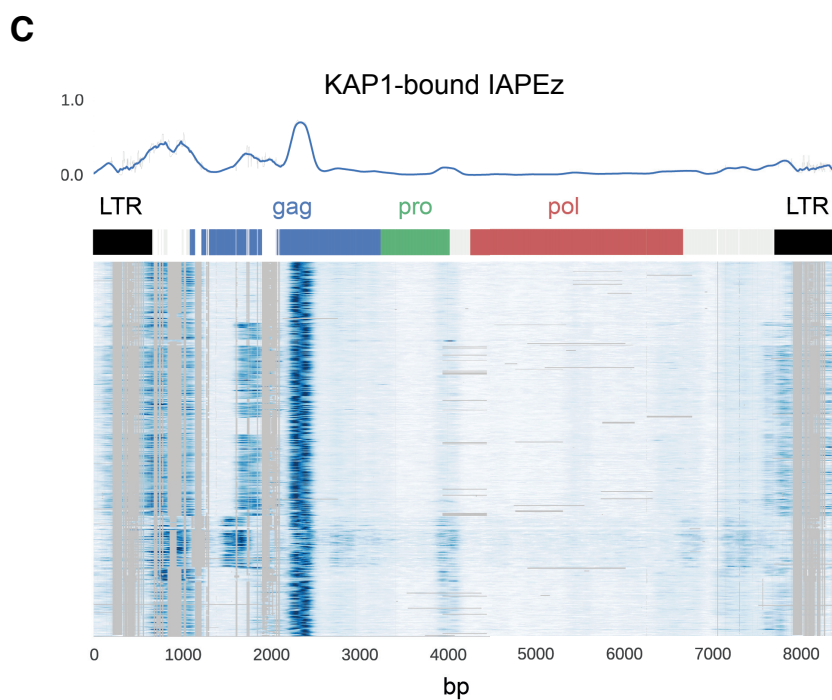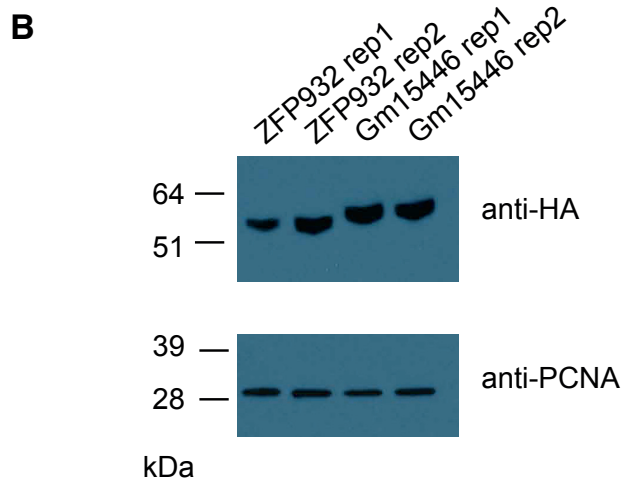
Figure S3, related to Figure 3. Differential regulation of TEs by the KRAB/KAP1 system.

(**A**) GFP repression assay of ZFP932/Gm15446 targets upon KRAB-ZFPs overexpression in 293T cell lines containing target *cis*-acting sequence upstream of PGK-GFP cassette. GFP mean fluorescence intensity is displayed and *P* value presented is relative to LacZ control, unless otherwise specified. (**B**) Western blot of ES cell lines used for ChIP-seq, overexpressing HA-tagged derivatives of ZFP932 or Gm15446. (**C**) Coverage plot of KAP1 ChIP-seq in ES cells on multiple alignment of "full length" (>6kb) IAPEz-int bound by KAP1. IAPEz repbase consensus is represented, and mean of binding coverage is depicted on top. Each row is individually normalized, with enrichment proportional to darkness of the blue color. (**D**) Two representative genomic regions containing full-length IAPEz. LTRs are highlighted and KAP1 ChIP-seq signal in murine ES is depicted. Error bars represent SD, ** p ≤0.01, *** p ≤ 0.001, ns= not significant, Student's t test.
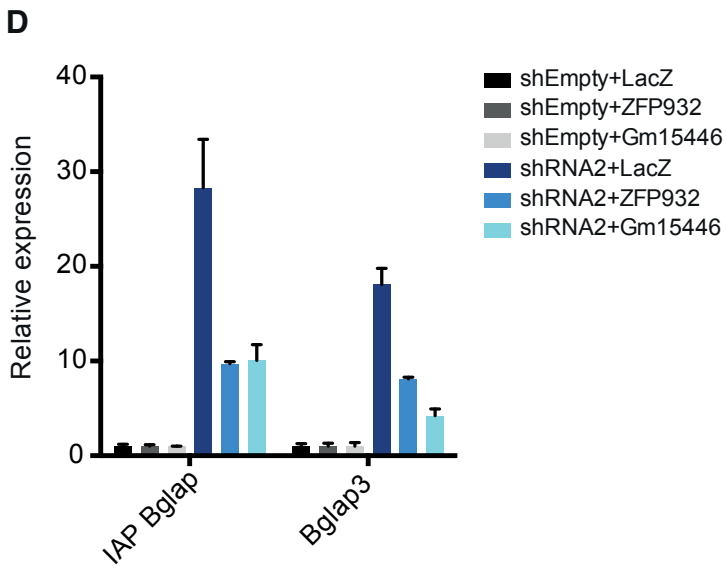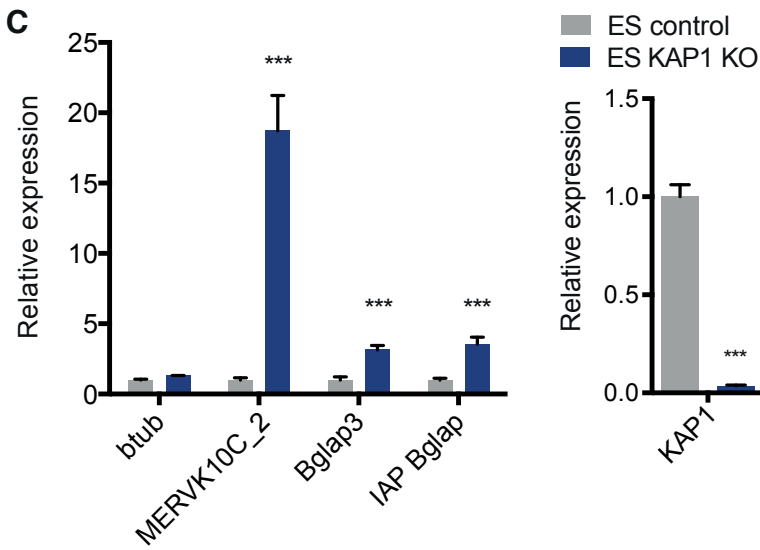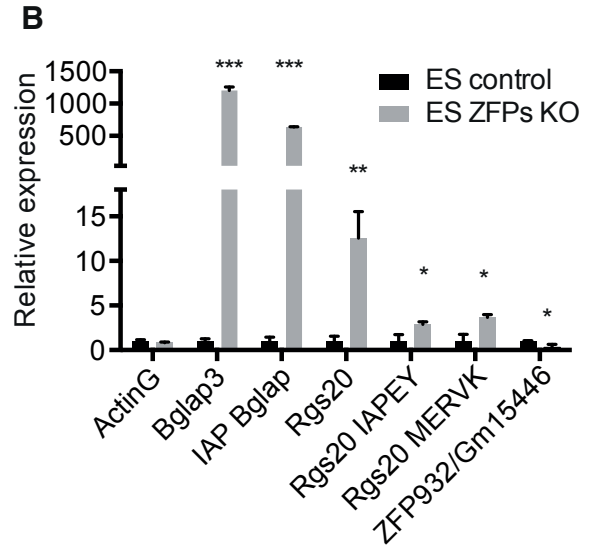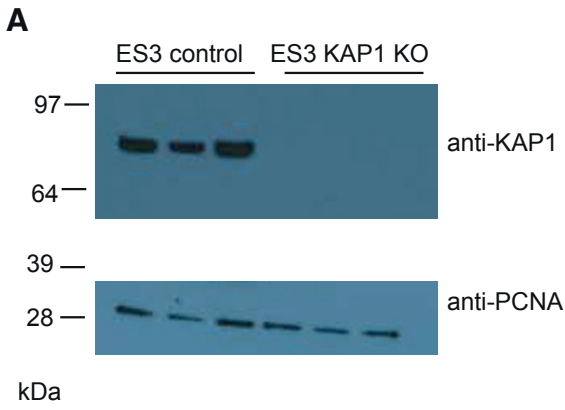
Figure S4, related to Figure 4. KAP1 and ZFP932/Gm15446 regulate the *Bglap3* locus in murine ESC.

(**A**) Western blot of *Kap1* KO ES cells used for RNA-seq. RT-qPCR analysis of mRNA expression of TE-gene pairs in *Zfp932/Gm15446* KO mES cells (**B**), *Kap1* KO in mES cells (**C**), and ZFP932/Gm15446 KD mES cells complemented with LacZ (control), ZFP932 or Gm15446 as indicated (**D**) (normalized to *Gapdh* and *ActinB*). For ZFP932/Gm15446 KD, ZFPs depletion was between 50-60%. Error bars represent SD, * p < 0.05, ** p ≤0.01, *** p ≤ 0.001, Student's t test.
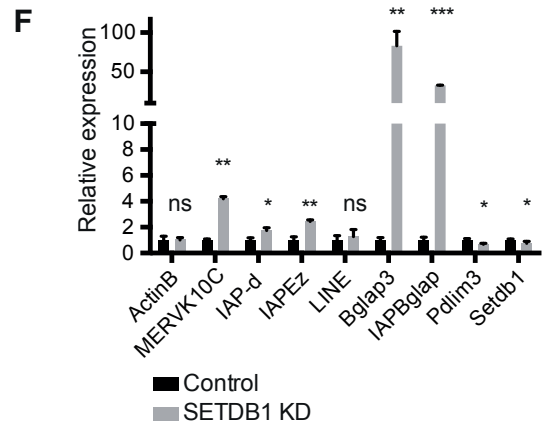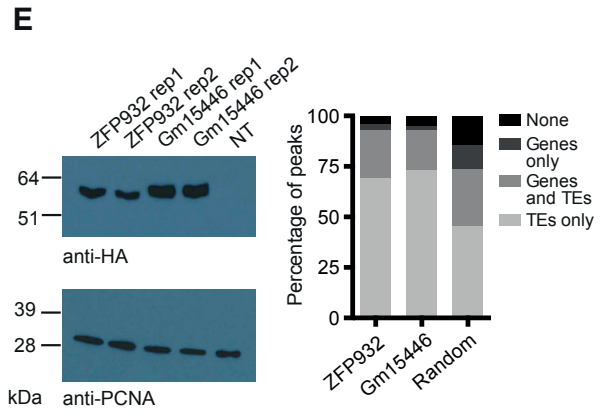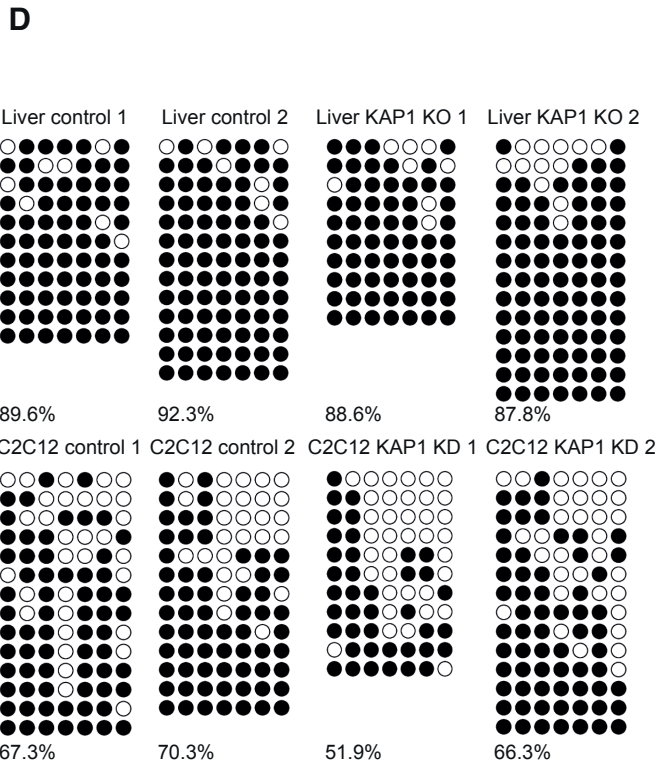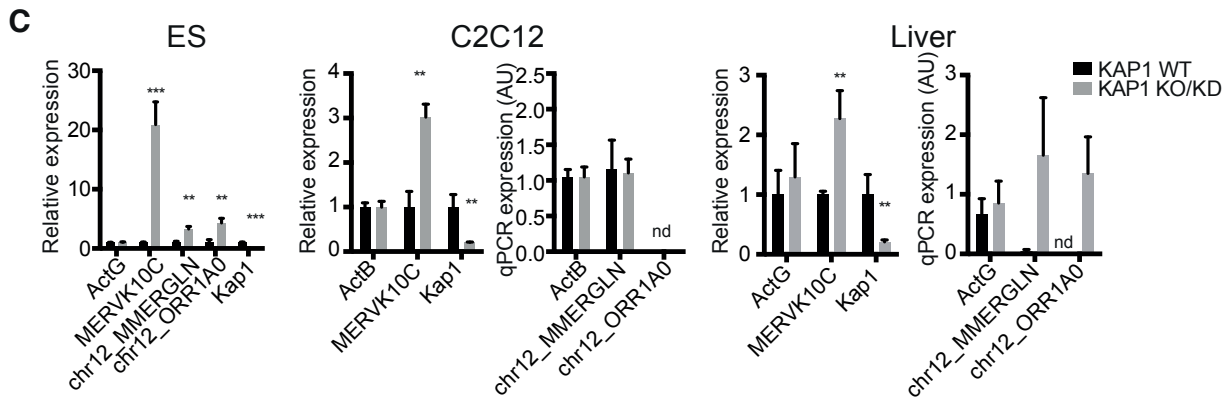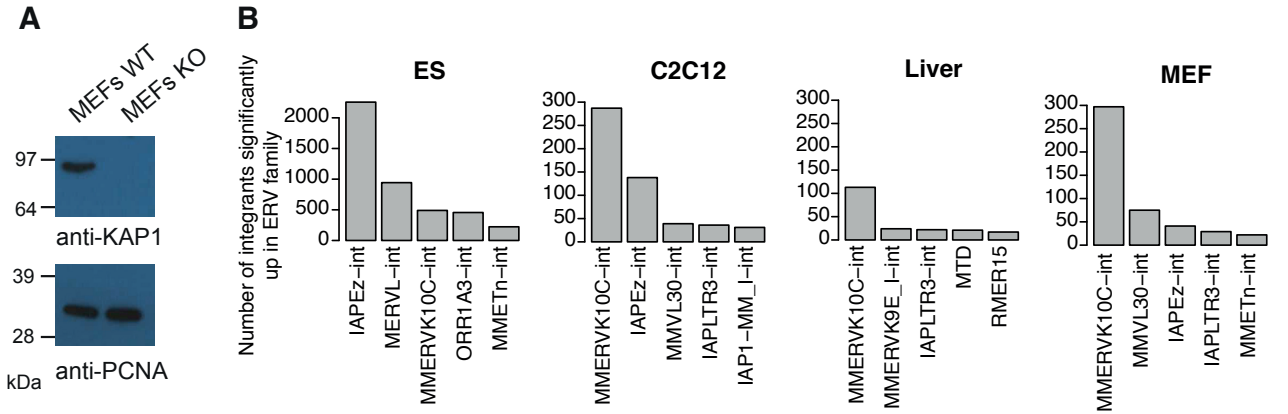
Figure S5, related to Figure 5. ERVs are controlled by KAP1 in differentiated cells.

(**A**) Western blot of MEFs *Kap1* KO cells used for RNA-seq. (**B**) Top 5 upregulated ERV families upon KAP1 depletion in different cells and tissues. Criteria for inclusion were more then 2-fold upregulation when compared to control cells and adjusted *P* value ≤ 0.05. (**C**) mRNA expression of ERVs in liver, C2C12, and ES KAP1-depleted cells versus control (ES and liver samples are normalized to *Gapdh* and *ActinB*; C2C12 samples are normalized to *ActinG* and *Tbp*). MERVK10C and MERGLN/ORR1A0 expression was measured with general and locus-specific primers, respectively. nd, not detected. (**D**) Bisulfite sequencing of MMERVK10C-int sequences in *Kap1* KO vs. wild type liver and in KAP1 KD vs. control C2C12 cells. Empty and full circles correspond to unmethylated and methylated CpG dinucleotides, respectively. (**E**) Western blot of C2C12 cell lines used for ChIP-seq, overexpressing HA-tagged derivatives of ZFP932 or Gm15446; and percentage of ZFP932, Gm15446, or random control peaks on genic and repeated regions. Random control is based on the mean of 100 random shuffling of the peaks of Gm15446 (KRAB-ZFP ChIP with more peaks). (**F**) RT-qPCR analysis of mRNA expression of general ERVKs, *Bglap3* gene, and IAP Bglap in SETDB1 KD C2C12 cells (samples are normalized to *ActinG* and *Tbp*). *Plim3* is a control gene known to be downregulated. Error bars represent SD, * p < 0.05, ** p ≤0.01, *** p ≤ 0.001, ns= not significant, Student's t test.

A. Ensembl genes, Repeats, KAP1 peaks ES, ES KAP1 KO, ES WT, MEFs KAP1 KO, MEFs WT tracks showing Fbxw19 gene, IAPEz and ERVB4 repeats, coverage read depth plots.

B.
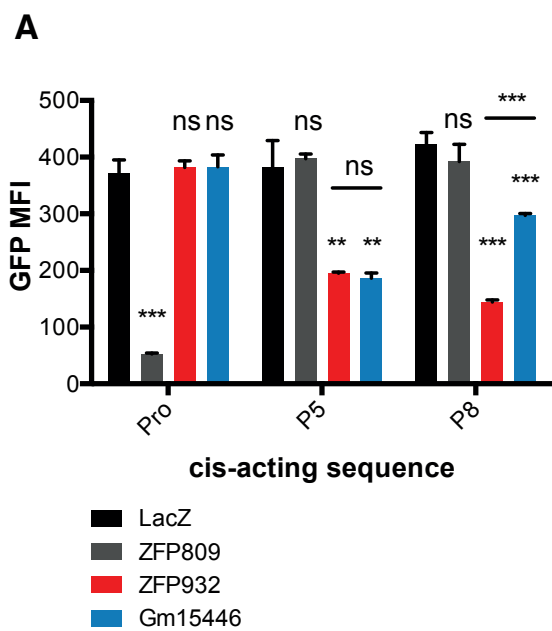C2C12 / Liver RT-qPCR expression (AU): ActinB, Fbxw19, ERVB4 / ActinG, Fbxw19, ERVB4
C2C12 / Liver Relative expression: ActinB, Gm13251, ETn, MERVK, Kap1 / ActinG, Gm13251, ETn, MERVK, Kap1

■ KAP1 WT
■ KAP1 KD
■ Liver KAP1 KO

C. chr3:88170706-88183241   12.5 kb
Ensembl Genes: Paqr6, Bglap3, Bglap2
Repeats: IAP-d-int
KAP1 IP [0-104], Input [0-104], ZFP932 IP1 [0-90], Input 1 [0-90], ZFP932 IP2 [0-230], Input 2 [0-230], Gm15446 IP1 [0-250], Input 1 [0-250], Gm15446 IP2 [0-150], Input 2 [0-150], H3K9me3 IP [0-26], Input [0-26]

D. IP/input (normalized to IAPEz): Gapdh, IAPEz, Primer 2, Primer 3 (ZFP/KAP1 peak), ZFP180
■ Control ■ ZFP shRNA2

E. Mean % DNA methylation: IAP 5LTR, IAP KAP1 peak
● Liver ctrl ■ Liver ZFP932/Gm15446 KD

F. IP/input: Gapdh, IAP Bglap, ZFP180 3'end
■ KAP1 ■ H3K9me3

G. Mean % DNA methylation: IAP 5LTR, IAP KAP1 peak
● Liver ctrl ■ Liver KAP1 KO

108

Figure S6, related to Figure 6. ZFP932/Gm15446 regulate the *Bglap3* locus in differentiated cells.

(**A**) GenomeBrowse screenshot of *Fbxw19* locus with coverage of stranded RNA-seq data in mESC and MEFs WT and KO for *Kap1*. Coverage in the sense strand is represented in green and in the anti-sense strand in blue. The orientation of the genes and TEs is represented with arrows. (**B**) Liver and C2C12 mRNA expression upon KAP1 depletion of two genomic regions with gene and TEs controlled by KAP1 in differentiated cells (liver samples are normalized to *Gapdh* and *ActinB*, and C2C12 samples to *ActinG* and *Tbp*). nd, not detected. AU, arbitrary units. (**C**) ChIP-seq density of ZFP932/Gm15446 and associated effectors on the *Bglap3* locus in C2C12 are illustrated. (**D**) ChIP-PCR of KAP1 in ZFP932/Gm15446 KD cells, normalized to IAPEz positive control. KD levels are of 65%. Primer 2 and 3 correspond to numbering present in Fig 4C. (**E**) Pyrosequencing of IAP Bglap in ZFP932/Gm15446 KD mouse liver, compared to control. (**F**) ChIP-PCR of KAP1 and H3K9me3 in mouse liver. (**G**) Pyrosequencing of IAP Bglap in *Kap1* KO mouse liver, compared to control. For DNA methylation analysis, mean CpG DNA methylation per sequence and per replicate was calculated. Error bars represent SD, * $p < 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, ns= not significant, Student's t test.

