

PRIMER

KRAB zinc finger proteins

Gabriela Ecco, Michael Imbeault and Didier Trono*

ABSTRACT

Krüppel-associated box domain zinc finger proteins (KRAB-ZFPs) are the largest family of transcriptional regulators in higher vertebrates. Characterized by an N-terminal KRAB domain and a C-terminal array of DNA-binding zinc fingers, they participate, together with their co-factor KAP1 (also known as TRIM28), in repression of sequences derived from transposable elements (TEs). Until recently, KRAB-ZFP/KAP1-mediated repression of TEs was thought to lead to irreversible silencing, and the evolutionary selection of KRAB-ZFPs was considered to be just the host component of an arms race against TEs. However, recent advances indicate that KRAB-ZFPs and their TE targets also partner up to establish species-specific regulatory networks. Here, we provide an overview of the KRAB-ZFP gene family, highlighting how its evolutionary history is linked to that of TEs, and how KRAB-ZFPs influence multiple aspects of development and physiology.

KEY WORDS: Endovirome, Transposable element, KRAB-ZFP, Co-option, Gene regulation, Speciation

Introduction

Biological events are regulated by complex transcriptional networks, with combinations of transcription factors interacting with *cis*-acting genomic sequences. Almost 70 years ago, Barbara McClintock proposed that some of these regulatory DNA sequences lay in mobile genetic elements (McClintock, 1950), and 20 years later Roy Britten and Eric Davidson outlined that the repetitive nature of these elements might explain how multiple changes in gene activity can so remarkably result from a single initiatory event (Britten and Davidson, 1969). Nevertheless, such transposable elements (TEs) kept being considered mostly as genetic threats in need of the strictest silencing, and were otherwise dismissed as purely selfish or junk DNA (Doolittle and Sapienza, 1980). However, the sequencing of the human genome at the dawn of the century changed this view, and it is increasingly recognized that some TEs are crucial components of transcriptional regulatory networks that play essential roles not only in the evolution but also the biology of most organisms (Garcia-Perez et al., 2016; Chuong et al., 2017; Thompson et al., 2016). Moreover, recent work indicates that a particular family of transcriptional regulators – Krüppel-associated box domain zinc finger proteins (KRAB-ZFPs) – controls TEs in higher vertebrates and, as such, exerts key influences on the biology of these organisms, including humans.

KRAB-ZFP genes first emerged more than 400 million years ago, and are now encoded in the hundreds by all modern tetrapods examined to date, with the notable exception of birds, in which they generally do not exceed ten (Emerson and Thomas, 2009; Liu et al.,

2014; Imbeault et al., 2017; Kauzlaric et al., 2017). KRAB-ZFPs are characterized by an N-terminal Krüppel-associated box (KRAB) domain and a C-terminal array of C₂H₂ zinc fingers (ZNFs) (Urrutia, 2003). Despite their abundance, the functions of KRAB-ZFPs have long remained ill-defined, although cumulated data have implicated some of them in processes as diverse as imprinting, cell differentiation, metabolic control and sexual dimorphism (reviewed by Lupo et al., 2013). This picture changed when the KRAB-binding co-factor KAP1 was demonstrated to be essential for the early embryonic repression of TEs in both mouse and human, and when a few individual KRAB-ZFPs could be linked to this function as well (Wolf and Goff, 2007, 2009; Wolf et al., 2015b; Rowe et al., 2010, 2013; Jacobs et al., 2014). It was then suspected that the primary role of KRAB-ZFPs was to silence TEs, and that their evolutionary selection represented the host component of an arms race against these genetic invaders (Jacobs et al., 2014; Castro-Diaz et al., 2014; Thomas and Schneider, 2011). More recent data, however, suggest that KRAB-ZFPs fulfill a role that is far more elaborate and, in some cases at least, can contribute to the domestication of their TE targets for the benefit of the host (Ecco et al., 2016; Imbeault et al., 2017).

In this Primer, we sum up our current understanding of the KRAB-ZFP family. We first provide an introduction to TEs and how they function. We also outline the structure, targets and general functions of KRAB-ZFPs. We then focus on the biological impact of the KRAB-ZFP gene family, highlighting its evolution, its role in controlling TEs, and how the selection of both TEs and KRAB-ZFPs might represent a dynamic partnership that generates the species-specific transcriptional networks that influence most aspects of human biology.

Transposable elements and their impact on the genome

TEs can be classified according to their transposition mechanism, overall genetic structure and phylogenetics. Most TEs present in the human genome are retroelements, whether endogenous retroviruses [e.g. human endogenous retrovirus (HERV) or long terminal repeat (LTR) retrotransposons] or non-LTR-retrotransposons of the long interspersed nuclear element (LINE), short interspersed nuclear element (SINE) and SINE-VNTR-Alu (SVA) subgroups. All retroelements spread via a copy-and-paste mechanism leading to their amplification. Given the functional and phylogenetic relationships between transposons and viruses, the sum of TEs present in the genome of an organism can be referred to as its ‘endovirome’, although it should be noted that not all TEs are *strictu sensu* derived from viruses. Some 4.5 million sequences derived from TEs can be readily identified in the human genome, accounting for about 50% of its DNA content. However, because TEs become unrecognizable over time owing to mutational drift, it is likely that this represents an underestimate of their contribution to our genetic make-up (de Koning et al., 2011; Hubley et al., 2016). Notably, as carriers of transcription factor-binding sites, TEs can impact the host genome in many ways (see Box 1). TEs thus fuel genetic diversity, but they can also induce deleterious mutations

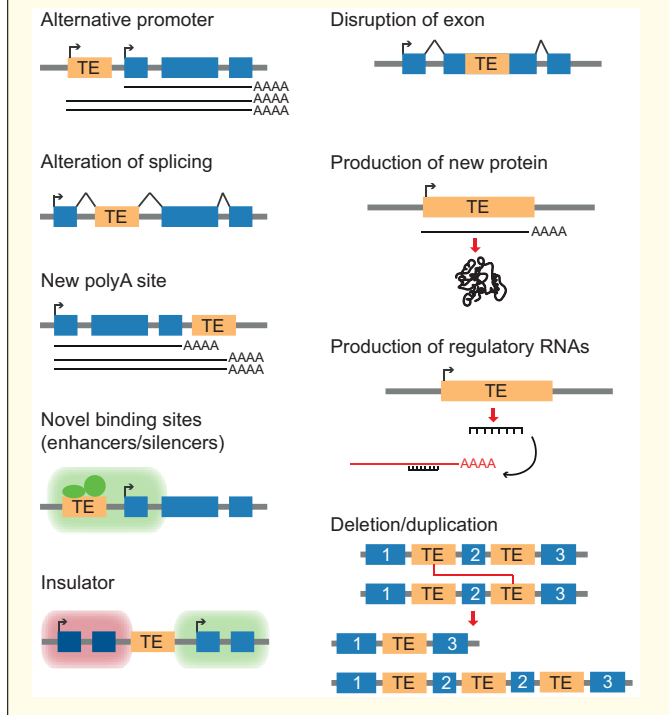
School of Life Sciences, École Polytechnique Fédérale de Lausanne (EPFL), Station 19, 1015 Lausanne, Switzerland.

*Author for correspondence (didier.