SUPPLEMENTAL TABLES

Table S1, related to Figure1. Sequences selected from ChIP-seq analysis to be tested in the screen.

Sequence ID	chromosome	start	end	size (bp)	TE	gene
P2	chr15	96071841	96072638	797	IAPEY_LTR, MIRc	ENSMUSG00000085460,ENSMUSG00000086809
P3	chr16	16373546	16374572	1026	RLTR44-int	-
P4	chr17	17449713	17450247	534	Lx4A, LTRIS2	-
P5	chr19	39721807	39722144	337	RLTR44-int	ENSMUSG00000062624
P6	chr2	24366644	24367245	601	LTRIS2	-
P7	chr2	28285331	28285800	469	LTRIS2	-
P8	chr4	43349284	43349917	633	RLTR9A3, MMERVK10D3_I-int, RLTR6-int	ENSMUSG00000028457
P9	chr4	43553863	43554400	537	B1F2	ENSMUSG00000028465
P10	chr4	55012441	55013092	651	L1MA5	ENSMUSG00000060206
P11	chr4	152526608	152527445	837	MTC-int	-
P12	chr4	155549797	155550889	1092	B1_Mur3, MT2B	ENSMUSG00000041936
P13	chr5	114074078	114074817	739	L2b, MTEa	ENSMUSG00000042190
P14	chr5	130162484	130163546	1063	-	ENSMUSG00000051034
P15	chr6	4873614	4874324	710	RLTRETN_Mm	ENSMUSG00000032827
P16	chr6	48546386	48547039	653	-	ENSMUSG00000052751
P17	chr7	26652988	26653803	816	RLTR44-int, ORR1E	-
P18	chr7	86361075	86361707	632	B2_Mm2, LTRIS2	-
P19	chr8	88210944	88211630	686	ORR1A3-int	ENSMUSG00000036902
P20	chr9	87086351	87086989	638	RLTR44-int	ENSMUSG00000056919

Table S2, related to Figure1. Library of KRAB-ZFPs and controls used in the screen, with index number.

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Plasmids and lentiviral vectors. For the screen, DNA targets were cloned into the pENTR/D/TOPO vector and transferred via gateway cloning into a pRRLSIN.cPPT.PGK-GFP.WPRE vector that was modified to contain R1-R2 gateway sites upstream the PGK-GFP cassette (pRRLSIN.cPPT.R1R2.PGK-GFP.WPRE). The KRAB-ZFP library was obtained partially from previous work (Gubelmann et al., 2013), and the remaining was codon optimized, synthetized into pENTR vectors, and further transferred via gateway cloning into a puromycin-selectable lentivector under a tetracyclin-inducible TRE promoter to obtain HA-tagged proteins (pSIN-TRE-R1R2-3xHA). pLKO.puro shRNA vectors or versions modified to contain the hygromycin resistance or tomato red cDNAs were used for KAP1 and ZFPs KD. The shRNA for *ZFP932/Gm15446* were obtained from the RNAi Consortium (http://www.broadinstitute.org/rnai/public/) or designed using the KRIBB siRNA AsiDesigner tool (http://sysbio.kribb.re.kr:8080/AsiDesigner/menuDesigner.jsf). For CRISPR-mediated KO, the lentiCRISPR vector (Shalem et al., 2014) was modified with an ubiquitin promoter in place of the EF1short promoter. sgRNA sequences were designed using the Zhang lab tool (http://crispr.mit.edu/). For KAP1 overexpression, pSicoR-KAP1-HA vector was used. Lentiviral vectors production protocols are detailed at http://tronolab.epfl.ch and backbones are available at Addgene (http://www.addgene.org).

Cell culture and cell-based assays. Mouse ES cells were cultured as previously described (Rowe et al., 2013a), using the *Kap1 loxP/loxP* ES3 line or its *Kap1* conditional KO derivative after transduction with a tamoxifen-inducible Cre vector (Rowe et al., 2010). For KO induction, cells were treated overnight with 1μM tamoxifen (Sigma) and collected at the latest 4-5 days later. *Kap1* KO and *loxP/loxP* control MEFs were generated and cultured as previously described (Rowe et al., 2013b). For KRAB-ZFP KO experiments, murine ES cells were transduced with an integrase defective lentiviral vector containing the pLentiCRISPR with an sgRNA for *Zfp932* and *Gm15446* or a control sgRNA against luciferase. Cells were selected for puromycin, and cloned from single cells. Two clones were selected and genotyped by PCR followed by TOPO cloning and Sanger sequencing. Clone 1 had a 909bp deletion in *Zfp932* disrupting exon 2 and a 4bp deletion in *Gm15446* (in exon 2); in clone 2 exons 2 and 3 of *Zfp932* and *Gm15446* were damaged. KRAB-ZFPs KD was induced with shRNA vectors for *Zfp932/Gm15446* on untransduced cells or on cells previously transduced with TRE vector containing LacZ control or ZFP for rescue experiments. For repression assays, cells were transduced at low multiplicity of infection (MOI), GFP fluorescence was assessed by flow cytometry, and ES cells pluripotency was monitored by anti-SSEA1 staining (BD Pharmingen, 560142). KAP1 KD was performed with shRNA vectors for *Kap1*. For KRAB-ZFPs ChIP experiments, C2C12 or ES cells were transduced with HA-tagged pSIN-TRE lentiviral vectors containing the codon-optimized KRAB-ZFP of interest, selected, and treated with doxycycline for 48h before harvesting. For KAP1 ChIP experiments in MEFs, MEF *Kap1* KO cells were complemented with HA-tagged KAP1 with levels similar to the endogenous protein. Cells were selected with 1μg/mL puromycin or 100 μg/mL hygromycin when necessary.

Mouse work. Lentiviral transgenesis was performed by perivitelline injection of pLKO.tomato vectors containing shRNA for *ZFP932/Gm15446* or empty control into fertilized oocytes (strain B6D2F1/J) that were transferred to foster mothers (strain NMRI). Lentiviral vectors for transgenesis were prepared as previously described (Barde et al., 2011), using Episerf medium (Invitrogen), the particle concentration obtained by p24 ELISA (PerkinElmer), and the infectious titer determined on HCT116 cells by tomato flow cytometry. Generated adults were genotyped by flow cytometry of peripheral blood cells and by quantitative PCR to assess the number of lentivector integration in the genome. *Kap1* KO and ZFP932/Gm15446 KD livers were harvested at 8 weeks of age.

Functional screen. DNA target sequences to be tested were chosen from the overlap of publicly available murine ESC ChIP-seq data for KAP1 (GSE41903), SETDB1 (GSE18371), H3K9me3 (GSE41903), and absence of ZFP57 (GSE31183) (Bilodeau et al., 2009; Quenneville et al., 2011; Rowe et al., 2013b). Nineteen sequences corresponding to KAP1 peaks were selected based on presence of TEs or of interesting KAP1 targets (such as 3'end of ZFP genes) (Table S2) and cloned into the pRRLSIN.cPPT.R1R2.PGK-GFP.WPRE vector. These sequences were tested for repression in murine ES and only repressed sequences were tested in the screen. PBS^{Lys1,2} sequence was also

cloned in the pRRLSIN.cPPT.R1R2.PGK-GFP.WPRE vector and tested in the screen. Lentiviral vectors with these sequences were used to transduce 293T cells at low MOI in order to favour integration of a single copy per cell. GFP positive cells were sorted and used in the screen. The ENSEMBL ID and amino acid sequences of the KRAB-ZFPs used to establish the screen (Table S1) were obtained from our previously published curation (Corsinotti et al., 2013) and cloned as described above. Reverse transfections and cell culture were done in triplicates in 96-well plates, using a Sciclone ALH 3000 (Caliper Life Sciences) liquid handling robot and a Multidrop Combi dispenser (Thermoscientific). For each well, $5x10^3$ cells in culture media were added to a 10μL mix containing 100-150 μg of KRAB-ZFP plasmid and 0.6 μl of Fugene 6 Transfection Reagent (Promega) in Opti-MEM (Life Technologies). All screen plates contained two LacZ plasmids and two empty wells as controls. For induction and selection of the TRE vector, doxycycline (final concentration 5 μg/mL) was added 16-20h after transfection, and on the following day cultures were supplemented with puromycin (final concentration 1 μg/mL). Cells were harvested on day 6 after transfection, resuspended in RIPA buffer containing protease inhibitors, and GFP was measured using a Tecan Infinite F500 plate reader (Tecan) (excitation at 485 nm and emission at 520 nm). Total protein content was quantified using BCA (BCA Protein Assay Reagent, Thermoscientific), and used to calculate normalized GFP fluorescence. Candidate hits were identified by selecting the 10 KRAB-ZFPs with the lowest normalized fluorescence values per 96-well plate, and only the ones that were present in all 3 replicates were considered. Hits were identified by transfection of the 293T cell line of interest with candidate KRAB-ZFPs in 24-well plates ($1.5x10^4$ cells/well transfected with 150-200 μg of DNA and 0.38 μL of fugene mix in Opti-MEM); doxycycline and puromycin were added at the same concentrations, with FACS readout after 6 days.

ChIP-PCR and ChIP-seq. Cells were harvested, washed with Episerf, fixed in 2mL per $1x10^7$ cells (10 min in 1% formaldehyde), quenched with glycine in 10mL (at 125 mM final), washed three times with PBS, and pelleted. Each pellet containing $1x10^7$ cells was lysed, resuspended in 1 mL of sonication buffer on ice (for KRAB-ZFPs: 10 mM Tris at pH 8, 1 mM EDTA, 0.1% SDS, and protease inhibitors; for all others: 10 mM Tris at pH 8, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% NaDOC, 0.25% NLS, and protease inhibitors), transferred to TC 12x12 tubes (Covaris), and sonicated (Covaris settings: 20 min, 5% duty cycle, 140W, 200 cycles). Sonication was assessed by reverse cross-linking (65°C, RNAse A at 1μg/μL, overnight), followed by DNA extraction. Fragment size (between 200-400bp) was checked on a Bioanalyzer (Agilent 2100). Immunoprecipitations were performed with chromatin from $1x10^7$ cells (for KAP1, histone marks, and PolII), or $4x10^7$ (for KRAB-ZFPs), with Dynabeads (ThermoFisher) in IP buffer (for KRAB-ZFPS: 10 mM Tris at pH 8, 1 mM EDTA, 0.1% SDS, 150 mM NaCl, 10% Triton X-100, and protease inhibitors; for all others: 16.25 mM Tris at pH 8.1, 137.5 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1.25% Triton X-100, and protease inhibitors) overnight. Chromatin was reversed cross-linked (65°C, Proteinase K at 400ng/μL, overnight) and DNA was further extracted for analysis. Antibodies used were HA (Covance, MMS-101P), KAP1 (Tronolab, rabbit polyclonal S23470), H3K9me3 (Diagenode, C15410056), H3K4me1 (Diagenode, pAb-037-050), H3K27ac (Abcam, ab4729), RNA PolII CTD repeat (Abcam, ab817), H3 (Abcam, ab1791). ChIP samples were used for SYBER Green qPCR (Aplied Biosystems) or library preparation for sequencing. Primers (were designed using Primer 3 (Untergasser et al., 2007) or GETPrime (Gubelmann et al., 2011). Libraries of immunoprecipitated chromatin and total input control from ChIP were performed with single-end adaptors as previously described (Rowe et al., 2013b). Sequencing was performed on an Illumina HiSeq 2500 (Illumina), with each library sequenced in 100-bp reads run.

RT-qPCR and RNA-seq. Total RNA was extracted and DNAse-I treated using a spin column-based RNA purification kit (Macherey-Nagel). Reverse transcription was performed with 500ng of RNA using random primers and SuperScriptII (Invitrogen). Primers were designed using Primer 3 (Untergasser et al., 2007) or GETPrime (Gubelmann et al., 2011), and used for SYBER Green qPCR (Aplied Biosystems). For mRNA sequencing, 100-bp single-end RNA-seq libraries were prepared using 200ng of total RNA and the Illumina TruSeq Stranded mRNA reagents (Illumina). Cluster generation was performed with the resulting libraries using the Illumina TruSeq SR Cluster Kit v4 reagents and sequenced on an Illumina HiSeq 2500 (Illumina).

Primers, shRNAs and sgRNAs.

Primers used in this study

Primer	Sequence	Purpose
Gapdh F	GCCCTTCTACAATGCCAAAG	ChIP-qPCR
Gapdh R	TTGTGATGGGTGTGAACCAC	ChIP-qPCR
P9 F	CTTGAGGCCAGCCAAGGA	ChIP-qPCR
P9 R	GCCTTAACAGCCTTACTTCTAGAATTG	ChIP-qPCR
P5 F	AGCCTTGGAAACAGGAACAG	ChIP-qPCR
P5 R	CACACTTTTGCCATCCTGTC	ChIP-qPCR
P8 F	ACACAATCTCCCCCCCTTTT	ChIP-qPCR
P8 F	GTGCCCCCTGTCCAGCTA	ChIP-qPCR
Bglap ChIP 1 F	CCCAGTGTCTGAAAGGGTAGG	ChIP-qPCR
Bglap ChIP 1 R	ATACTGGCCAACAGGAATGC	ChIP-qPCR
Bglap ChIP 2 F	TCATGGTGTCTGCTAGGTGTG	ChIP-qPCR
Bglap ChIP 2 R	TCAGAATCAGAGGCAACAGG	ChIP-qPCR
Bglap ChIP 4 F	TTGGTGCACTGTTTGACCTG	ChIP-qPCR
Bglap ChIP 4 R	AATAAGGTTCCCGGTCTTGG	ChIP-qPCR
Bglap ChIP 5 F	CAGCCCAACTGTGTGTTTTC	ChIP-qPCR
Bglap ChIP 5 R	ACATTTGGCCACGACCTATG	ChIP-qPCR
Bglap ChIP 6 F	TCTCTGATGTAAGCAGGAGGAG	ChIP-qPCR
Bglap ChIP 6 R	CAATCACCAACCACAGCATC	ChIP-qPCR
Bglap ChIP 7 F	CACACTGTACAAGAGGCTCCAG	ChIP-qPCR
Bglap ChIP 7 R	TTGTGCTGGAGTGGTCTCTATG	ChIP-qPCR
ZFP180 3'end F	CCGTACAGGTGCAATCTGTG	ChIP-qPCR
ZFP180 3'end R	GTTTGTAGCTCTGGCGGAAC	ChIP-qPCR
Gapdh F	TCCATGACAACTTTGGCATTG	RT-qPCR
Gapdh R	AGTCTTCTGGGTGGCAGTGA	RT-qPCR
Actin F (ActinB)	CTAAGGCCAACCGTGAAAAGAT	RT-qPCR
Actin R (ActinB)	CACAGCCTGGATGGCTACGT	RT-qPCR
ActinG F	TGGATCAGCAAGCAGGAGTATG	RT-qPCR
ActinG R	CCTGCTCAGTCCATCTAGAAGCA	RT-qPCR
TBP F	TTGACCTAAAGACCATTGCACTTC	RT-qPCR
TBP R	TTCTCATGATGACTGCAGCAAA	RT-qPCR
btub F	GCAGTGCGGCAACCAGAT	RT-qPCR
btub R	AGTGGGATCAATGCCATGCT	RT-qPCR
Bglap3 F	CTGACAAAGCCTTCATGTCC	RT-qPCR
Bglap3 R	TCAAGCTCACATAGCTCCC	RT-qPCR
IAP Bglap F/Bglap ChIP 3 F	AGGTGTTGCAGAGGTTTTGG	RT-qPCR and ChIP-qPCR
IAP Bglap R/Bglap ChIP 3 F	AATATCGGACACAGGGCAAG	RT-qPCR and ChIP-qPCR
Rgs20 F	CTACTTGTGGCCTCAATGG	RT-qPCR
Rgs20 R	GACAGTGAGGCAAGAACAG	RT-qPCR
Rgs20 IAPEY F	GTTCCTGCAAAACAGACTGC	RT-qPCR
Rgs20 IAPEY R	ATTGCTCTGGTCAGCCATTC	RT-qPCR
Rgs20 MERVK F	ACTGGAGGTCCTTGTCCTATG	RT-qPCR

Rgs20 MERVK R	AGGTTCGATGTGCTCTTTCC	RT-qPCR
ZFP932/Gm15446 F	TTGCACATCATTGTCATCTCC	RT-qPCR
ZFP932/Gm15446 R	CTGACCTACAAAGGCTTTACCAC	RT-qPCR
MERVK10C F	GCCCCTCAATTGGTAGAATG	RT-qPCR
MERVK10C R	TTTCCGGCAGTCTCTAATGC	RT-qPCR
MERVK10C_2 F	CAAATAGCCCTACCATATGTCAG	RT-qPCR
MERVK10C_2 R	GTATACTTTCTTCTTCAGGTCCAC	RT-qPCR
KAP1 F	CGGAAATGTGAGCGTGTTCTC	RT-qPCR
KAP1 R	CGGTAGCCAGCTGATGCAA	RT-qPCR
chr12_ORR1A0 F	GGTTGGAATGGGTGTTTCAC	RT-qPCR
chr12_ORR1A0 R	TCGTCCAACTTTCCAAGTCC	RT-qPCR
chr12_MMERGLN F	ACCCACAGTCTGCAAAATCC	RT-qPCR
chr12_MMERGLN R	AGTGATGCGGATTCCAACTC	RT-qPCR
Fbxw19 F	TGTGTACGTGTGGGAGGAGA	RT-qPCR
Fbxw19 R	AGAAAGCAGGGAATGGGACT	RT-qPCR
Fbxw19 ERVB4 F	TTAAAGCAGGGGAGGTGTTG	RT-qPCR
Fbxw19 ERVB4 R	GACCCCCTTTTCTTTTCTGG	RT-qPCR
Gm13251 F	GATGTGAAGTGTGCTTCGAGT	RT-qPCR
Gm13251 R	CACAACAGGACCAGACACCA	RT-qPCR
Gm13251 Etn F	TAATCTTTGGGCCAGGACTC	RT-qPCR
Gm13251 Etn R	CCAAAGAAATGCCACACCTG	RT-qPCR
Gm13251 MERVK F	AGCCCTTGGGATGATAACAG	RT-qPCR
Gm13251 MERVK R	GGATAACGCAATGCTGTGTG	RT-qPCR
Gm6871 F	ACCTACAGGAATCTCACCAC	RT-qPCR
Gm6871 R	GTTTGGTGCCTTCCATGTC	RT-qPCR
IAP-d F	CAGCTGAACACAATCACTCATC	RT-qPCR
IAP-d R	TCCAGTGCGGGAATCTATG	RT-qPCR
IAPEz F	CTTGCCCTTAAAGGTCTAAAAGCA	RT-qPCR and ChIP-qPCR
IAPEz R	GCGGTATAAGGTACAATTAAAAGATATGG	RT-qPCR and ChIP-qPCR
LINE F	TTTGGGACACAATGAAAGCA	RT-qPCR
LINE R	CTGCCGTCTACTCCTCTTGG	RT-qPCR
Pdlim3 F	AACCACAGGAATTCAAACCC	RT-qPCR
Pdlim3 R	TGTCATCAATGTTTGCTGCT	RT-qPCR
Setdb1 F	GATGTCCCCACTTCCTCTGA	RT-qPCR
Setdb1 R	GCATAGCTACGCCACACTGA	RT-qPCR
KVDMR F	GGGGTTTAAAGGGTTTTAAGATTAT	pyrosequencing
KVDMR R	TCATAACCTCCCCCTCCTC	pyrosequencing
KVDMR Seq	TGTAAGTTTGGGTTATAAAGA	pyrosequencing
IAP 5LTR F	GGAATAGTAGTTTATTGGTTAGAATTAT	pyrosequencing
IAP 5LTR R	CCCCACCACTCCTACTTACATCAAT	pyrosequencing
IAP 5LTR Seq	GGTTAGAATTATTATTTGTTATATG	pyrosequencing
KAP1 peak F	GGTTGGTGATAAGTTTAGGGAGTTT	pyrosequencing
KAP1 peak R	TACTAACAACTATCCCCCTCT	pyrosequencing

KAP1 peak Seq	GGTGATAAGTTTAGGGAGTTTTA	pyrosequencing
bisMERVK10C.1F	ATAGTTTAATTTAAGATATGGGGTT	bisulfite
bisMERVK10C.1R	ACAATAATCAATACCACTCTACAAC	bisulfite
GAG F	GGAGCTAGAACGATTCGCAGTTA	qPCR - mice genotyping
GAG R	GGTGTAGCTGTCCCAGTATTTGTC	qPCR - mice genotyping
GAG T (probe)	ACAGCCTTCTGATGTTTCTAACAGGCCAGG	qPCR - mice genotyping
Albumin F	GCTGTCATCTCTTGTGGGCTGT	qPCR - mice genotyping
Albumin R	ACTCATGGGAGCTGCTGGTTC	qPCR - mice genotyping
Albumin T (probe)	CCTGTCATGCCCACACAAATCTCTCC	qPCR - mice genotyping
ZFP932 F	TGACTTTTTAAAATAAGGGAACAACTG	CRISPR genotyping
ZFP932 R	TGGCTCACATTTGTAGCATCA	CRISPR genotyping
Gm15446 F	TTTCTCTCTCCCTTTCTCTCTCC	CRISPR genotyping
Gm15446 R	CATGGCTCACATTTGTACATCA	CRISPR genotyping
ZFP932/Gm15446 F	TGCAACATATATCCTATACAAAGAGCA	CRISPR genotyping
ZFP932/Gm15446 R	CATGCCTGCAAAGATAATTGG	CRISPR genotyping

shRNAs and sgRNAs used in this study

shRNA or sgRNA	sequence (hairpin or sgRNA with PAM)	Target
shRNA1	CCGGCCTCTCATGGTCAACTTCAAACTCGAGTTTGAAGTTGACCATGAGAGGTTTTTG	ZFP932/Gm15446
shRNA2	CCGGGCCTATTCACGACACAGCATTCTCGAGAATGCTGTGTCGTGAATAGGCTTTTTG	ZFP932/Gm15446
shRNA3	CCGGTATGATAACGTCTTCACATATCTCGAGATATGTGAAGACGTTATCATATTTTTG	Gm6871
shRNA4	CCGGAACGTCTTCACATATCACAGTCTCGAGACTGTGATATGTGAAGACGTTTTTTTG	Gm6871
shKAP1	CCGGCCGCATGTTCAAACAGTTCAACTCGAGTTGAACTGTTTGAACATGCGGTTTTTG	KAP1
shSETDB1	CCGGGCCTTGATCTTCCATGTCATTCTCGAGAATGACATGGAAGATCAAGGCTTTTTG	SETDB1
sg932Gm	TGTGAAAGCTCCAGAAGACATGG	ZFP932/Gm15446
sgLuc	ACGCTGGGCGTTAATCAAAGAGG	luciferase (control)

Bioinformatics analyses

Merged repeats track. All TE analyses were performed using a merged repeats track generated in-house. As the RepeatMasker annotation often contain TEs fragmented into smaller regions that not always correspond to their biological organization, we generated a new annotation that merges some of this fragments based on ERVs structure. The RepeatMasker 3.2.8 (Repeat Library 20090604, http://www.repeatmasker.org/species/mm.html) was used as basis, and we computed a frequency table of LTRs surrounding internal ERV ("ERV-int") parts for each ERV family. To infer the significance of the LTR-int associations we computed the frequencies distributions of the putative LTR for each "-int", and performed a Wald test to assign a *P* value to each LTR-int pair. The LTR was attributed to an "-int" family when the *P* value was smaller than 0.001, the frequency of occurrence next to an "-int" was higher than 2%, and it was annotated as an "LTR". ERV-int integrants that shared the same name, or attributed LTRs, were merged with their neighbouring elements when the distance was shorter than 400 bp and there were no SINEs or LINEs in between. We also kept the fragmented information in our new annotated list. For the final track, the other families of repeats were not modified and were added to the merged ERV/LTR elements.

ChIP-seq analyses. Reads were mapped to the mouse genome assembly mm9 using Bowtie 2 short read aligner (Langmead and Salzberg, 2012), using the --sensitive-local mode. The peaks were called using either the MACS program v1.4.2.1 (Zhang et al., 2008) or the SICER software

v1.1 for histone modification marks (Zang et al., 2009), with the total input chromatin coverage as control. For MACS we used the default software parameters and selected MACS score above 100. For SICER we used the recommended parameters for histone marks (redundancy threshold: 1, window size: 200, fragment size: 150, effective genome fraction: 0.74, gap size: 400, FDR: 0.01). For KAP1 in ES cells, PeakSplitter was also used (Salmon-Divon et al., 2010) with the default software parameters. For KRAB-ZFPs, all ChIP-seq experiments were performed in duplicates and only the overlap of peaks of both duplicates was considered for the analyses. For KAP1 ChIP-seq in MEFs, experiments were performed in duplicates and the union of both replicates was used for further analyses. Bigwig tracks were normalized to reads per 100 million mapped reads. Repeats enrichment of Fig 3C and Fig S5 were calculated from the overlap of repeats with ZFP932 or Gm15446 peaks containing KAP1 using the intersect tool (with default parameters) from the BEDtools 2.25.0 software (Quinlan and Hall, 2010) (minimum of 1 bp overlap). As control, we randomly assigned the peaks in the genome and used them in overlaps with repeats (this process was repeated 100 times). Motif search was performed using all ZFP932 and Gm15446 peaks with RSAT (Medina-Rivera et al., 2015), using unbound ERVK sequences as background control.

Coverage plots. ChIP-seq signal on each feature of interest were extracted from the bigwig file, and each sequence signal was scaled between 0 and 1. Features were aligned on their 3' end, and plotted on a heatmap using the matplotlib library of Python. For coverage plots with alignments, sequences were multiply aligned using the MAFFT software v. 7.245 (auto parameters and gap extension penalty of 0.123), and the coverage of each sequence was plotted on top of the alignment normalizing it as before. When gaps were created by only one sequence, it was removed from the alignment prior to plotting. Each consensus was then added to the corresponding alignment.

ERV annotation. For MMERVK10C-int, consensus sequence and coordinates were taken from Repbase (http://www.girinst.org/repbase/). For IAPez, consensus sequence was taken from Repbase. Coordinates for IAPez were estimated from an alignment with IAP1-MM_I (using IAP1-MM_I Repbase coordinates) and were checked using previous publications (Carmi et al., 2011). IAP Bglap LTR annotation was done by comparison with previously annotated IAP LTRs (Christy et al., 1985).

RNA-seq analysis. The RNA-seq reads were mapped to the mm9 genome using TopHat (Kim et al., 2013) with the following parameters: --b2-sensitive --no-novel-juncs --no-novel-indels. These settings allow multimapped reads to be randomly assigned once among the mapped loci. Gene counts were generated using the HTseq-count program with default parameters. TE counts were computed using the multicov script from the BEDtools software with the -split option. Only genes or TEs that had at least as many reads as samples present in the analysis were considered further. Sequencing depth normalization and differential expression analyses were performed using the voom function from the R package LIMMA (Law et al., 2014) from Bioconductor (Gentleman et al., 2004). The gene library sizes as given by voom were used to normalize the TEs counts. To be considered significantly upregulated, a gene or a TE had to have 2-fold increased expression (P value adjusted $\leq$ 0.05 for *Kap1* KO/KD; P value $\leq$ 0.01 for *Zfps* KO). P values were computed using a moderated t-test and corrected for multiple testing using the Benjamini-Hochberg method (Benjamini and Hochberg, 1995). MA plots were generated from normalized count values as given by VOOM and plotted using the matplotlib library of Python.

Public sequencing data. Raw reads were downloaded from publicly available ChIP-seq (GEO IDs: GSE41903, GSE18371, GSE31183, GSE48519, GSE62664, GSE58323) and RNA-seq data (GEO IDs: GSE62664). Mapping and data processing was done as described above. CAGE peaks based expression table for mouse of FANTOM 5 data was downloaded from the FANTOM consortium website (http://fantom.gsc.riken.jp), and their TSS peaks or max count were obtained from UCSC genome browser public tracks (Lizio et al., 2015). Representative samples were selected and expression in TPM (tags per million) was plotted using R software (http://www.R-project.org).

Heatmaps and boxplots. Heatmaps were generated with R, using the heatmap.2 function. Both rows and columns were clustered using hierarchical clustering. The agglomeration method was "complete", and the distance metric used was Pearson distance. For heatmaps of ERV expression, only integrants that had 10 or more reads on average per set of replicates were included. Boxplots were generated using R, and binding to transcription factors was determined using BEDtools with a minimum of 1 bp overlap.

SUPPLEMENTAL REFERENCES

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society. Series B (Methodological) *57*, 289-300.

Carmi, S., Church, G.M., and Levanon, E.Y. (2011). Large-scale DNA editing of retrotransposons accelerates mammalian genome evolution. Nature communications *2*, 519.

Christy, R.J., Brown, A.R., Gourlie, B.B., and Huang, R.C. (1985). Nucleotide sequences of murine intracisternal A-particle gene LTRs have extensive variability within the R region. Nucleic acids research *13*, 289-302.

Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J.*, et al.* (2004). Bioconductor: open software development for computational biology and bioinformatics. Genome biology *5*, R80.

Gubelmann, C., Gattiker, A., Massouras, A., Hens, K., David, F., Decouttere, F., Rougemont, J., and Deplancke, B. (2011). GETPrime: a gene- or transcript-specific primer database for quantitative real-time PCR. Database : the journal of biological databases and curation *2011*, bar040.

Gubelmann, C., Waszak, S.M., Isakova, A., Holcombe, W., Hens, K., Iagovitina, A., Feuz, J.D., Raghav, S.K., Simicevic, J., and Deplancke, B. (2013). A yeast one-hybrid and microfluidics-based pipeline to map mammalian gene regulatory networks. Molecular systems biology *9*, 682.

Quenneville, S., Verde, G., Corsinotti, A., Kapopoulou, A., Jakobsson, J., Offner, S., Baglivo, I., Pedone, P.V., Grimaldi, G., Riccio, A.*, et al.* (2011). In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. Molecular cell *44*, 361-372.

Salmon-Divon, M., Dvinge, H., Tammoja, K., and Bertone, P. (2010). PeakAnalyzer: genome-wide annotation of chromatin binding and modification loci. BMC bioinformatics *11*, 415.

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

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

V. Conclusions and perspectives

Although TEs represent almost half of our genome, there is still a great gap of knowledge regarding how they impact our biology and how they are regulated. In the last decades we have learned that these elements can have beneficial or detrimental effects for the host. For these reasons, the host has developed several mechanisms of TE control, and one of the most important ones involves KRAB-ZFPs and KAP1. In this thesis we have studied the KRAB/KAP1-mediated regulation of TEs in pluripotent and somatic cells. Understanding how these elements are regulated in various tissues helps us comprehend vertebrate evolution, development, and physiology.

1. A large scale screen to identify KRAB-ZFPs binding to specific sequences

We have developed a large-scale screening system to identify KRAB-ZFPs interacting with specific DNA targets, with emphasis on TE sequences. Our method confidently identified ZFP809 amongst the 211 tested murine KRAB-ZFPs as the one specifically interacting with its previously known target in the MLV genome (PBS^{Pro}). The screen further allowed the discovery of two *bona fide* novel KRAB-ZFPs – Gm6871 and ZFP932, bound to SINE/LINE and ERVK sequences, respectively. These results thus show that our screen is an efficient method for the identification of KRAB-ZFP/DNA targets pairs, opening the way to functional analyses with this major class of transcriptional repressors.

Identifying proteins bound to DNA has been a challenge for many years. The availability of sequenced genomes, coupled with high-throughput sequencing

technologies allowed many advances, but there are still several difficulties (Deplancke, 2009; Guillen-Ahlers et al., 2014). ChIP-seq is one of the main methodologies to evaluate protein-DNA interactions, allowing the characterization of *in vivo* interactions between proteins and specific DNA sequences in the genome (Schones and Zhao, 2008). It is a powerful technique that enables the identification of genome-wide DNA loci bound by a certain protein of interest. However, this technique requires one to have a specific protein to then look for its targets, like many others called protein-centered methods. Most of the current techniques that allow high-throughput approaches are protein-centered (such as ChIP-seq, SELEX, and MITOMI-seq), likely due to the myriad range of technologies available for DNA amplification and sequencing.

On the other hand, finding the protein that regulates a specific DNA sequence can be more challenging. The so-called DNA-centered techniques include electrophoretic mobility shift assay (EMSA), DNA affinity purification followed by mass spectrometry (such as PiCh and ChAP-MS), and yeast one-hybrid assays, and are less adaptable to large-scale approaches (Dey et al., 2012; Simicevic and Deplancke, 2010). A good example of DNA-centered technique that can be successfully applied at high-throughput scale is the yeast one-hybrid screen (Deplancke et al., 2004; Hens et al., 2011; Reece-Hoyes et al., 2011). Techniques combining DNA affinity purification with mass spectrometry are promising and evolving fast (including some with the use of CRISPR proteins to direct DNA purification), however they are still limited to the sensitivity of mass spectrometry, which restricts good performances to a few types of repetitive sequences or organisms (Byrum et al., 2012; Dejardin and Kingston, 2009; Fujita and Fujii, 2013; Ide and Dejardin, 2015). Nonetheless, none of these methods are based on a functional approach. As a consequence, they are more susceptible to unspecific interactions and false positives when compared to our screen.

While the main advantages of our system are the facts that it is DNA-centered and functional, one of its disadvantages is its limitation solely to KRAB-ZFPs or other repressive factors. The yeast one-hybrid screen, as it consists on a score of the binding interaction between the transcription factor and the DNA sequence,

can be used for a wider variety of regulators. Notwithstanding, our methodology can be adapted to other repressive regulators and other species, which broadens its applications. In fact, we obtained a complete human library of KRAB-ZFPs and our method could be optimized also in this setting.

Finally, our screen could be further improved by decreasing the level of noise observed in the results at the plate reader step. We tried different optimizations and modifications to decrease the noise (such as different transfection reagents and different reporter proteins), but we were not successful in decreasing the noise further. Despite that, the current settings allowed us to confidently identify several candidates, with a success rate of 30-40%. Even if we may miss some true positives, the subsequent FACS validation step included in the process is very stringent and so far excluded all false positives. The use of other reporter proteins (such as luciferase) or different transfection reagents that increase transfection efficiency could further ameliorate the plate reader stage.

Altogether we have developed an efficient tool to identify KRAB-ZFPs binding to specific DNA sequences. Our screen could be applied to several murine DNA targets, ranging from interesting classes of TEs to specific integrants of interest (such as the agouti IAP). It could be further adapted to human or other species in order to characterize, for instance, specific KRAB-ZFPs regulating TEs associated with diseases. While large-scale ChIP-seq screens have been established to identify KRAB-ZFPs genome-wide targets (*M. Imbeault, personal communication*, and Najafabadi et al., 2015), ours is the first large-scale effort identifying TE/KRAB-ZFP pairs with DNA target as bait, opening the way to the characterization of this important family of transcription factors.

2. KRAB/ KAP1 and the evolutionary arms race hypothesis

One of the candidates obtained in the screen was Gm6871, initially identified to repress a B2/SINE sequence and later identified to regulate LINE-1s in murine ES cells. We incorporated this candidate and its characterization – as well as the

subsequent evaluation of KAP1 role in regulating LINE-1s in mouse – into another project of the lab, which investigated the role of KRAB/KAP1 in the control of LINE-1s in human ES cells. In this work, we showed for the first time that KAP1 and KRAB-ZFPs are responsible for regulating LINEs. Also, it was the first extensive demonstration that KRAB-ZFPs regulate endogenous retroelements. Collectively, we showed that while KRAB-ZFPs and KAP1 regulate ancient LINE-1s, DNA methylation (likely via piRNAs) controls young integrants. These results suggest an evolutionally dynamic regulation of LINEs in pluripotent cells and support the hypothesis of an arms race between KRAB-ZFPs and TEs.

Previous studies performed in mouse established KAP1 role in controlling ERVs, but its function in the regulation of LINEs remained unclear (Matsui et al., 2010; Rowe et al., 2010). Upon *Kap1* KO in murine ES cells, only a modest increase of 3-fold upregulation had been observed for LINE-1s, while IAPs had more than 60-fold augmented expression. Also, another group reported that shRNA-mediated depletion of ZFP819 led to higher expression of LINE-1s, IAPs, and SINEs. As no binding evidence was demonstrated for non-LTR elements, it remained unknown whether KRAB/KAP1 regulate LINE-1s. Our work, hence, established that KRAB-ZFPs together with KAP1 represses this class of TEs in stem cells. Our results were soon corroborated by the demonstration that ZNF93 controls LINE-1s in human ES cells. It is important to know that the effect observed by us and others on LINE-1 expression are still mild, corroborating the hypothesis that other regulatory layers repress these elements.

Together with the work performed on ZNF93, our data strongly suggest the existence of an evolutionary arms race between KRAB-ZFPs and TEs (Jacobs et al., 2014). The abundance of KRAB-ZFPs in vertebrate genomes, associated with a correspondence of these proteins and TE numbers across genomes, were some of the first evidence to support a co-evolution hypothesis (Thomas and Schneider, 2011). The amplification pattern of LINE-1 families was also in line with this idea. Differently from ERVs, LINE-1s amplify in an unusual way, alternating periods of high and low activity and with normally only one family active at a time: a family of elements emerges, amplifies in the genome, becomes

extinct, and is replaced by a more recent one (Khan et al., 2006; Sookdeo et al., 2013). This pattern led many groups to suggest this was due to a co-evolutionary arms race with the host. Our work, and particularly the one of Jacobs and colleagues, finally provided evidence of such fact (Jacobs et al., 2014).

On our side, we characterized Gm6871, a mouse-specific KRAB-ZFPs that regulates mainly elements of the murine-specific L1MdF2 family. At the same time, we observed that KAP1 was particularly enriched at L1MdF and L1MdF2 elements, families that emerged between 5.6 and 3.8 million years ago. In human ES cells, KAP1 targeted primate-specific LINE-1 families, while young human-specific elements were regulated by DNA methylation – possibly via piRNAs mechanisms as previously reported (Marchetto et al., 2013). The identification of a mouse-specific KRAB-ZFP regulating LINE-1s and the dynamic of KAP1-binding to the different families are strongly suggestive of the evolutionary arms race between KRAB-ZFPs and LINEs.

The results of Jacobs and colleagues regarding ZNF91/91 and LINEs/SVAs consolidated our hypothesis (Jacobs et al., 2014). After observing increased expression of primate-specific retrotransposons in murine ES cells containing a copy of the human chromosome 11, they screened for primate specific KRAB-ZFPs and identified ZNF91 and ZNF93 as regulators of SVAs and L1PA3-6 elements, respectively. They went on to characterize the evolutionary history of ZNF93 and its LINE-1 targets. Not only the KRAB-ZFP underwent structural changes allowing the recognition of its TE targets, but also LINE elements progressively lost the ZNF binding site, escaping the restriction imposed by the repressor. These data elegantly demonstrate the presence of an ongoing evolutionary arms race between KRAB-ZFPs and TEs.

The other candidate we obtained in our screen was ZFP932, which led us to the discovery of its paralog Gm15446. We not only characterized these mouse-specific KRAB-ZFPs role in regulating ERVKs and neighboring genes in ES cells, but we further discovered that these regulators and their TE targets control gene expression in somatic cells as well. Some of the data we collected about these proteins also advocate for the arms race hypothesis. The existence of two KRAB-

ZFP paralogs structurally so similar and with overlapping but distinct set of targets is in accordance with an ongoing co-evolution between KRAB-ZFPs and TEs. MERVK10C-int elements, for instance, were almost exclusively bound by Gm15446. We could speculate that *Gm15446* derived from a duplication event from *Zfp932* gene, and accumulated mutations that allowed it to repress a new set of elements. It would be interesting to date the two paralogs and their different targets in order to reconstruct their evolutionary history. We have tried to do such characterization, but the fact that both proteins and TEs are mouse-specific, coupled with the low availability of related genomes, challenged our quest. Also, it would be informative to know which of the zinc fingers of ZFP932 and Gm15446 are responsible for the binding to the different families.

3. KRAB/KAP1-mediated regulation of TE/gene expression in adult cells and the co-option hypothesis

The most important implication of the results we obtained from ZFP932/Gm15446 was the discovery that the KRAB/KAP1 complex regulates TEs and secondarily genes in somatic cells. Our results suggest that TEs and KRAB-ZFPs establish transcriptional networks that regulate not only development but probably also many physiological events. This view further supports the hypothesis that KRAB-ZFPs would participate in the co-option of TEs by the host.

Our observations invalidate the current model that KRAB-ZFPs/KAP1 irreversibly silence TEs in early embryogenesis and are then dispensable in adult cells (Maksakova et al., 2008; Mikkelsen et al., 2007; Rowe and Trono, 2011; Wolf et al., 2015). In this model, KRAB-ZFPs bind specifically to the TE in early embryo, interact with KAP1, which then recruits chromatin-modifying enzymes, leading to repressive chromatin changes such as H3K9me3 and ultimately DNA methylation. Upon differentiation, once these TE sequences are DNA methylated, they would be permanently silenced, and the KRAB/KAP1 complex would be dispensable (Mikkelsen et al., 2007; Rowe et al., 2013a; Walsh et al., 1998).

Only recently, some evidence supported the possibility that KRAB/KAP1 could regulate TEs also in differentiated cells. In *Kap1*-depleted neuronal progenitor cells, we observed upregulation of some families of ERVs in KO cells when compared to control (Fasching et al., 2015). Neurons are known to be more permissive to retrotransposons, so those cells were thought to be an exception at the time. Also, as KAP1 binding was not assessed, a direct role could not be confirmed. Moreover, we had previously observed KAP1 bound to TEs in human CD4+ cells, and increasing evidence showed that many KRAB-ZFPs were expressed in differentiated cells (Barde et al., 2013; Lizio et al., 2015). The most compelling evidence came from a report that *Setdb1*-depleted B lymphocytes have increased expression of ERVs (Collins et al., 2015). Together with these data, our work establishes a role for the KRAB/KAP1 complex in repressing TEs also in differentiated cells.

Previous studies that characterized the transcriptome of KAP1-depleted adult cells registered no changes in expression of TEs, including cells in which we now observe deregulation, such as liver and MEFs (Bojkowska et al., 2012; Rowe et al., 2010; Rowe et al., 2013b; Santoni de Sio et al., 2012). We believe this difference is due to the advances in sequencing technologies (longer reads and better quality) and the improvement of our ability to map single TE integrants instead of bulk populations. As the impact of KAP1-depletion on TEs is smaller in differentiated than in pluripotent cells, we believe that looking at single integrants versus family consensus can make a difference in this case.

Another possible explanation for the smaller number of deregulated TEs in KAP1-depleted adult cells could be the need for activating factors for transposon expression. This would also clarify why (1) not all KAP1-bound TEs are upregulated upon depletion of the co-repressor, and (2) different subsets of TEs are activated in different cells. In the study of *Setdb1*-depleted B lymphocytes, Collins and colleagues observed that only some of the epigenetically de-repressed ERVs are actually transcriptionally active. The transcribed TEs were bound by PAX5, a transcription factor important for B cells (Collins et al., 2015). We speculate that for KRAB/KAP1 the scenario would be similar, and presence

or absence of an activator would dictate TE expression upon removal of the repressor.

Our results further provide insights on the repressive mechanisms of the KRAB/KAP1 complex in somatic tissues. In somatic cells KRAB/KAP1 repression of ERVs also occurs through H3K9me3, via SETDB1, but without deposition of DNA methylation. Hence, in differentiated cells histone-mediated repression is not dispensable as initially thought (Mikkelsen et al., 2007). More interestingly, we showed that in differentiated cells ERVs are expressed even if highly DNA methylated, as it was the case of the IAP Bglap. Methylation is commonly perceived as a permanent repressive mark and adult cells are believed to keep TEs repressed mainly via DNA methylation. However, several emerging evidence show that this mark is probably more diversified than initially envisioned (Jones 2012, Schubeler 2012). Of note, we previously observed that in human ES cells, transcription can occur from highly methylated KAP1-controlled promoters; also in *Setdb1* KO B lymphocytes some elements are expressed despite remaining DNA methylated (Collins et al., 2015; Turelli et al., 2014). We believe DNA methylation does keep many TEs silenced in adult cells, but our results suggest that the regulation of these elements is complex and likely composed of several intertwined layers of control.

It is important to note, however, that our methylation and expression analysis were done in bulk populations. Hence, we cannot ensure that the methylation and expression status are the same in a given cell, or if, for instance, a few cells that are demethylated could be responsible for the observed upregulation. The best way to investigate this question would be to perform single cell analysis of expression and DNA methylation in the same cells, which is challenging with the current technologies. Another possibility would be to use reporter systems and transgenesis experiments *in vivo* in order to differentiate the cells, sort them, and observe both methylation and expression. Investigating further this mechanistic question would be certainly interesting and would challenge our perception of DNA methylation as a permanent repressive mark.

As explored in the introduction of this thesis, there are several examples of co-option of TEs by the host (reviewed in Rebollo et al., 2012c; Warren et al., 2015). There are fascinating cases of TEs acting as enhancers or promoters of cellular genes, or even co-option of full transposon genes. Several cases have been associated with tissue-specific transcription factors, such HERV-E insertion upstream amylase genes or RLTR13D5 ERVs as placental-specific enhancers (Chuong et al., 2013; Ting et al., 1992). However, KRAB-ZFPs have never been implicated in this process. In our work, we show that also in differentiated tissues KRAB-ZFPs can modulate the expression of TE-driven gene expression. These proteins are broadly expressed in somatic cells, and recent evidence demonstrates that they are mainly implicated in binding to TEs (*M. Imbeault, personal communication;* Lizio et al., 2015; Najafabadi et al., 2015). Hence, it is tempting to speculate that the KRAB/KAP1 complex participates in different physiological processes, via co-option of TEs.

We have actually generated *in vivo* data that could support further the co-option hypothesis. In the course of our experiments, we have generated ZFP932/Gm15446 KD mice, and about 50% of the born mutants had a kinked-tail phenotype (**Figure 7**). This phenotype presented incomplete penetrance and was confirmed only with one shRNA sequence, but it is promising as it has been associated with IAP deregulation in multiple cases. The axin-fused mouse is characterized by the insertion of an IAP inside the *Axin* gene, and transcripts generated from the 3'UTR of this IAP disrupt protein function, causing the kinked-tail phenotype (Vasicek et al., 1997). Another group has correlated the presence of the kinked-tail phenotype with demethylation of IAPs and MERV-L elements (Ramirez et al., 2006).

A

Control

79%, 4.2cp

Zfp932/Gm15446 KD

2%, 0.1cp

61%, 4.3cp

8%, 0.4cp

14%, 0.5cp

11%, 2.2cp

0.2%, 1.3cp

B

Control   *Zfp932/Gm15446* KD

C
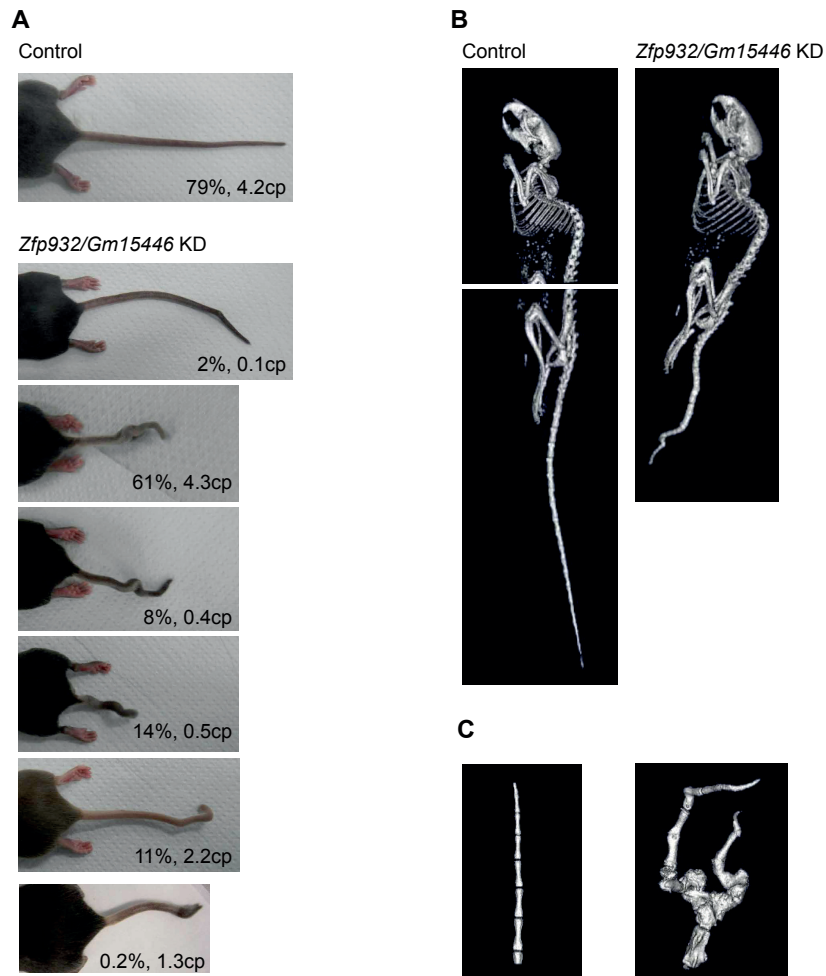
Figure 7. Some Zfp932/Gm15446 KD mice display a kinked tail phenotype. Transgenic animals were generated by lentiviral vector transduction of oocytes with Zfp932/Gm15446 shRNA or control vector. (**A**) Picture of the tail of the transgenic animals. Percentage of tomato positive cells (proxy for vector expression) detected in the blood and *gag* copy number in each animal is present in the picture. (**B**) and (**C**) Micro-CT x-ray imaging of control and KD animals.

Bglap3 is one of the three genes of the *Bglap* cluster in the mouse (Desbois et al., 1994). Bglap stands for "bone gamma-carboxyglutamic acid-containing protein", and BGLAP1 (also called osteocalcin) is an osteoblast-specific protein that binds to calcium and hydroxyapatite. Osteocalcin acts as a hormone, with implications for bone metabolism and formation, glucose homeostasis, and male fertility

127

(Ducy et al., 1996; Ferron et al., 2010; Lee et al., 2007; Oury et al., 2011). *Bglap3* functions, however, are not well defined, and while some report it could have functions different from *Bglap1*, both proteins are structurally very similar (Desbois et al., 1994; Petrucci et al., 2006; Sato and Tada, 1995). In view of these facts, we could speculate that ectopic expression of a protein so similar to osteocalcin could lead to bone formation defects, as the ones observed in the kinked-tail mouse. Our efforts to correlate *Bglap3* overexpression with the kinked-tail phenotype so far were not successful. Generation of *Zfp932/Gm15446* KO mice and characterization of the animals would be ideal to investigate (1) if the kinked-tail phenotype is specific of ZFP932/Gm15446 depletion, and (2) if IAP-mediated ectopic expression of Bglap3 is associated with the phenotype. Interestingly, ZFP932 could be implicated in osteoblast differentiation, as overexpression of this protein inhibits the differentiation of murine mesenchymal stem cells towards osteoblasts (Huang et al., 2012). We are also investigating this possibility and evaluating the role of the IAP-mediated *Bglap3* overexpression in this process. If proven correct, this would strengthen the hypothesis that KRAB-ZFPs participate in a domestication process of TEs by the host and are involved in species-specific regulatory networks.

Another interesting fact that is in accordance with the domestication of TEs is our observation that KAP1 binds to multiple places, possibly via different KRAB-ZFPs, on TEs. This could be due to the variety of transcriptional features present in these elements, which are true transcriptional landmines. Repression via one or two KRAB-ZFPs would be sufficient to block TE expression and transposition. Multiple KRAB-ZFPs binding and deposition of repressive chromatin would impair, for instance, transcription factors recruitment and enhancer effects, which could be later co-opted by the host.

Finally, an important experiment that could still be done to better understand ZFP932/Gm15446 regulation of genes via TEs would be to demonstrate a direct TE-gene interaction. While our *Bglap3* example demonstrates well that KRAB-ZFPs can impact directly on genes via promoter effects – and we accumulated relevant data to show that the KRAB/KAP1 complex could also repress enhancer effects – we have not provided direct evidence of KRAB-ZFP-mediated enhancer

activity. Two interesting possibilities would be to (1) perform chromosome capture techniques in KO versus control cells, and (2) remove KRAB/KAP1-bound TEs via CRISPR/Cas9 and monitor epigenetic and transcriptional consequences on neighboring genes.

4. Arms race versus TEs domestication

Collectively, the results generated from this thesis support both the existence of an evolutionary arms race between TEs and KRAB-ZFPs and an ongoing process of KRAB-ZFP/TE domestication (**Figure 8**). Upon entry of a new TE in the genome, the transposon would be expressed and initially controlled via small RNA mechanisms, such as piRNAs. Over time, the fast-evolving KRAB-ZFP family could generate new paralogs, one of which would be capable of binding the TE family and repressing it. A co-evolution would follow, in which both KRAB-ZFPs and TEs would accumulate mutations and would be selected for repression or escaping, respectively. At any stage of this process, a newly inserted TE could land in the vicinity of a gene. In the event that this gene could be important for physiologically or phenotypically relevant function, the KRAB-ZFP, TE, and gene would be susceptible to high selective pressure. In the hypothetical example presented here, the TE would be inserted next to a gene that is relevant for fur color. Upon insertion of the TE and presence of the repressor, the mouse would have a phenotype that confers a selective advantage regarding predation (grey fur color on a grey environment). In that case, the presence of this TE/KRAB-ZFPs pair could be then selected and fixed in the population. We propose that both arms race and domestication coexist, fueling the genetic diversity of the host, its KRAB-ZFPs, and of TE sequences.
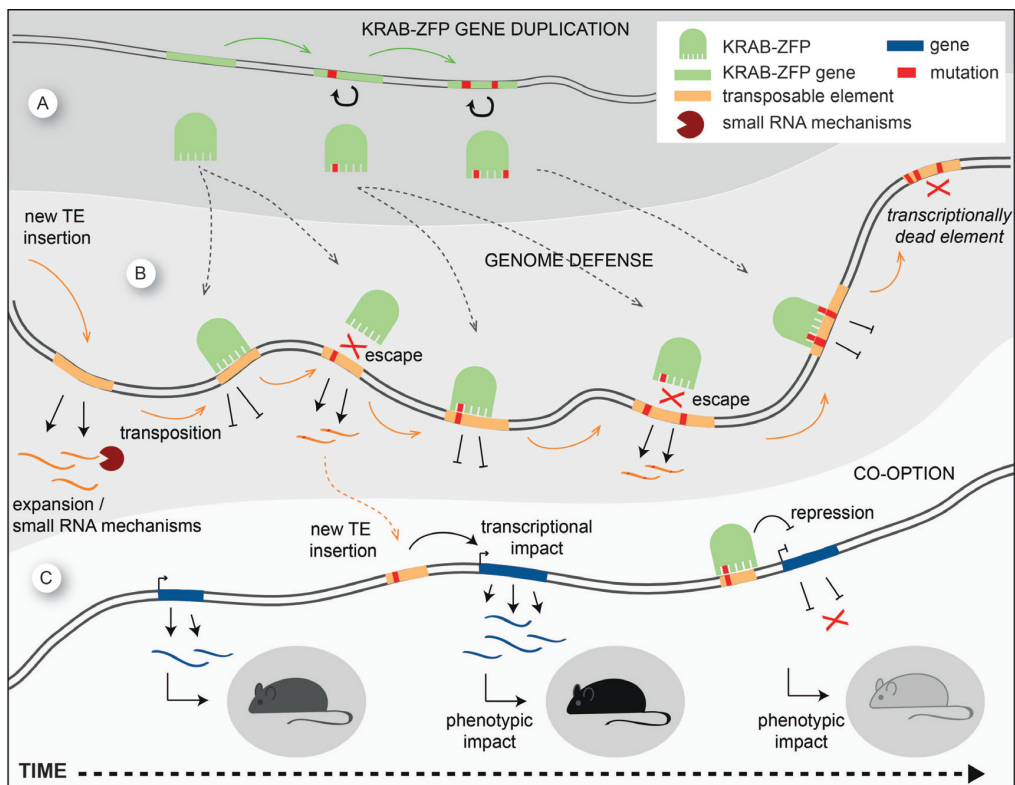
Figure 8. Co-existence of evolutionary arms race between KRAB-ZFPs and TEs and TE domestication by the host via KRAB-ZFPs. In this model, when a novel TE enters the host, it starts to be expressed and to transpose, and the first line of defense normally comprises small RNA-mediated mechanisms (**B**). Over time, KRAB-ZFPs genes duplicate and are selected to bind novel replicating TEs (**A**). In parallel, transposons accumulate mutations and escape repression (**B**), which triggers KRAB-ZFPs structural evolution to suppress expression of new escapees (**A**). The accumulation of mutations in transposons eventually leads to deleterious mutations in the TE, which impair transcription. At the same time, new transposon insertions near genes can lead to transcriptional impacts that could have phenotypic effects (**C**). In this scenario KRAB-ZFPs could then repress the effects of the TE and the TE/KRAB-ZFP pair could be co-opted in a physiological process, as illustrated here hypothetically for fur color.

130

VI. References

Amouroux, R., Nashun, B., Shirane, K., Nakagawa, S., Hill, P.W., D'Souza, Z., Nakayama, M., Matsuda, M., Turp, A., Ndjetehe, E.*, et al.* (2016). De novo DNA methylation drives 5hmC accumulation in mouse zygotes. Nature cell biology *18*, 225-233.

Antony, J.M., Deslauriers, A.M., Bhat, R.K., Ellestad, K.K., and Power, C. (2011). Human endogenous retroviruses and multiple sclerosis: innocent bystanders or disease determinants? Biochimica et biophysica acta *1812*, 162-176.

Aravin, A.A., and Bourc'his, D. (2008). Small RNA guides for de novo DNA methylation in mammalian germ cells. Genes & development *22*, 970-975.

Aravin, A.A., Sachidanandam, R., Bourc'his, D., Schaefer, C., Pezic, D., Toth, K.F., Bestor, T., and Hannon, G.J. (2008). A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. Molecular cell *31*, 785-799.

Ayarpadikannan, S., Lee, H.-E., Han, K., and Kim, H.-S. (2015). Transposable element-driven transcript diversification and its relevance to genetic disorders. Gene *558*, 187-194.

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

Barde, I., Verp, S., Offner, S., and Trono, D. (2011). Lentiviral Vector Mediated Transgenesis. Current protocols in mouse biology *1*, 169-184.

Baudino, L., Yoshinobu, K., Morito, N., Santiago-Raber, M.L., and Izui, S. (2010). Role of endogenous retroviruses in murine SLE. Autoimmunity reviews *10*, 27-34.

Baylin, S.B., and Jones, P.A. (2011). A decade of exploring the cancer epigenome - biological and translational implications. Nature reviews. Cancer *11*, 726-734.

Beck, C.R., Garcia-Perez, J.L., Badge, R.M., and Moran, J.V. (2011). LINE-1 elements in structural variation and disease. Annual review of genomics and human genetics *12*, 187-215.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society. Series B (Methodological) *57*, 289-300.

Bierhoff, H., Dammert, M.A., Brocks, D., Dambacher, S., Schotta, G., and Grummt, I. (2014). Quiescence-induced LncRNAs trigger H4K20 trimethylation and transcriptional silencing. Molecular cell *54*, 675-682.

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

Bloom, K., and Joglekar, A. (2010). Towards building a chromosome segregation machine. Nature *463*, 446-456.

Bojkowska, K., Aloisio, F., Cassano, M., Kapopoulou, A., Santoni de Sio, F., Zangger, N., Offner, S., Cartoni, C., Thomas, C., Quenneville, S.*, et al.* (2012). Liver-specific ablation of Kruppel-associated box-associated protein 1 in mice leads to male-predominant hepatosteatosis and development of liver adenoma. Hepatology (Baltimore, Md.) *56*, 1279-1290.

Bourque, G., Leong, B., Vega, V.B., Chen, X., Lee, Y.L., Srinivasan, K.G., Chew, J.L., Ruan, Y., Wei, C.L., Ng, H.H.*, et al.* (2008). Evolution of the mammalian transcription factor binding repertoire via transposable elements. Genome research *18*, 1752-1762.

Branco, M.R., Ficz, G., and Reik, W. (2012). Uncovering the role of 5-hydroxymethylcytosine in the epigenome. Nature reviews. Genetics *13*, 7-13.

Britten, R.J., and Davidson, E.H. (1971). Repetitive and Non-Repetitive DNA Sequences and a Speculation on the Origins of Evolutionary Novelty. The Quarterly Review of Biology *46*, 111-138.

Bulut-Karslioglu, A., De La Rosa-Velazquez, I.A., Ramirez, F., Barenboim, M., Onishi-Seebacher, M., Arand, J., Galan, C., Winter, G.E., Engist, B., Gerle, B.*, et al.* (2014). Suv39h-dependent H3K9me3 marks intact retrotransposons and silences LINE elements in mouse embryonic stem cells. Molecular cell *55*, 277-290.

Byrum, S.D., Raman, A., Taverna, S.D., and Tackett, A.J. (2012). ChAP-MS: a method for identification of proteins and histone posttranslational modifications at a single genomic locus. Cell reports *2*, 198-205.

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

Carmell, M.A., Girard, A., van de Kant, H.J., Bourc'his, D., Bestor, T.H., de Rooij, D.G., and Hannon, G.J. (2007). MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. Developmental cell *12*, 503-514.

Carmi, S., Church, G.M., and Levanon, E.Y. (2011). Large-scale DNA editing of retrotransposons accelerates mammalian genome evolution. Nature communications *2*, 519.

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

Chalopin, D., Naville, M., Plard, F., Galiana, D., and Volff, J.N. (2015). Comparative analysis of transposable elements highlights mobilome diversity and evolution in vertebrates. Genome biology and evolution *7*, 567-580.

Christy, R.J., Brown, A.R., Gourlie, B.B., and Huang, R.C. (1985). Nucleotide sequences of murine intracisternal A-particle gene LTRs have extensive variability within the R region. Nucleic acids research *13*, 289-302.

Chuong, E.B. (2013). Retroviruses facilitate the rapid evolution of the mammalian placenta. BioEssays : news and reviews in molecular, cellular and developmental biology *35*, 853-861.

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

Chuong, E.B., Rumi, M.A., Soares, M.J., and Baker, J.C. (2013). Endogenous retroviruses function as species-specific enhancer elements in the placenta. Nature genetics *45*, 325-329.

Ciabrelli, F., and Cavalli, G. (2015). Chromatin-driven behavior of topologically associating domains. Journal of molecular biology *427*, 608-625.

Cohen, C.J., Lock, W.M., and Mager, D.L. (2009). Endogenous retroviral LTRs as promoters for human genes: a critical assessment. Gene *448*, 105-114.

Collins, 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. Proceedings of the National Academy of Sciences of the United States of America *112*, 8367-8372.

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

Dejardin, J., and Kingston, R.E. (2009). Purification of proteins associated with specific genomic Loci. Cell *136*, 175-186.

Deplancke, B. (2009). Experimental advances in the characterization of metazoan gene regulatory networks. Briefings in functional genomics & proteomics *8*, 12-27.

Deplancke, B., Dupuy, D., Vidal, M., and Walhout, A.J. (2004). A gateway-compatible yeast one-hybrid system. Genome research *14*, 2093-2101.

Desbois, C., Hogue, D.A., and Karsenty, G. (1994). The mouse osteocalcin gene cluster contains three genes with two separate spatial and temporal patterns of expression. The Journal of biological chemistry *269*, 1183-1190.

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

Dey, B., Thukral, S., Krishnan, S., Chakrobarty, M., Gupta, S., Manghani, C., and Rani, V. (2012). DNA-protein interactions: methods for detection and analysis. Molecular and cellular biochemistry *365*, 279-299.

Dodge, J.E., Kang, Y.K., Beppu, H., Lei, H., and Li, E. (2004). Histone H3-K9 methyltransferase ESET is essential for early development. Molecular and cellular biology *24*, 2478-2486.

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

Doolittle, W.F., and Sapienza, C. (1980). Selfish genes, the phenotype paradigm and genome evolution. Nature *284*, 601-603.

Ducy, P., Desbois, C., Boyce, B., Pinero, G., Story, B., Dunstan, C., Smith, E., Bonadio, J., Goldstein, S., Gundberg, C.*, et al.* (1996). Increased bone formation in osteocalcin-deficient mice. Nature *382*, 448-452.

Dupressoir, A., Lavialle, C., and Heidmann, T. (2012). From ancestral infectious retroviruses to bona fide cellular genes: role of the captured syncytins in placentation. Placenta *33*, 663-671.

Dupressoir, A., Vernochet, C., Bawa, O., Harper, F., Pierron, G., Opolon, P., and Heidmann, T. (2009). Syncytin-A knockout mice demonstrate the critical role in placentation of a

fusogenic, endogenous retrovirus-derived, envelope gene. Proceedings of the National Academy of Sciences of the United States of America *106*, 12127-12132.

Dupressoir, A., Vernochet, C., Harper, F., Guegan, J., Dessen, P., Pierron, G., and Heidmann, T. (2011). A pair of co-opted retroviral envelope syncytin genes is required for formation of the two-layered murine placental syncytiotrophoblast. Proceedings of the National Academy of Sciences of the United States of America *108*, E1164-1173.

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

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

Ferron, M., Wei, J., Yoshizawa, T., Del Fattore, A., DePinho, R.A., Teti, A., Ducy, P., and Karsenty, G. (2010). Insulin signaling in osteoblasts integrates bone remodeling and energy metabolism. Cell *142*, 296-308.

Feschotte, C., and Pritham, E.J. (2007). DNA transposons and the evolution of eukaryotic genomes. Annual review of genetics *41*, 331-368.

Fort, A., Hashimoto, K., Yamada, D., Salimullah, M., Keya, C.A., Saxena, A., Bonetti, A., Voineagu, I., Bertin, N., Kratz, A*., et al.* (2014). Deep transcriptome profiling of mammalian stem cells supports a regulatory role for retrotransposons in pluripotency maintenance. Nature genetics *46*, 558-566.

Franchini, D.M., Schmitz, K.M., and Petersen-Mahrt, S.K. (2012). 5-Methylcytosine DNA demethylation: more than losing a methyl group. Annual review of genetics *46*, 419-441.

Friedli, M., and Trono, D. (2015a). The developmental control of transposable elements and the evolution of higher species. Annual Review of Cellular and Developmental Biology *31*, 13.11-13.23.

Friedli, M., and Trono, D. (2015b). The developmental control of transposable elements and the evolution of higher species. Annual review of cell and developmental biology *31*, 429-451.

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

Fujita, T., and Fujii, H. (2013). Efficient isolation of specific genomic regions and identification of associated proteins by engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP) using CRISPR. Biochemical and biophysical research communications *439*, 132-136.

Gautsch, J.W., and Wilson, M.C. (1983). Delayed de novo methylation in teratocarcinoma suggests additional tissue-specific mechanisms for controlling gene expression. Nature *301*, 32-37.

Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J*., et al.* (2004). Bioconductor: open software development for computational biology and bioinformatics. Genome biology *5*, R80.

Ghildiyal, M., and Zamore, P.D. (2009). Small silencing RNAs: an expanding universe. Nature reviews. Genetics *10*, 94-108.

Goll, M.G., and Bestor, T.H. (2005). Eukaryotic cytosine methyltransferases. Annual review of biochemistry *74*, 481-514.

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

Gubelmann, C., Gattiker, A., Massouras, A., Hens, K., David, F., Decouttere, F., Rougemont, J., and Deplancke, B. (2011). GETPrime: a gene- or transcript-specific primer database for quantitative real-time PCR. Database : the journal of biological databases and curation *2011*, bar040.

Gubelmann, C., Waszak, S.M., Isakova, A., Holcombe, W., Hens, K., Iagovitina, A., Feuz, J.D., Raghav, S.K., Simicevic, J., and Deplancke, B. (2013). A yeast one-hybrid and microfluidics-based pipeline to map mammalian gene regulatory networks. Molecular systems biology *9*, 682.

Guillen-Ahlers, H., Shortreed, M.R., Smith, L.M., and Olivier, M. (2014). Advanced methods for the analysis of chromatin-associated proteins. Physiological genomics *46*, 441-447.

Hancks, D.C., and Kazazian, H.H., Jr. (2010). SVA retrotransposons: Evolution and genetic instability. Seminars in cancer biology *20*, 234-245.

Hancks, D.C., and Kazazian, H.H., Jr. (2012). Active human retrotransposons: variation and disease. Current opinion in genetics & development *22*, 191-203.

He, Z., Cai, J., Lim, J.W., Kroll, K., and Ma, L. (2011). A novel KRAB domain-containing zinc finger transcription factor ZNF431 directly represses Patched1 transcription. The Journal of biological chemistry *286*, 7279-7289.

Hens, K., Feuz, J.D., Isakova, A., Iagovitina, A., Massouras, A., Bryois, J., Callaerts, P., Celniker, S.E., and Deplancke, B. (2011). Automated protein-DNA interaction screening of Drosophila regulatory elements. Nature methods *8*, 1065-1070.

Heras, S.R., Macias, S., Plass, M., Fernandez, N., Cano, D., Eyras, E., Garcia-Perez, J.L., and Caceres, J.F. (2013). The Microprocessor controls the activity of mammalian retrotransposons. Nature structural & molecular biology *20*, 1173-1181.

Hu, G., Kim, J., Xu, Q., Leng, Y., Orkin, S.H., and Elledge, S.J. (2009). A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal. Genes & development *23*, 837-848.

Huang, G.J., He, Z., and Ma, L. (2012). ZFP932 suppresses cellular Hedgehog response and Patched1 transcription. Vitamins and hormones *88*, 309-332.

Hutnick, L.K., Huang, X., Loo, T.C., Ma, Z., and Fan, G. (2010). Repression of retrotransposal elements in mouse embryonic stem cells is primarily mediated by a DNA methylation-independent mechanism. The Journal of biological chemistry *285*, 21082-21091.

Ide, S., and Dejardin, J. (2015). End-targeting proteomics of isolated chromatin segments of a mammalian ribosomal RNA gene promoter. Nature communications *6*, 6674.

Ito, S., D'Alessio, A.C., Taranova, O.V., Hong, K., Sowers, L.C., and Zhang, Y. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature *466*, 1129-1133.

Iuchi, S. (2001). Three classes of C2H2 zinc finger proteins. Cellular and molecular life sciences : CMLS *58*, 625-635.

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

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

Jackson-Grusby, L., Beard, C., Possemato, R., Tudor, M., Fambrough, D., Csankovszki, G., Dausman, J., Lee, P., Wilson, C., Lander, E*., et al.* (2001). Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. Nature genetics *27*, 31-39.

Jacobs, F.M., Greenberg, D., Nguyen, N., Haeussler, M., Ewing, A.D., Katzman, S., Paten, B., Salama, S.R., and Haussler, D. (2014). An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. Nature *516*, 242-245.

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

Jahner, D., Stuhlmann, H., Stewart, C.L., Harbers, K., Lohler, J., Simon, I., and Jaenisch, R. (1982). De novo methylation and expression of retroviral genomes during mouse embryogenesis. Nature *298*, 623-628.

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

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

Kanellopoulou, C., Muljo, S.A., Kung, A.L., Ganesan, S., Drapkin, R., Jenuwein, T., Livingston, D.M., and Rajewsky, K. (2005). Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. Genes & development *19*, 489-501.

Kapusta, A., Kronenberg, Z., Lynch, V.J., Zhuo, X., Ramsay, L., Bourque, G., Yandell, M., and Feschotte, C. (2013). Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. PLoS genetics *9*, e1003470.

Karimi, M.M., Goyal, P., Maksakova, I.A., Bilenky, M., Leung, D., Tang, J.X., Shinkai, Y., Mager, D.L., Jones, S., Hirst, M*., et al.* (2011). DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mESCs. Cell stem cell *8*, 676-687.

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

Kazazian, H.H., Jr., Wong, C., Youssoufian, H., Scott, A.F., Phillips, D.G., and Antonarakis, S.E. (1988). Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. Nature *332*, 164-166.

Kempler, G., Freitag, B., Berwin, B., Nanassy, O., and Barklis, E. (1993). Characterization of the Moloney murine leukemia virus stem cell-specific repressor binding site. Virology *193*, 690-699.

Khan, H., Smit, A., and Boissinot, S. (2006). Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. Genome research *16*, 78-87.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome biology *14*, R36.

Kohli, R.M., and Zhang, Y. (2013). TET enzymes, TDG and the dynamics of DNA demethylation. Nature *502*, 472-479.

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

Kuramochi-Miyagawa, S., Watanabe, T., Gotoh, K., Totoki, Y., Toyoda, A., Ikawa, M., Asada, N., Kojima, K., Yamaguchi, Y., Ijiri, T.W*., et al.* (2008). DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. Genes & development *22*, 908-917.

Lamprecht, B., Walter, K., Kreher, S., Kumar, R., Hummel, M., Lenze, D., Kochert, K., Bouhlel, M.A., Richter, J., Soler, E*., et al.* (2010). Derepression of an endogenous long terminal repeat activates the CSF1R proto-oncogene in human lymphoma. Nature medicine *16*, 571-579, 571p following 579.

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

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

Lavialle, C., Cornelis, G., Dupressoir, A., Esnault, C., Heidmann, O., Vernochet, C., and Heidmann, T. (2013). Paleovirology of 'syncytins', retroviral env genes exapted for a role in placentation. Philosophical transactions of the Royal Society of London. Series B, Biological sciences *368*, 20120507.

Law, C.W., Chen, Y., Shi, W., and Smyth, G.K. (2014). voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome biology *15*, R29.

Lawrence, M., Daujat, S., and Schneider, R. (2016). Lateral Thinking: How Histone Modifications Regulate Gene Expression. Trends in genetics : TIG *32*, 42-56.

Lee, N.K., Sowa, H., Hinoi, E., Ferron, M., Ahn, J.D., Confavreux, C., Dacquin, R., Mee, P.J., McKee, M.D., Jung, D.Y*., et al.* (2007). Endocrine regulation of energy metabolism by the skeleton. Cell *130*, 456-469.

Leung, D.C., Dong, K.B., Maksakova, I.A., Goyal, P., Appanah, R., Lee, S., Tachibana, M., Shinkai, Y., Lehnertz, B., Mager, D.L*., et al.* (2011). Lysine methyltransferase G9a is required for de novo DNA methylation and the establishment, but not the maintenance, of proviral silencing. Proceedings of the National Academy of Sciences of the United States of America *108*, 5718-5723.

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

Li, E., Bestor, T.H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell *69*, 915-926.

Li, X., Scaringe, W.A., Hill, K.A., Roberts, S., Mengos, A., Careri, D., Pinto, M.T., Kasper, C.K., and Sommer, S.S. (2001). Frequency of recent retrotransposition events in the human factor IX gene. Human mutation *17*, 511-519.

Liu, H., Chang, L.H., Sun, Y., Lu, X., and Stubbs, L. (2014). Deep vertebrate roots for mammalian zinc finger transcription factor subfamilies. Genome biology and evolution *6*, 510-525.

Lizio, M., Harshbarger, J., Shimoji, H., Severin, J., Kasukawa, T., Sahin, S., Abugessaisa, I., Fukuda, S., Hori, F., Ishikawa-Kato, S*., et al.* (2015). Gateways to the FANTOM5 promoter level mammalian expression atlas. Genome biology *16*, 22.

Macfarlan, T.S., Gifford, W.D., Driscoll, S., Lettieri, K., Rowe, H.M., Bonanomi, D., Firth, A., Singer, O., Trono, D., and Pfaff, S.L. (2012). Embryonic stem cell potency fluctuates with endogenous retrovirus activity. Nature *487*, 57-63.

Mager, D.L., and Stoye, J.P. (2015). Mammalian Endogenous Retroviruses. Microbiology spectrum *3*, MDNA3-0009-2014.

Maksakova, I.A., Mager, D.L., and Reiss, D. (2008). Keeping active endogenous retroviral-like elements in check: the epigenetic perspective. Cellular and molecular life sciences : CMLS *65*, 3329-3347.

Maksakova, I.A., Romanish, M.T., Gagnier, L., Dunn, C.A., van de Lagemaat, L.N., and Mager, D.L. (2006). Retroviral Elements and Their Hosts: Insertional Mutagenesis in the Mouse Germ Line. PLoS genetics *2*, e2.

Maksakova, I.A., Thompson, P.J., Goyal, P., Jones, S.J., Singh, P.B., Karimi, M.M., and Lorincz, M.C. (2013). Distinct roles of KAP1, HP1 and G9a/GLP in silencing of the two-cell-specific retrotransposon MERVL in mouse ES cells. Epigenetics & chromatin *6*, 15.

Malone, C.D., and Hannon, G.J. (2009). Small RNAs as guardians of the genome. Cell *136*, 656-668.

Marchetto, M.C., Narvaiza, I., Denli, A.M., Benner, C., Lazzarini, T.A., Nathanson, J.L., Paquola, A.C., Desai, K.N., Herai, R.H., Weitzman, M.D.*, et al.* (2013). Differential L1 regulation in pluripotent stem cells of humans and apes. Nature *503*, 525-529.

Marchi, E., Kanapin, A., Magiorkinis, G., and Belshaw, R. (2014). Unfixed endogenous retroviral insertions in the human population. Journal of virology *88*, 9529-9537.

Margueron, R., and Reinberg, D. (2010). Chromatin structure and the inheritance of epigenetic information. Nature reviews. Genetics *11*, 285-296.

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

McClintock, B. (1950). The Origin and Behavior of Mutable Loci in Maize. Proceedings of the National Academy of Sciences of the United States of America *36*, 344-355.

McNamara, R.P., Reeder, J.E., McMillan, E.A., Bacon, C.W., McCann, J.L., and D'Orso, I. (2016). KAP1 Recruitment of the 7SK snRNP Complex to Promoters Enables Transcription Elongation by RNA Polymerase II. Molecular cell *61*, 39-53.

Medina-Rivera, A., Defrance, M., Sand, O., Herrmann, C., Castro-Mondragon, J.A., Delerce, J., Jaeger, S., Blanchet, C., Vincens, P., Caron, C.*, et al.* (2015). RSAT 2015: Regulatory Sequence Analysis Tools. Nucleic acids research *43*, W50-56.

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

Mi, S., Lee, X., Li, X., Veldman, G.M., Finnerty, H., Racie, L., LaVallie, E., Tang, X.Y., Edouard, P., Howes, S.*, et al.* (2000). Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. Nature *403*, 785-789.

Miki, Y., Katagiri, T., Kasumi, F., Yoshimoto, T., and Nakamura, Y. (1996). Mutation analysis in the BRCA2 gene in primary breast cancers. Nature genetics *13*, 245-247.

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

Morgan, H.D., Sutherland, H.G., Martin, D.I., and Whitelaw, E. (1999). Epigenetic inheritance at the agouti locus in the mouse. Nature genetics *23*, 314-318.

Najafabadi, H.S., Mnaimneh, S., Schmitges, F.W., Garton, M., Lam, K.N., Yang, A., Albu, M., Weirauch, M.T., Radovani, E., Kim, P.M., *et al.* (2015). C2H2 zinc finger proteins greatly expand the human regulatory lexicon. Nature biotechnology *33*, 555-562.

Nakanishi, A., Kobayashi, N., Suzuki-Hirano, A., Nishihara, H., Sasaki, T., Hirakawa, M., Sumiyama, K., Shimogori, T., and Okada, N. (2012). A SINE-derived element constitutes a unique modular enhancer for mammalian diencephalic Fgf8. PloS one *7*, e43785.

Ng, S.Y., Johnson, R., and Stanton, L.W. (2012). Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors. The EMBO journal *31*, 522-533.

Nielsen, A.L., Ortiz, J.A., You, J., Oulad-Abdelghani, M., Khechumian, R., Gansmuller, A., Chambon, P., and Losson, R. (1999). Interaction with members of the heterochromatin protein 1 (HP1) family and histone deacetylation are differentially involved in transcriptional silencing by members of the TIF1 family. The EMBO journal *18*, 6385-6395.

Niwa, O., Yokota, Y., Ishida, H., and Sugahara, T. (1983). Independent mechanisms involved in suppression of the Moloney leukemia virus genome during differentiation of murine teratocarcinoma cells. Cell *32*, 1105-1113.

Notwell, J.H., Chung, T., Heavner, W., and Bejerano, G. (2015). A family of transposable elements co-opted into developmental enhancers in the mouse neocortex. Nature communications *6*, 6644.

Ohnuki, M., Tanabe, K., Sutou, K., Teramoto, I., Sawamura, Y., Narita, M., Nakamura, M., Tokunaga, Y., Nakamura, M., Watanabe, A., *et al.* (2014). Dynamic regulation of human endogenous retroviruses mediates factor-induced reprogramming and differentiation potential. Proceedings of the National Academy of Sciences of the United States of America *111*, 12426-12431.

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

Orgel, L.E., and Crick, F.H. (1980). Selfish DNA: the ultimate parasite. Nature *284*, 604-607.

Oury, F., Sumara, G., Sumara, O., Ferron, M., Chang, H., Smith, C.E., Hermo, L., Suarez, S., Roth, B.L., Ducy, P., *et al.* (2011). Endocrine regulation of male fertility by the skeleton. Cell *144*, 796-809.

Pace, J.K., 2nd, and Feschotte, C. (2007). The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage. Genome research *17*, 422-432.

Padeken, J., Zeller, P., and Gasser, S.M. (2015). Repeat DNA in genome organization and stability. Current opinion in genetics & development *31*, 12-19.

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

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

Petrucci, M., Paquette, Y., Leblond, F.A., Pichette, V., and Bonnardeaux, A. (2006). Evidence that the mouse osteocalcin-related gene does not encode nephrocalcin. Nephron. Experimental nephrology *104*, e140-146.

Pezic, D., Manakov, S.A., Sachidanandam, R., and Aravin, A.A. (2014). piRNA pathway targets active LINE1 elements to establish the repressive H3K9me3 mark in germ cells. Genes & development *28*, 1410-1428.

Pi, W., Zhu, X., Wu, M., Wang, Y., Fulzele, S., Eroglu, A., Ling, J., and Tuan, D. (2010). Long-range function of an intergenic retrotransposon. Proceedings of the National Academy of Sciences of the United States of America *107*, 12992-12997.

Puget, N., Sinilnikova, O.M., Stoppa-Lyonnet, D., Audoynaud, C., Pages, S., Lynch, H.T., Goldgar, D., Lenoir, G.M., and Mazoyer, S. (1999). An Alu-mediated 6-kb duplication in the BRCA1 gene: a new founder mutation? American journal of human genetics *64*, 300-302.

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

Quenneville, S., Verde, G., Corsinotti, A., Kapopoulou, A., Jakobsson, J., Offner, S., Baglivo, I., Pedone, P.V., Grimaldi, G., Riccio, A.*, et al.* (2011). In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. Molecular cell *44*, 361-372.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics (Oxford, England) *26*, 841-842.

Raiz, J., Damert, A., Chira, S., Held, U., Klawitter, S., Hamdorf, M., Lower, J., Stratling, W.H., Lower, R., and Schumann, G.G. (2012). The non-autonomous retrotransposon SVA is trans-mobilized by the human LINE-1 protein machinery. Nucleic acids research *40*, 1666-1683.

Ramirez, M.A., Pericuesta, E., Fernandez-Gonzalez, R., Moreira, P., Pintado, B., and Gutierrez-Adan, A. (2006). Transcriptional and post-transcriptional regulation of retrotransposons IAP and MuERV-L affect pluripotency of mice ES cells. Reproductive biology and endocrinology : RB&E *4*, 55.

Rebollo, R., Farivar, S., and Mager, D.L. (2012a). C-GATE - catalogue of genes affected by transposable elements. Mobile DNA *3*, 9.

Rebollo, R., Miceli-Royer, K., Zhang, Y., Farivar, S., Gagnier, L., and Mager, D.L. (2012b). Epigenetic interplay between mouse endogenous retroviruses and host genes. Genome biology *13*, R89.

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

Reece-Hoyes, J.S., Barutcu, A.R., McCord, R.P., Jeong, J.S., Jiang, L., MacWilliams, A., Yang, X., Salehi-Ashtiani, K., Hill, D.E., Blackshaw, S.*, et al.* (2011). Yeast one-hybrid assays for gene-centered human gene regulatory network mapping. Nature methods *8*, 1050-1052.

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

Romanish, M.T., Cohen, C.J., and Mager, D.L. (2010). Potential mechanisms of endogenous retroviral-mediated genomic instability in human cancer. Seminars in cancer biology *20*, 246-253.

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

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

Rowe, H.M., Kapopoulou, A., Corsinotti, A., Fasching, L., Macfarlan, T.S., Tarabay, Y., Viville, S., Jakobsson, J., Pfaff, S.L., and Trono, D. (2013b). TRIM28 repression of retrotransposon-based enhancers is necessary to preserve transcriptional dynamics in embryonic stem cells. Genome research *23*, 452-461.

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

Salmon-Divon, M., Dvinge, H., Tammoja, K., and Bertone, P. (2010). PeakAnalyzer: genome-wide annotation of chromatin binding and modification loci. BMC bioinformatics *11*, 415.

Samuelson, L.C., Phillips, R.S., and Swanberg, L.J. (1996). Amylase gene structures in primates: retroposon insertions and promoter evolution. Molecular biology and evolution *13*, 767-779.

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

Sato, M., and Tada, N. (1995). Preferential expression of osteocalcin-related protein mRNA in gonadal tissues of male mice. Biochemical and biophysical research communications *215*, 412-421.

Schones, D.E., and Zhao, K. (2008). Genome-wide approaches to studying chromatin modifications. Nature reviews. Genetics *9*, 179-191.

Schubeler, D. (2012). Molecular biology. Epigenetic islands in a genetic ocean. Science (New York, N.Y.) *338*, 756-757.

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

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

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

Simicevic, J., and Deplancke, B. (2010). DNA-centered approaches to characterize regulatory protein-DNA interaction complexes. Molecular bioSystems *6*, 462-468.

Singh, K., Cassano, M., Planet, E., Sebastian, S., Jang, S.M., Sohi, G., Faralli, H., Choi, J., Youn, H.D., Dilworth, F.J.*, et al.* (2015). A KAP1 phosphorylation switch controls MyoD function during skeletal muscle differentiation. Genes & development *29*, 513-525.

Sookdeo, A., Hepp, C.M., McClure, M.A., and Boissinot, S. (2013). Revisiting the evolution of mouse LINE-1 in the genomic era. Mobile DNA *4*, 3.

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

Stavenhagen, J.B., and Robins, D.M. (1988). An ancient provirus has imposed androgen regulation on the adjacent mouse sex-limited protein gene. Cell *55*, 247-254.

Stubbs, L., Sun, Y., and Caetano-Anolles, D. (2011). Function and Evolution of C2H2 Zinc Finger Arrays. Sub-cellular biochemistry *52*, 75-94.

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

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

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

Tan, X., Xu, X., Elkenani, M., Smorag, L., Zechner, U., Nolte, J., Engel, W., and Pantakani, D.V. (2013). Zfp819, a novel KRAB-zinc finger protein, interacts with KAP1 and functions in genomic integrity maintenance of mouse embryonic stem cells. Stem cell research *11*, 1045-1059.

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

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

Tsumura, A., Hayakawa, T., Kumaki, Y., Takebayashi, S., Sakaue, M., Matsuoka, C., Shimotohno, K., Ishikawa, F., Li, E., Ueda, H.R.*, et al.* (2006). Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. Genes to cells : devoted to molecular & cellular mechanisms *11*, 805-814.

Turelli, P., Castro-Diaz, N., Marzetta, F., Kapopoulou, A., Raclot, C., Duc, J., Tieng, V., Quenneville, S., and Trono, D. (2014). Interplay of TRIM28 and DNA methylation in controlling human endogenous retroelements. Genome research *24*, 1260-1270.

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

Urrutia, R. (2003). KRAB-containing zinc-finger repressor proteins. Genome biology *4*, 231.

Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A., and Luscombe, N.M. (2009). A census of human transcription factors: function, expression and evolution. Nature reviews. Genetics *10*, 252-263.

Vasicek, T.J., Zeng, L., Guan, X.J., Zhang, T., Costantini, F., and Tilghman, S.M. (1997). Two dominant mutations in the mouse fused gene are the result of transposon insertions. Genetics *147*, 777-786.

Venkatesh, S., and Workman, J.L. (2015). Histone exchange, chromatin structure and the regulation of transcription. Nature reviews. Molecular cell biology *16*, 178-189.

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

Wang, T., Zeng, J., Lowe, C.B., Sellers, R.G., Salama, S.R., Yang, M., Burgess, S.M., Brachmann, R.K., and Haussler, D. (2007). Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53. Proceedings of the National Academy of Sciences of the United States of America *104*, 18613-18618.

Warren, I.A., Naville, M., Chalopin, D., Levin, P., Berger, C.S., Galiana, D., and Volff, J.N. (2015). Evolutionary impact of transposable elements on genomic diversity and lineage-specific innovation in vertebrates. Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology *23*, 505-531.

Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., *et al.* (2002). Initial sequencing and comparative analysis of the mouse genome. Nature *420*, 520-562.

Weber, A.R., Krawczyk, C., Robertson, A.B., Kusnierczyk, A., Vagbo, C.B., Schuermann, D., Klungland, A., and Schar, P. (2016). Biochemical reconstitution of TET1-TDG-BER-dependent active DNA demethylation reveals a highly coordinated mechanism. Nature communications *7*, 10806.

Wissing, S., Munoz-Lopez, M., Macia, A., Yang, Z., Montano, M., Collins, W., Garcia-Perez, J.L., Moran, J.V., and Greene, W.C. (2012). Reprogramming somatic cells into iPS cells activates LINE-1 retroelement mobility. Human molecular genetics *21*, 208-218.

Wiznerowicz, M., Jakobsson, J., Szulc, J., Liao, S., Quazzola, A., Beermann, F., Aebischer, P., and Trono, D. (2007). The Kruppel-associated box repressor domain can trigger de novo promoter methylation during mouse early embryogenesis. The Journal of biological chemistry *282*, 34535-34541.

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

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

Wolf, D., Hug, K., and Goff, S.P. (2008). TRIM28 mediates primer binding site-targeted silencing of Lys1,2 tRNA-utilizing retroviruses in embryonic cells. Proceedings of the National Academy of Sciences of the United States of America *105*, 12521-12526.

Wolf, G., Yang, P., Fuchtbauer, A.C., Fuchtbauer, E.M., Silva, A.M., Park, C., Wu, W., Nielsen, A.L., Pedersen, F.S., and Macfarlan, T.S. (2015). The KRAB zinc finger protein ZFP809 is required to initiate epigenetic silencing of endogenous retroviruses. Genes & development *29*, 538-554.

Yang, N., and Kazazian, H.H., Jr. (2006). L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. Nature structural & molecular biology *13*, 763-771.

Zang, C., Schones, D.E., Zeng, C., Cui, K., Zhao, K., and Peng, W. (2009). A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. Bioinformatics (Oxford, England) *25*, 1952-1958.

Zeng, L., Fagotto, F., Zhang, T., Hsu, W., Vasicek, T.J., Perry, W.L., 3rd, Lee, J.J., Tilghman, S.M., Gumbiner, B.M., and Costantini, F. (1997). The mouse Fused locus encodes Axin, an inhibitor of the Wnt signaling pathway that regulates embryonic axis formation. Cell *90*, 181-192.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W.*, et al.* (2008). Model-based analysis of ChIP-Seq (MACS). Genome biology *9*, R137.

Zhou, V.W., Goren, A., and Bernstein, B.E. (2011). Charting histone modifications and the functional organization of mammalian genomes. Nature reviews. Genetics *12*, 7-18.

Zuo, X., Sheng, J., Lau, H.T., McDonald, C.M., Andrade, M., Cullen, D.E., Bell, F.T., Iacovino, M., Kyba, M., Xu, G.*, et al.* (2012). Zinc finger protein ZFP57 requires its co-factor to recruit DNA methyltransferases and maintains DNA methylation imprint in embryonic stem cells via its transcriptional repression domain. The Journal of biological chemistry *287*, 2107-2118.

VII. Appendices

M. Friedli, P. Turelli, A. Kapopoulou, B. Rauwel, N. Castro-Diaz, H. M. Rowe, G. Ecco, C. Unzu, E. Planet, A. Lombardo, B. Mangeat, B. E. Wildhaber, L. Naldini, and D. Trono, '**Loss of Transcriptional Control over Endogenous Retroelements During Reprogramming to Pluripotency**', *Genome Res,* 24 (2014), 1251-9.

Research

Loss of transcriptional control over endogenous retroelements during reprogramming to pluripotency

Marc Friedli,[1] Priscilla Turelli,[1] Adamandia Kapopoulou,[1] Benjamin Rauwel,[1] Nathaly Castro-Díaz,[1] Helen M. Rowe,[1,5] Gabriela Ecco,[1] Carmen Unzu,[2] Evarist Planet,[1] Angelo Lombardo,[3] Bastien Mangeat,[4] Barbara E. Wildhaber,[2] Luigi Naldini,[3] and Didier Trono[1]

[1]School of Life Sciences and "Frontiers in Genetics" National Program, Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland; [2]Pediatric Surgery Laboratory, Department of Pathology and Immunology, Faculty of Medicine, University of Geneva, 1211 Geneva, Switzerland; [3]San Raffaele Telethon Institute for Gene Therapy and Vita Salute San Raffaele University, 20132 Milan, Italy; [4]Department of Pathology and Immunology, Faculty of Medicine, University of Geneva, 1211 Geneva, Switzerland

Endogenous retroelements (EREs) account for about half of the mouse or human genome, and their potential as insertional mutagens and transcriptional perturbators is suppressed by early embryonic epigenetic silencing. Here, we asked how ERE control is maintained during the generation of induced pluripotent stem cells (iPSCs), as this procedure involves profound epigenetic remodeling. We found that all EREs tested were markedly up-regulated during the reprogramming of either mouse embryonic fibroblasts, human CD34+ cells, or human primary hepatocytes. At the iPSC stage, EREs of some classes were repressed, whereas others remained highly expressed, yielding a pattern somewhat reminiscent of that recorded in embryonic stem cells. However, variability persisted between individual iPSC clones in the control of specific ERE integrants. Both during reprogramming and in iPS cells, the up-regulation of specific EREs significantly impacted on the transcription of nearby cellular genes. While transcription triggered by specific ERE integrants at highly precise developmental stages may be an essential step toward obtaining pluripotent cells, the broad and unspecific unleashing of the repetitive genome observed here may contribute to the inefficiency of the reprogramming process and to the phenotypic heterogeneity of iPSCs.

[Supplemental material is available for this article.]

The forced expression of a combination of transcription factors such as POU5F1 (also known as OCT4), KLF4, and SOX2 can result in the reprogramming of somatic cells into induced pluripotent stem cells (iPSCs) (Takahashi and Yamanaka 2006; Takahashi et al. 2007). The efficiency of this process varies according to the cells chosen as starting material and the protocols used for their modification, but without further manipulation it does not exceed a few percent and it implies a latency of 1 wk to several weeks, suggesting that a cascade of events, some of which are probably stochastic, is required for full reprogramming (Jaenisch and Young 2008; Hanna et al. 2009; Yamanaka 2009). During this period, a complex sequence of still incompletely characterized epigenetic changes takes place, whereby the expression of pluripotency genes is ultimately induced whereas that of differentiation genes is repressed (Koche et al. 2011; Polo et al. 2012).

The development of a totipotent fertilized egg into a nascent embryo is a paradigm for the reverse process. It too stems from epigenetic mechanisms, which are essential not only for the establishment of specialized lineages but also for events such as imprinting and, most importantly, for the silencing of endogenous retroelements (EREs). EREs collectively account for more than half of the genome of either humans or mice, with thousands to millions of copies of endogenous retroviruses (ERVs), long interspersed elements (LINEs), short interspersed elements (SINEs; which in humans include *Alu* repeats), or SVAs (SINE-VNTR-*Alu*, a hominoid-specific ERE family) (de Koning et al. 2011). These genetic invaders, which multiply by the copy-and-paste process that defines retrotransposons, are targeted during the first few days of embryogenesis by silencing mechanisms notably involving their recognition by sequence-specific protein- or RNA-based repressors and the secondary recruitment of heterochromatin-inducing complexes (Rowe and Trono 2011). Histone methylation, histone deacetylation, and DNA methylation ensue, which inactivate their potential as insertional mutagens (Kaer and Speek 2013; Shukla et al. 2013) and repress their *cis*-acting transcriptional components, which would otherwise activate neighboring genes via promoter or enhancer effects (Bourque et al. 2008; Kunarso et al. 2010; Macfarlan et al. 2011; Rebollo et al. 2012; Schmidt et al. 2012; Chuong et al. 2013; Rowe et al. 2013).

The epigenetic control of EREs is a rigorously orchestrated process, with some of these elements never expressed, and others transcribed in low-cellularity embryos or in embryonic stem cells (ESCs) to be silenced only later (Lane et al. 2003; Macfarlan et al. 2011). It has recently emerged that transient transcription driven

by specific coopted ERE integrants that briefly evade repression seems critical to the identity of the pluripotent state both in mice (Macfarlan et al. 2012) and humans (Santoni et al. 2012; Lu et al. 2014). This layered repression process likely reflects at least in part the orderly action of cognate repressors, including KRAB-containing zinc finger proteins (KRAB-ZFPs), which together with their cofactor TRIM28 (also known as KAP1 or TIF1B) are key to the early embryonic control of a broad spectrum of retrotransposons (Wolf and Goff 2007; Matsui et al. 2010; Rowe et al. 2010). Interestingly, KRAB-ZFPs, encoded in the hundreds by the genomes of both mouse and human, are widely expressed in ESCs but subsequently adopt highly tissue-, stage-, and cell-specific patterns of expression (Barde et al. 2013; Corsinotti et al. 2013). Whether the same holds true for other yet-to-be-identified early embryonic controllers of EREs is unknown. However, it is likely that any lag between the de-repression of specific ERE integrants and the reactivation of their sequence- or class-specific repressors will open a window of opportunity for ERE-originating transcriptional perturbations and, at least for the small fraction of these elements still endowed with retrotransposition ability (Finnegan 2012), for insertional mutagenesis. Accordingly, the present study was undertaken to examine how the transcriptional control of EREs is maintained during the reprogramming of either murine or human somatic cells, and whether it is fully reestablished in induced pluripotent stem cells.

Results

Global ERE de-repression during the reprogramming of mouse and human cells

In order to assess the control of EREs during reprogramming to pluripotency, we transduced mouse embryonic fibroblasts (MEFs) harboring a *Pou5f1*-GFP transgene with a lentiviral vector expressing *POU5F1 (OCT4)*, *KLF4*, and *SOX2* (hereafter called OKS) as a single polycistronic transcript from a spleen focus forming virus (SFFV) promoter (Supplemental Fig. S1A; Pasi et al. 2011). As expected, OKS-transduced MEFs formed GFP-positive colonies after ~12 d (Supplemental Fig. S1B) and silenced the OKS vector (Supplemental Fig. S1C). We picked and replated a series of iPSC clones and verified that they expressed *Nanog* at similar levels as ESCs (Supplemental Fig. S1D) and could be differentiated into embryoid bodies (EBs) (Pasi et al. 2011), indicating successful reprogramming. We then used real-time quantitative PCR (RT-QPCR) to measure the expression of families of EREs at various times of the reprogramming process and in the resulting iPSCs. We consistently observed marked increases in the transcript levels of all tested EREs in independent reprogramming experiments (Fig. 1A–D). The timing of the up-regulation was variable, usually initiating 6–10 d post-transduction with a burst of expression occasionally observed around day 2. In all iPSC clones tested, LINE1 and the ERV MusD remained highly expressed, whereas IAP (intracisternal A particle, another ERV) exhibited a fully repressed state, so that for these families of retroelements expression patterns were roughly comparable between iPSC and ESC, as previously noted (Wissing et al. 2012). This is consistent with a model whereby the *trans*-acting factors controlling IAPs, but not those responsible for silencing LINE1 and MusD, are reactivated during the late stages of reprogramming. MERVL, another ERV, exhibited little change in expression during reprogramming itself, but rose sharply in iPSCs, where its levels were markedly above those measured in ESCs. Rather than a defect in controlling factors, this could reflect higher fractions of cells cycling into an early post-zygotic-like state where the MERVL long terminal repeat (LTR), which serves as a

promoter for many genes restricted to the 2/4 cell stage of embryonic development, is particularly active (Macfarlan et al. 2012).

We next assessed the control of EREs during the reprogramming of human somatic cells. Upon transduction of cord blood CD34+ cells with the OKS vector, iPSC clones were efficiently obtained (Supplemental Fig. S1B), which exhibited a morphology comparable to that of ESCs, expressed a full set of pluripotency genes, induced the formation of teratomas when injected into immunodeficient mice, and were karyotypically euploid (Supplemental Fig. S1E–H). However, de-repression of all tested human ERE families was detected by RT-QPCR during the reprogramming process, with sharp increases around day 19 post-transduction across three independent experiments performed with cells from different donors (Fig. 1E,H; Supplemental Fig. S2). Similar to their murine counterparts, human LINE1s remained highly expressed in pluripotent cells, as did SVAs (Fig. 1E,F). In contrast, the LTR-containing HERVK, including its HERVK14ci strain, was silenced in iPSCs, mimicking the behavior of IAPs in the murine system (Fig. 1G,H).

To confirm and extend these observations, we performed RNA deep sequencing (RNA-seq) at multiple time points of an independent reprogramming experiment, including in our analysis the parental CD34+ cells, six of the resulting iPSC clones, and a human ESC clone (H1) (for RNA-seq data, see Supplemental Tables 1–3, and GEO accession number GSE57866). A comparison of the transcriptomes of cells harvested at day 19 post-transduction with that of the starting population confirmed the de-repression of multiple EREs, and further identified specific subclasses of dysregulated retrotransposons (Fig. 2A), with HERVH, HERVK, and their associated LTRs LTR7Y and LTR5-Hs displaying the most pronounced up-regulation (Fig. 2B). In iPSCs, some of these elements, such as HERVK and S71 (LTR6b), displayed a repressed state comparable to that found in ESCs, but others (e.g., HERVW, LTR17, LTR7Y, L1Hs) did not or did only partially (Fig. 2B,C). Of note, while significant expression of HERVH was detected in iPSCs and ESCs, as previously reported (Santoni et al. 2012), close to a third of the HERVH integrants up-regulated at day 19 were not among those ultimately detected in these cells (data not shown). Finally, in addition to the d19 peak of expression, we occasionally observed a much earlier burst around d2–d4 for some retroelements (e.g., HERVH and HERVK) (Fig. 2C). We further RNA-sequenced eight independent human ES cell lines (UCLA1-6 [Diaz Perez et al. 2012], H19, and an independent H1 sample) to verify that the elevated ERE expression levels detected in individual iPS clones were not a general feature of pluripotent cells. We found very little heterogeneity between all tested human ES cell lines, which sharply contrasted with the inappropriate control of EREs in iPS clones and even more with their marked up-regulation during the reprogramming process (Fig. 2C). To ascertain that the observed unleashing of EREs during reprogramming is not restricted to fibroblasts or blood cells, we assessed ERE expression during reprogramming of human liver cells. We thus reprogrammed primary human hepatocytes with OKS or with a vector that also included MYC (OKSM) (Sommer et al. 2010) and measured ERE-specific transcription by real time quantitative PCR. Confirming the generality of this phenomenon, all ERE classes recapitulated the burst of expression during reprogramming previously observed in MEFs and CD34+ cells (Fig. 2D).

Deregulation of ERE-close gene transcripts during reprogramming

We next asked whether the up-regulation of ERVs that occurred during reprogramming impacted the expression of nearby genes.
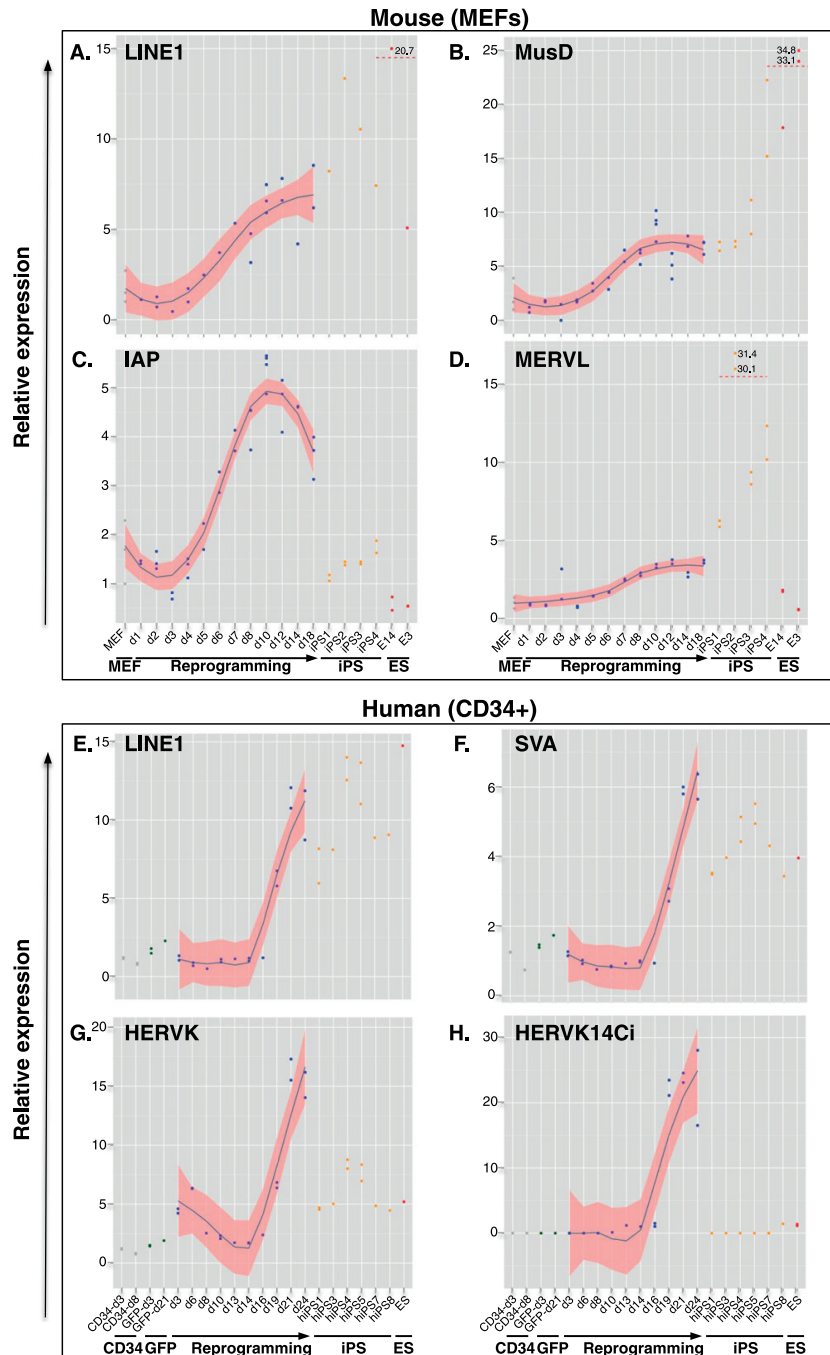
Figure 1. Global up-regulation of EREs during reprogramming. qPCR quantification of transcripts from indicated murine (*A–D*) and human (*E–H*) EREs during reprogramming of MEFs and cord blood CD34[+] cells, respectively. Expression levels are indicated for parental cells (gray dots), cells transduced with an SFFV-GFP control vector (green dots), OKS-induced reprogramming time points (blue dots), individual iPS clones (orange dots), and ES cells (red dots). For the smoothing pattern across OKS-induced reprogramming time points, we computed a locally weighted polynomial regression (LOESS) with a 95% confidence interval.

148
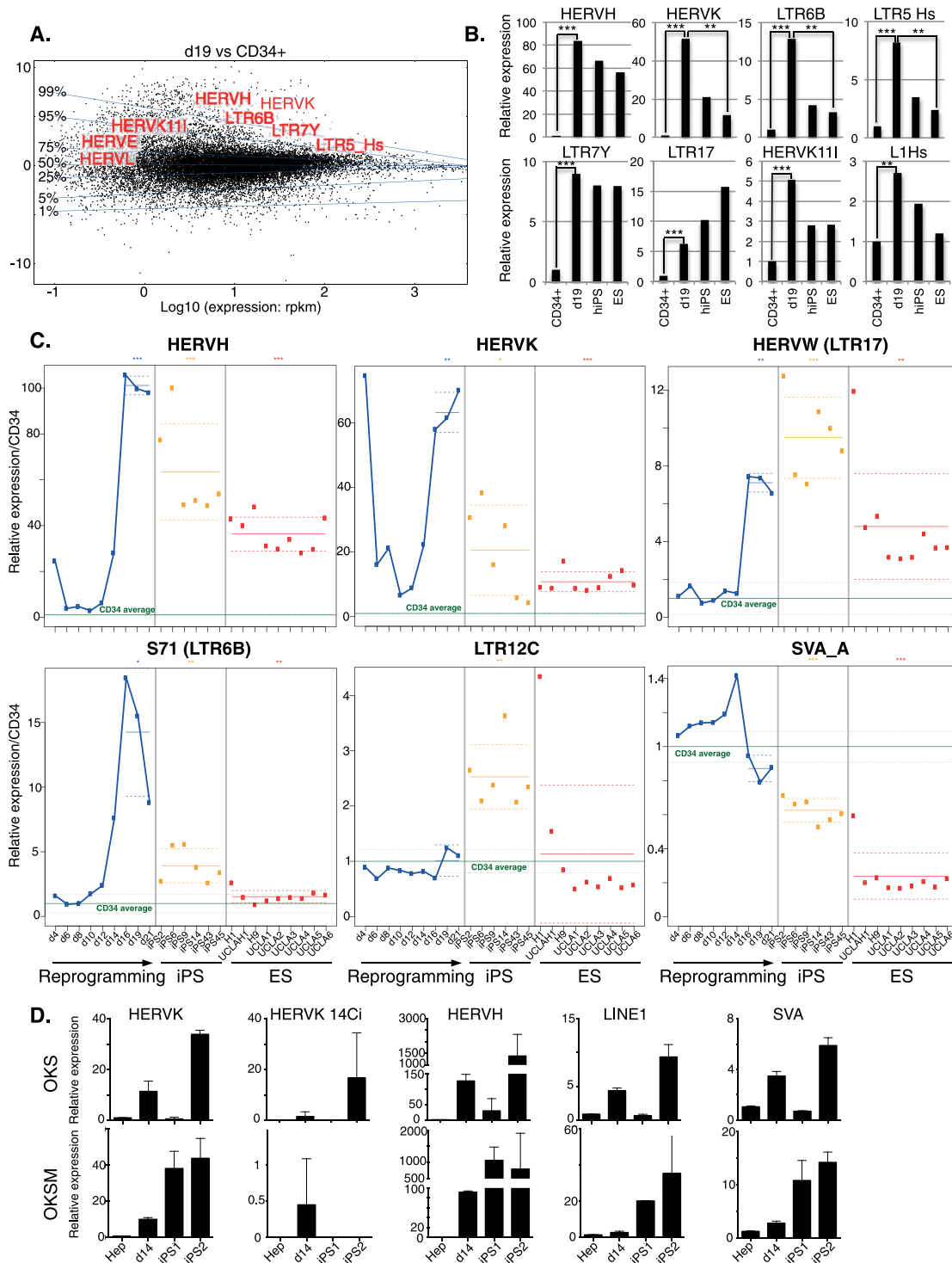
Figure 2. (Legend on next page)

149

Using a twofold change cutoff and an adjusted P-value of <0.05, we identified 3703 genes up-regulated at d19 in OKS-transduced CD34$^+$ cells compared with nontransduced cells. Genes near up-regulated HERVH (<40 kb between TSS and HERV, n = 365 genes) were far more likely ($P = 2.603 \times 10^{-05}$) to be up-regulated than genes more distant from these elements (>80 kb between TSS and HERV, n = 18,145 genes) (Fig. 3A). A similar analysis could not be performed for HERVK due to insufficient numbers. However, two additional lines of evidence confirmed that it was the HERV that influenced the gene and not the reverse. First, for both HERVH and HERVK, expression was not influenced by distance from an up-regulated gene (Fig. 3A). Second, intragenic HERVs were not more often up-regulated than their intergenic counterparts (Fig. 3B); in fact, the reverse trend was observed (P = 0.024 for HERVK and $P = 4.26 \times 10^{-10}$ for HERVH).

Heterogeneity of human iPSC clones in repression of specific EREs and induction of prototypic ERE controllers

Importantly, when we compared levels of expression of individual EREs between iPSC clones derived from a single donor and issued from the same reprogramming experiment, we noticed striking differences, notably for HERVH, HERVK, and L1Hs (e.g., cf. clone 6 and clone 43 or cf. clone 2 and clone 45) (Figs. 2C, 4A). HERVK, for instance, was fivefold to ninefold more expressed in iPS2 and iPS6 compared with iPS43, iPS45, and human ES cells (Fig. 4A, top), whereas L1Hs was threefold more expressed in iPS2 than in iPS14. Interestingly, several members of the KRAB-ZFP gene family (e.g., ZNF492, ZNF649, ZNF208) exhibited marked differences in expression between iPSC clones (Supplemental Figs. S3A, S4C). A direct comparison of the transcriptomes of two iPSC clones confirmed that specific EREs and some KRAB-ZFPs were among the most discordant transcripts, suggesting that a failure to reactivate sequence-specific repressors during reprogramming might account for the lack of silencing of their target EREs in iPSCs (Supplemental Fig. S3B). Consistent with this hypothesis, *Trim28* expression increased gradually during reprogramming, and *Trim28* knockout MEFs failed to reprogram (data not shown). Noteworthy, known post-transcriptional controllers of retroelements (e.g., APOBEC3A and SAMHD1) (Bogerd et al. 2006; Hrecka et al. 2011; Laguette et al. 2011) were transiently induced during reprogramming (Supplemental Fig. S3A), indicating the activation of at least some genome defense mechanisms along this process.

Incomplete control of specific EREs activates the transcription of nearby genes in iPSCs

Upon scoring the expression of specific ERE integrants, we again detected considerable heterogeneity between iPSC clones (Fig. 4A,

bottom). For example, a HERVH located on chromosome 8 was four- to fivefold more expressed in iPS6 compared with other iPSC clones and to human ES, while a HERVK inserted on chromosome 22 was four- to eightfold more expressed in clones 2, 6, and 14 compared with clones 43, 45, and human ESCs (Fig. 4A, bottom). EREs can epigenetically affect the local genomic landscape owing to their content in a variety of *cis*-acting regulatory sequences (Bourque et al. 2008; Kunarso et al. 2010; Rebollo et al. 2012; Schmidt et al. 2012; Chuong et al. 2013; Rowe et al. 2013). Correspondingly, we found numerous instances where lack of repression of a specific ERE was accompanied by up-regulation of the adjacent gene (Fig. 4B,C; Supplemental Fig. S4A,B). For example, the *HHLA1* gene, situated next to the above-mentioned chromosome-8 HERVH, was significantly expressed only in iPS6 (Spearman correlation P = 0.93, P = 0.007), while expression of the *PRODH* gene correlated perfectly with that of the adjacent chromosome-22 HERVK among iPSC clones (Spearman correlation P = 0.96, P = 0.003) (Fig. 4B,C). Likewise, expression of *KLKB1* and *C9orf129* paralleled that of HERVs situated nearby (Supplemental Fig. S4A,B). Interestingly, with the exception of the HERV near *PRODH* (for which mapping of ChIP-seq reads was not possible), all these EREs were previously identified as bound by TRIM28 and adorned with the H3K9me3 repressive mark in human ESCs (Fig. 4B,C; Supplemental Fig. S4A,B; data not shown). Noteworthy, when differentiated to EBs or neural committed, some ERE de-repressed iPS clones regained control of specific integrants while others remained uncontrolled (Supplemental Fig. S5). For example, the ERE near *PRODH* in iPS6 was still locked in the highly expressed state after neural commitment even though expression of *PRODH* itself was reduced about threefold (Supplemental Fig. S5).

To further explore this phenomenon, we examined the state of the chromatin at several of these loci by chromatin immunoprecipitation followed by quantitative PCR (ChIP-PCR). We found histone marks typical of active enhancers (H3K4me1 and H3K27ac) near the LTRs of the HERVs situated next to *HHLA1* and *PRODH* in iPSCs exhibiting an up-regulation of the corresponding ERE–gene pairs, but not in clones where these units were fully repressed (Fig. 4D). Conversely H3K9me3, a repressive mark that is a hallmark of KRAB/TRIM28-mediated silencing (Schultz et al. 2002), displayed the opposite pattern at these loci.

Discussion

These data demonstrate that the reprogramming of somatic cells to induced pluripotent stem cells is accompanied by a marked de-repression of endogenous retroelements from all known classes. It has recently been suggested that transcription from EREs is important to drive ES-specific transcripts, in particular MERVL in

---

Figure 2. De-repression of individual EREs during reprogramming and in iPSCs. (*A*) MA-plot comparing RNA-seq–determined transcriptome of day 19 (d19) OKS-transduced vs. parental CD34$^+$ cells. Transcripts (RefSeq) are plotted in black with the ratio (d19/CD34$^+$) on the *y*-axis and expression levels on the *x*-axis. Representative up-regulated Repbase families are shown in red. Transversal blue lines depict magnitude of gene deregulation (e.g., only 1% of genes lie *above* the 99% line). (*B*) Expression levels of indicated HERVs in parental (average of three samples) or d19 OKS-transduced CD34$^+$ cells, human iPS cells (average of six clones from the same reprogramming experiment), and the H1 ES cell line (average of two samples). Fold changes compared with CD34$^+$ triplicates and P-values are calculated using the DESeq package (Anders and Huber 2010). (*C*) Relative expression of indicated HERVs during reprogramming of CD34$^+$ cells, in six resulting iPSC clones (orange dots) and nine independent samples of hES cells (red dots). Green line indicates parental cells (average of three, same donor); dotted green lines, plus and minus standard deviation; and solid blue line, reprogramming time points. The horizontal (blue, orange, and red) solid lines show the mean of each group of samples. Dotted lines show a 95% confidence interval for each mean. We performed a Wilcoxon test for each mean to test if it was different from one. (*D*) qPCR quantification of transcripts from indicated human EREs during reprogramming of primary hepatocytes. (*Top*) Reprogramming with OKS. (*Bottom*) Reprogramming with OKSM. (Hep) Average of four nontransduced hepatocyte samples.
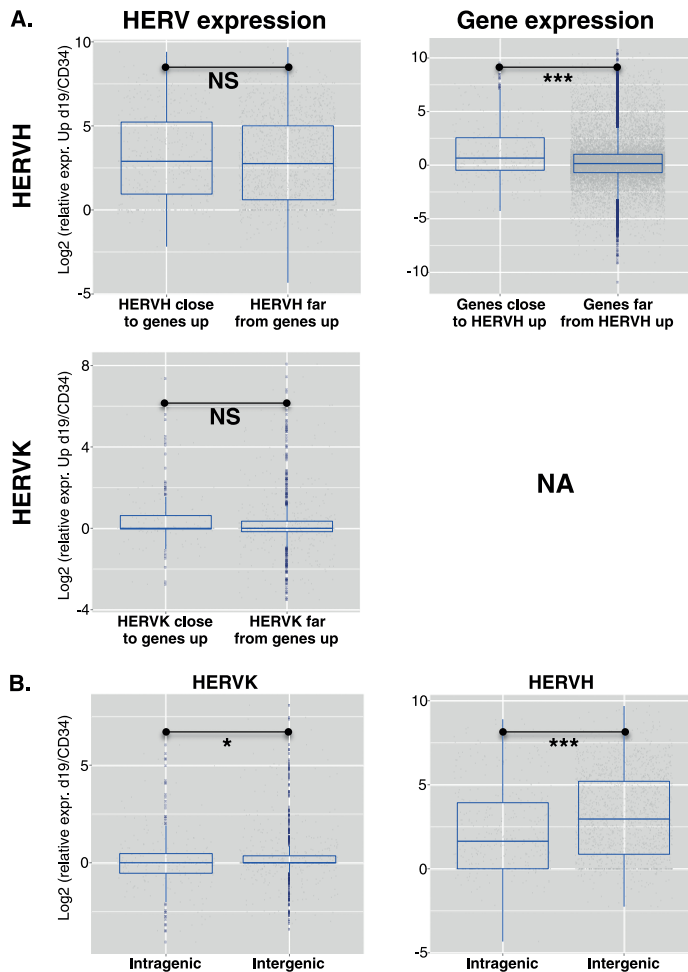
Figure 3. Up-regulation of ERE-close genes during reprogramming. (*A*) Using a twofold change cutoff and an adjusted *P*-value of <0.05, 3703 genes were found up-regulated in CD34$^+$ cells at day 19 post-OKS transduction, compared with untransduced cells. (*Left*) Relative expression of HERVH (*top*) or HERVK (*bottom*) integrants situated close (<40 kb) or far (>80 kb) from these genes. No significant difference was detected (Wilcoxon test *P* = 0.14 for HERVK, 0.11 for HERVH). (*Right*) Reverse analysis, revealing that HERVH-close transcripts (*n* = 365 genes, or 695 transcripts) were more likely to be up-regulated than HERVH-distant transcripts (*n* = 18145 genes, or 30,053 transcripts). (NA) Not available due to insufficient numbers. (*B*) Expression levels of intragenic vs. intergenic HERVHs or HERVKs at d19 of CD34$^+$ cells reprogramming showing no bias toward intragenic HERVs. In fact, the reverse trend was observed, with more intergenic than intragenic HERVs up-regulated at d19 of reprogramming.

murine cells (Macfarlan et al. 2012) and HERVH in human cells (Santoni et al. 2012; Lu et al. 2014). However, while we indeed found numerous HERVH integrants up-regulated in human ES cells compared with cord blood CD34$^+$ cells, we also noted that more than a third of the members of this group found to be activated at day 19 of reprogramming were fully repressed in ESCs or iPSCs. Thus the possible requirement of HERVH-mediated transcription during iPS reprogramming and to maintain the pluripotent state comes at a price, as control is broadly released on this family of elements. Furthermore, HERVKs, including HERVK14ci, were markedly up-regulated during reprogramming but fully

repressed in pluripotent cells. For most EREs, activation reached its peak shortly before the emergence of fully reprogrammed iPSCs. Our results suggest that this phenomenon might be due at least partly to a lack of synchronization between erasure of repressive chromatin marks at EREs on the one hand and reactivation of sequence-specific *trans*-repressors, for instance, KRAB-ZFPs, on the other hand. Importantly, at that stage, genes situated near up-regulated EREs had a greater chance of being themselves induced, consistent with ERE-based promoter or enhancer effects (Rowe et al. 2013). It could be that this contributes to the inefficiency of reprogramming, if it results in the stochastic activation of genes affecting the path to pluripotency (Polo et al. 2012). It will be interesting to determine whether ERE activation also occurs when reprogramming efficiency is increased by depletion of the MBD3 repressor (deterministic reprogramming) (Rais et al. 2013), by coexpression of CEBPA (Di Stefano et al. 2014), or following nuclear transfer. These faster reprogramming methods may be accompanied by timely reactivation of cognate KRAB-ZFPs, which may minimize aberrant ERE reactivation.

ERE de-repression is predicted to result only rarely in mutagenic transposition events, owing both to the paucity of retroelements still endowed with transposition capacity (Quinlan et al. 2011; Finnegan 2012) and to the presence of restriction factors blocking their spread at a post-transcriptional level (Wolf and Goff 2008). However, the transcriptional perturbation of ERE-close genes may confer iPSCs or their progeny with phenotypic anomalies difficult to detect through conventional assays, such as blockade of differentiation to particular lineages, predisposition to oncogenic changes, aberrant release of bioactive molecules, or altered immunogenicity. Supporting this model, a recent comparative analysis of 49 iPSC lines derived from several human tissues detected an aberrant up-regulation of some LTR7/HERVH transcripts and neighboring genes, including *HHLA1*, in several differentiation-defective clones (Koyanagi-Aoi et al. 2013). Furthermore, the potential for more distal phenotypic anomalies resulting from inappropriate ERE-induced gene activation is illustrated by the recent observation that the schizophrenia-linked *PRODH* gene (Kempf et al. 2008) is controlled by the nearby HERVK (Suntsova et al. 2013), which we found deregulated in some iPS clones. Our findings thus warrant an in-depth survey of the genomic, transcriptional, and epigenetic state of the repetitive genome of iPSC clones, whether these are to be used for basic research or are envisioned for clinical applications.

151

Figure 4. Clonal variability of ERE control in iPSCs. (*A*) Comparative expression of indicated ERE families (*top*) in parental CD34[+] cells (average of three samples), in six iPSC clones simultaneously derived from their reprogramming, and in control H1 ES cells. (*Bottom*) Expression levels of individual HERV integrants in same cells as *A* (*top*). (*B*,*C*) Two examples of dysregulated HERV integrants and expression level of closest genes, *HHLA1* for a HERVH on chromosome 8 (*B*) and *PRODH* for a HERVK on chromosome 22 (*C*). (*Left*) Expression level tracks, as well as hES H3K9me3 and TRIM28 binding data (obtained by ChIP-seq). Note that for the *PRODH* locus, ChIP-seq reads could not be mapped due to the high redundancy of this region, preventing binding site calling at this location. Red arrowheads indicate HERV orientation. (*Right*) Expression levels of ERE–gene pairs, with Spearman correlation calculated. (*D*) Activation (H3K4me1 and H3K27ac) and repressive (H3K9me3) histone marks at HERVs situated close to *HHLA1* and *PRODH* and controls (activation mark negative controls, *ZNF180* and *ZNF420*; positive controls, *POU5F1* [*OCT4*] and *NANOG* promoters; H3K9me3 negative controls, *EVX1* and *GAPDH*; positive controls, *ZNF180* and *ZNF420*) depicted as a heatmap from least (blue) to most (brown) enriched. Two qPCR assays (e.g., ERE *PRODH* 5′ and ERE *PRODH* 5′ #2) were designed on the 5′ and 3′ of each ERE. Note active enhancer marks on the ERE near *PRODH* in iPS clones 2, 6, and 14 compared with iPS clones 43 and 45, and on the ERE near *HHLA1* in iPS clone 6 compared with other clones. Up-regulated ERE–gene pairs are highlighted with a red box for appropriate iPS clones.

152

Methods

Reprogramming

MEFs: Primary *Pou5f1*-GFP MEFs were prepared from E12 embryos (http://jaxmice.jax.org/strain/008214.html) and reprogrammed by transduction with OKS using four HCT116-transducing units (HC-TU) per cell as previously described (Pasi et al. 2011). CD34$^+$ cells from human cord blood were obtained and prepared as previously described (Barde et al. 2013), before transducing 250,000 cells with OKS using 100 HC-TU per cell. After 5 d, cells were switched to mTeSR1 medium (Stemcell Technologies no. 05859) and grown on a mouse fibroblast feeder layer until reprogrammed colonies emerged (~21 d). Individual human iPSC clones were then picked and expanded. Primary human hepatocytes were isolated from liver biopsies as previously described (Birraux et al. 2009) and plated on Matrigel or collagen before being transduced with OKS or OKSM using 20 HC-TU per cell. After 5 d, cells were switched to mTeSR1 medium (Stemcell Technologies no. 05859) and grown until reprogrammed colonies emerged (~25 d).

qPCR

Total RNA from cells at different reprogramming time points, iPSC clones, and ES cells was extracted with TRIzol reagent (Life Technologies no. 15596-018) and PureLink micro-to-midi total RNA Purification System (Life Technologies no.12183018). cDNA was prepared with SuperScript II reverse transcriptase and real-time PCR quantification was performed with FastStart Universal SYBR Green Master (Rox; Roche no. 04913914001). Normalization was done to two or three housekeeping genes (mouse: *Gapdh*, *Cox6a1*, *Tfrc*; human: *TFRC*, *B2M*).

Chromatin immunoprecipitation (ChIP)-PCR

Ten million iPSCs or ESCs were immunoprecipitated as previously described (Barde et al. 2013), with TRIM28- (Abcam ab10483), H3K9me3- (Diagenode pAb-056-050), H3K27ac- (Abcam ab4729), or H3K4me1- (Diagenode pAb-037-050) specific antibodies. SYBR green qPCR was performed to quantify enrichment at specific loci.

RNA-seq

RNA-seq was performed on an Illumina HiSeq 2000 sequencer (single-read 100-cycles assay). The library was generated from 250 ng total RNA using the TruSeq RNA Sample Preparation v2 kit (Illumina). Raw reads (100-bp single-end) were mapped to the human transcriptome (RefSeq), to the human genome (hg19 assembly), and to Repbase consensus sequences using the Bowtie short-read aligner (Langmead et al. 2009) and allowing up to three mismatches, and counts were normalized to the transcript length and to the total number of reads (RPKM). Differentially expressed RefSeq transcripts and EREs were defined using the DESeq Bioconductor package (Anders and Huber 2010).

Data access

The RNA-seq data from this study have been submitted to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE57866.

Acknowledgments

References

Anders S, Huber W 2010. Differential expression analysis for sequence count data. *Genome Biol* **11:** R106.

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

Birraux J, Menzel O, Wildhaber B, Jond C, Nguyen TH, Chardot C. 2009. A step toward liver gene therapy: efficient correction of the genetic defect of hepatocytes isolated from a patient with Crigler-Najjar syndrome type 1 with lentiviral vectors. *Transplantation* **87:** 1006–1012.

Bogerd HP, Wiegand HL, Doehle BP, Lueders KK, Cullen BR. 2006. APOBEC3A and APOBEC3B are potent inhibitors of LTR-retrotransposon function in human cells. *Nucleic Acids Res* **34:** 89–95.

Bourque G, Leong B, Vega VB, Chen X, Lee YL, Srinivasan KG, Chew JL, Ruan Y, Wei CL, Ng HH, et al. 2008. Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Res* **18:** 1752–1762.

Chuong EB, Rumi MA, Soares MJ, Baker JC. 2013. Endogenous retroviruses function as species-specific enhancer elements in the placenta. *Nat Genet* **45:** 325–329.

Corsinotti A, Kapopoulou A, Gubelmann C, Imbeault M, Santoni de Sio FR, Rowe HM, Mouscaz Y, Deplancke B, Trono D. 2013. Global and stage specific patterns of Kruppel-associated-box zinc finger protein gene expression in murine early embryonic cells. *PLoS ONE* **8:** e56721.

de Koning AP, Gu W, Castoe TA, Batzer MA, Pollock DD. 2011. Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet* **7:** e1002384.

Di Stefano B, Sardina JL, van Oevelen C, Collombet S, Kallin EM, Vicent GP, Lu J, Thieffry D, Beato M, Graf T. 2014. C/EBPα poises B cells for rapid reprogramming into induced pluripotent stem cells. *Nature* **506:** 235–239.

Diaz Perez SV, Kim R, Li Z, Marquez VE, Patel S, Plath K, Clark AT. 2012. Derivation of new human embryonic stem cell lines reveals rapid epigenetic progression in vitro that can be prevented by chemical modification of chromatin. *Hum Mol Genet* **21:** 751–764.

Finnegan DJ. 2012. Retrotransposons. *Curr Biol* **22:** R432–R437.

Hanna J, Saha K, Pando B, van Zon J, Lengner CJ, Creyghton MP, van Oudenaarden A, Jaenisch R. 2009. Direct cell reprogramming is a stochastic process amenable to acceleration. *Nature* **462:** 595–601.

Hrecka K, Hao C, Gierszewska M, Swanson SK, Kesik-Brodacka M, Srivastava S, Florens L, Washburn MP, Skowronski J. 2011. Vpx relieves inhibition of HIV-1 infection of macrophages mediated by the SAMHD1 protein. *Nature* **474:** 658–661.

Jaenisch R, Young R. 2008. Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. *Cell* **132:** 567–582.

Kaer K, Speek M. 2013. Retroelements in human disease. *Gene* **518:** 231–241.

Kempf L, Nicodemus KK, Kolachana B, Vakkalanka R, Verchinski BA, Egan MF, Straub RE, Mattay VA, Callicott JH, Weinberger DR, et al. 2008. Functional polymorphisms in PRODH are associated with risk and protection for schizophrenia and fronto-striatal structure and function. *PLoS Genet* **4:** e1000252.

Koche RP, Smith ZD, Adli M, Gu H, Ku M, Gnirke A, Bernstein BE, Meissner A. 2011. Reprogramming factor expression initiates widespread targeted chromatin remodeling. *Cell Stem Cell* **8:** 96–105.

Koyanagi-Aoi M, Ohnuki M, Takahashi K, Okita K, Noma H, Sawamura Y, Teramoto I, Narita M, Sato Y, Ichisaka T, et al. 2013. Differentiation-defective phenotypes revealed by large-scale analyses of human pluripotent stem cells. *Proc Natl Acad Sci* **110:** 20569–20574.

Kunarso G, Chia NY, Jeyakani J, Hwang C, Lu X, Chan YS, Ng HH, Bourque G. 2010. Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat Genet* **42:** 631–634.

Laguette N, Sobhian B, Casartelli N, Ringeard M, Chable-Bessia C, Segeral E, Yatim A, Emiliani S, Schwartz O, Benkirane M. 2011. SAMHD1 is the dendritic- and myeloid-cell-specific HIV-1 restriction factor counteracted by Vpx. *Nature* **474:** 654–657.

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

Lane N, Dean W, Erhardt S, Hajkova P, Surani A, Walter J, Reik W. 2003. Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. *Genesis* **35:** 88–93.

Lu X, Sachs F, Ramsay L, Jacques PE, Goke J, Bourque G, Ng HH. 2014. The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. *Nat Struct Mol Biol* **21:** 423–425.

Macfarlan TS, Gifford WD, Agarwal S, Driscoll S, Lettieri K, Wang J, Andrews SE, Franco L, Rosenfeld MG, Ren B, et al. 2011. Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A. *Genes Dev* **25:** 594–607.

Macfarlan TS, Gifford WD, Driscoll S, Lettieri K, Rowe HM, Bonanomi D, Firth A, Singer O, Trono D, Pfaff SL. 2012. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487:** 57–63.

Matsui T, Leung D, Miyashita H, Maksakova IA, Miyachi H, Kimura H, Tachibana M, Lorincz MC, Shinkai Y. 2010. Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature* **464:** 927–931.

Pasi CE, Dereli-Oz A, Negrini S, Friedli M, Fragola G, Lombardo A, Van Houwe G, Naldini L, Casola S, Testa G, et al. 2011. Genomic instability in induced stem cells. *Cell Death Differ* **18:** 745–753.

Polo JM, Anderssen E, Walsh RM, Schwarz BA, Nefzger CM, Lim SM, Borkent M, Apostolou E, Alaei S, Cloutier J, et al. 2012. A molecular roadmap of reprogramming somatic cells into iPS cells. *Cell* **151:** 1617–1632.

Quinlan AR, Boland MJ, Leibowitz ML, Shumilina S, Pehrson SM, Baldwin KK, Hall IM. 2011. Genome sequencing of mouse induced pluripotent stem cells reveals retroelement stability and infrequent DNA rearrangement during reprogramming. *Cell Stem Cell* **9:** 366–373.

Rais Y, Zviran A, Geula S, Gafni O, Chomsky E, Viukov S, Mansour AA, Caspi I, Krupalnik V, Zerbib M, et al. 2013. Deterministic direct reprogramming of somatic cells to pluripotency. *Nature* **502:** 65–70.

Rebollo R, Miceli-Royer K, Zhang Y, Farivar S, Gagnier L, Mager DL. 2012. Epigenetic interplay between mouse endogenous retroviruses and host genes. *Genome Biol* **13:** R89.

Rowe HM, Trono D. 2011. Dynamic control of endogenous retroviruses during development. *Virology* **411:** 273–287.

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

Rowe HM, Kapopoulou A, Corsinotti A, Fasching L, Macfarlan TS, Tarabay Y, Viville S, Jakobsson J, Pfaff SL, Trono D. 2013. TRIM28 repression of retrotransposon-based enhancers is necessary to preserve transcriptional dynamics in embryonic stem cells. *Genome Res* **23:** 452–461.

Santoni FA, Guerra J, Luban J. 2012. HERV-H RNA is abundant in human embryonic stem cells and a precise marker for pluripotency. *Retrovirology* **9:** 111.

Schmidt D, Schwalie PC, Wilson MD, Ballester B, Goncalves A, Kutter C, Brown GD, Marshall A, Flicek P, Odom DT. 2012. Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell* **148:** 335–348.

Schultz DC, Ayyanathan K, Negorev D, Maul GG, Rauscher FJ III. 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.

Shukla R, Upton KR, Munoz-Lopez M, Gerhardt DJ, Fisher ME, Nguyen T, Brennan PM, Baillie JK, Collino A, Ghisletti S, et al. 2013. Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma. *Cell* **153:** 101–111.

Sommer CA, Sommer AG, Longmire TA, Christodoulou C, Thomas DD, Gostissa M, Alt FW, Murphy GJ, Kotton DN, Mostoslavsky G. 2010. Excision of reprogramming transgenes improves the differentiation potential of iPS cells generated with a single excisable vector. *Stem Cells* **28:** 64–74.

Suntsova M, Gogvadze EV, Salozhin S, Gaifullin N, Eroshkin F, Dmitriev SE, Martynova N, Kulikov K, Malakhova G, Tukhbatova G, et al. 2013. Human-specific endogenous retroviral insert serves as an enhancer for the schizophrenia-linked gene PRODH. *Proc Natl Acad Sci* **110:** 19472–19477.

Takahashi K, Yamanaka S. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126:** 663–676.

Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S. 2007. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131:** 861–872.

Wissing S, Munoz-Lopez M, Macia A, Yang Z, Montano M, Collins W, Garcia-Perez JL, Moran JV, Greene WC 2012. Reprogramming somatic cells into iPS cells activates LINE-1 retroelement mobility. *Hum Mol Genet* **21:** 208–218.

Wolf D, Goff SP 2007. TRIM28 mediates primer binding site-targeted silencing of murine leukemia virus in embryonic cells. *Cell* **131:** 46–57.

Wolf D, Goff SP 2008. Host restriction factors blocking retroviral replication. *Annu Rev Genet* **42:** 143–163.

Yamanaka S 2009. Elite and stochastic models for induced pluripotent stem cell generation. *Nature* **460:** 49–52.

Loss of transcriptional control over endogenous retroelements during reprogramming to pluripotency

Marc Friedli, Priscilla Turelli, Adamandia Kapopoulou, et al.

Supplemental Material	http://genome.cshlp.org/content/suppl/2014/06/12/gr.172809.114.DC1.html
References	This article cites 46 articles, 10 of which can be accessed free at: http://genome.cshlp.org/content/24/8/1251.full.html#ref-list-1
Creative Commons License	This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see http://genome.cshlp.org/site/misc/terms.xhtml). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/.
Email Alerting Service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**

To subscribe to *Genome Research* go to:
http://genome.cshlp.org/subscriptions

155

Gabriela Ecco

Date of birth: December 11th 1986
Nationalities: Brazilian, Italian
Address: Chemin de la Vulliette 7
        CH1000   Lausanne, Switzerland
Contact: gabriela.ecco@epfl.ch
        +41 787943700

Education

PhD IN BIOTECHNOLOGY AND BIOENGINEERING, EPFL; LAUSANNE, SWITZERLAND – 2011-PRESENT

Expected graduation date: July 2016
Supervisor: Prof Didier Trono, MD
PhD thesis: Unraveling the KRAB/KAP1 control of transposable elements in pluripotent and somatic cells

BACHELOR OF BIOLOGICAL SCIENCES, UFSC; FLORIANÓPOLIS, BRAZIL – 2006-2011

GPA: 9.75/10.00, *cum laude*
Bachelor thesis: Effect of S-nitrosylation on *Mycobacterium tuberculosis* tyrosine phosphatases PtpA and PtpB

Research Experience

PHD THESIS: JULY 2011 - PRESENT

Laboratory of Virology and Genetics, École Polytechnique Fédérale de Lausanne, Switzerland
Supervisor: Prof. Didier Trono, MD
- Developed a screening system to identify transcription factors binding to specific DNA sequences and characterized candidate proteins function in pluripotent and differentiated cells

PHD INTERNSHIP: JULY 2015- AUGUST 2015

Department of Oral and Maxillofacial Surgery and Division of Tissue Engineering, The University of Tokyo Hospital, Tokyo
Supervisor: Prof Tsuyoshi Takato, MD
- Evaluated the effects of hydrogels in cartilage tissue engineering

UNDERGRADUATE RESEARCH: MARCH 2006 – FEBRUARY 2011

Center of Structural and Molecular Biology, Federal University of Santa Catarina, Brazil
Supervisor: Prof. Hernán Terenzi, PhD
- Participated on several projects on the characterisation of proteins function and structure

UNDERGRADUATE RESEARCH: MARCH 2009 – DECEMBER 2009

Laboratory of Cytogenetics, University Hospital, Federal University of Santa Catarina, Brazil
Supervisor: Prof. Eliana Ternes Pereira, MD
- Assisted with karyotype determination

UNDERGRADUATE RESEARCH: JULY 2008 – SEPTEMBER 2008

Summer Research Program 2008, Laboratory of Virology and Genetics, École Polytechnique Fédérale de Lausanne, Switzerland
Supervisor: Prof. Didier Trono, MD

- Investigated the importance of Trim5a protein domains in the restriction of HIV, SIV, N-MLV, and B-MLV virus

UNDERGRADUATE RESEARCH: MARCH 2008 – DECEMBER 2008
Laboratory of Cytogenetics, Federal University of Santa Catarina, Brazil
Supervisor: Prof. Maria Cecília Menks Ribeiro, PhD
- Assisted with the standardization of RT-PCR detection of hybrid transcripts in leukemic cells for the University Hospital

Oral Presentations

2015:
Endogenous retroelements and their KRAB-ZFP/KAP1 control system influence gene expression in adult tissues. FASEB meeting on Mobile DNA in Mammalian Genomes. Palm Beach, USA.

2011:
Mycobacterium tuberculosis tyrosine phosphatase A (PtpA) activity is modulated by Cys53 S-nitrosylation. SBBq Cone Sul Symposium, Annual meeting of the Brazilian Society for Biochemistry and Molecular Biology. Foz do Iguaçu, Brazil.

2008:
Inhibition of AtMYB2 DNA-binding by nitric oxide involves cysteine S-nitrosylation. XVIII Seminário de Iniciação Científica da UFSC [XVIII UFSC's Scientific Initiation Symposium]. Florianópolis, Brazil.

Posters

2014:
A functional screen identifies KRAB-containing zinc finger proteins responsible for controlling specific endogenous retroelements in murine embryonic stem cells. **Keystone meeting on Stem Cells and Reprogramming.** Olympic Valley, USA.

2013:
A functional screen for the identification and characterization of KRAB-containing zinc finger proteins. **Abcam meeting on Chromatin Structure and Function.** Grand Cayman, Cayman Islands.

2010:
Cys53 S-nitrosylation inhibits AtMYB2 DNA-binding. **XXXIX Annual Meeting of the Brazilian Society for Biochemistry and Molecular Biology.** Foz do Iguaçu, Brazil.

2008:
Expression, Purification and Initial Characterization of a Kynureninase from *Trypanossoma cruzi*. **XXXVII Annual Meeting of the Brazilian Society for Biochemistry and Molecular Biology and XI Congress of the PABMB.** Águas de Lindóia, Brazil.

2007:
Expression, Purification and Initial Characterization of a Kynureninase from *Trypanosoma cruzi*. **53rd Brazilian Conference of Genetics.** Águas de Lindóia, Brazil.

Awards and Fellowships

- Swiss Society for Molecular and Cellular Biosciences (SSMCB) Travel Grant 2015.
- Fellowship PIBIC/CNPq (Institutional Program for Scientific Initiation Fellowships/Brazilian National Council for Scientific and Technological Development), 2006-2011.
- 3rd place in the 6th Award Highlight of the Year in Scientific Initiation Life Sciences, 2008, CNPq.

Publications

1. **G. Ecco**, M. Cassano, A. Kauzlaric, J. Duc, A. Coluccio, S. Offner, M. Imbeault, H. M. Rowe, P. Turelli, and D. Trono, 'Transposable elements and their KRAB-ZFP controllers regulate gene expression in adult tissues', *Dev Cell*, (2016), *In press*.
2. **G. Ecco**, H. M. Rowe, and D. Trono, 'A Large-Scale Functional Screen to Identify Epigenetic Repressors of Retrotransposon Expression', *Methods Mol Biol*, 1400 (2016), 403-17.
3. N. Castro-Diaz, **G. Ecco**, A. Coluccio, A. Kapopoulou, B. Yazdanpanah, M. Friedli, J. Duc, S. M. Jang, P. Turelli, and D. Trono, 'Evolutionly Dynamic L1 Regulation in Embryonic Stem Cells', *Genes Dev,* 28 (2014), 1397-409.
4. M. Friedli, P. Turelli, A. Kapopoulou, B. Rauwel, N. Castro-Diaz, H. M. Rowe, **G. Ecco**, C. Unzu, E. Planet, A. Lombardo, B. Mangeat, B. E. Wildhaber, L. Naldini, and D. Trono, 'Loss of Transcriptional Control over Endogenous Retroelements During Reprogramming to Pluripotency', *Genome Res,* 24 (2014), 1251-9.
5. C. Matiollo, **G. Ecco**, A. C. Menegatti, G. Razzera, J. Vernal, and H. Terenzi, 'S-Nitrosylation of Mycobacterium Tuberculosis Tyrosine Phosphatase a (Ptpa) Induces Its Structural Instability', *Biochim Biophys Acta,* 1834 (2013), 191-6.
6. L. D. Chiaradia, P. G. Martins, M. N. Cordeiro, R. V. Guido, **G. Ecco**, A. D. Andricopulo, R. A. Yunes, J. Vernal, R. J. Nunes, and H. Terenzi, 'Synthesis, Biological Evaluation, and Molecular Modeling of Chalcone Derivatives as Potent Inhibitors of Mycobacterium Tuberculosis Protein Tyrosine Phosphatases (Ptpa and Ptpb)', *J Med Chem,* 55 (2012), 390-402.
7. **G. Ecco**, J. Vernal, G. Razzera, P. A. Martins, C. Matiollo, and H. Terenzi, 'Mycobacterium Tuberculosis Tyrosine Phosphatase a (Ptpa) Activity Is Modulated by S-Nitrosylation', *Chem Commun (Camb),* 46 (2010), 7501-3.
8. P. V. Maillard, **G. Ecco**, M. Ortiz, and D. Trono, 'The Specificity of Trim5 Alpha-Mediated Restriction Is Influenced by Its Coiled-Coil Domain', *J Virol,* 84 (2010), 5790-801.
9. **G. Ecco**, J. Vernal, G. Razzera, C. Tavares, V. I. Serpa, S. Arias, F. K. Marchini, M. A. Krieger, S. Goldenberg, and H. Terenzi, 'Initial Characterization of a Recombinant Kynureninase from Trypanosoma Cruzi Identified from an Est Database', *Gene,* 448 (2009), 1-6.
10. C. Matiollo, J. Vernal, **G. Ecco**, J. B. Bertoldo, G. Razzera, E. M. de Souza, F. O. Pedrosa, and H. Terenzi, 'A Transthyretin-Related Protein Is Functionally Expressed in Herbaspirillum Seropedicae', *Biochem Biophys Res Commun,* 387 (2009), 712-6.

Skills

Gene and protein expression assessment (qPCR, RNA-seq, flow cytometry, Western Blot), genome engineering (cloning, CRISPR knock-out, siRNA knock-down, lentiviral vector production), screening system design and development, biochemistry techniques (protein purification, IP, ChIP-seq), cell culture (mouse and human ES and iPS cells), animal experimentation, bioinformatics analysis (R programming language). Project management experience and good communication skills.

Languages

English (fluent, TOEFL iBT score: 111), Portuguese (native), French (fluent), Italian (intermediate), Spanish (intermediate), German (intermediate).

Additional

- Supervision of a Bachelor's student's research summer project – EPFL, Switzerland (August, 2012)
- EBI-Wellcome Trust Summer School in Bioinformatics – England (June 2012)
- Member of the Bioscience Network Lausanne (BSNL) association of PhD students

References

1. Prof. Didier Trono, MD
EPFL-SV-GHI-LVG, SV 3512 (Bâtiment SV), Station 19, CH-1015 Lausanne, Switzerland
+41 21 69 31751, didier.trono@epfl.ch

2. Dr. Priscilla Turelli
EPFL-SV-GHI-LVG, AI 3151 (Bâtiment AI), Station 19, CH-1015 Lausanne, Switzerland
+41 21 69 31811, priscilla.turelli@epfl.ch

3. Dr. Helen Rowe
Department of Infection and Immunity, Cruciform Building (Floor 3), University College London (UCL), 90 Gower Street, London, WC1E 6BT
+44 20 76796022, h.rowe@ucl.ac.uk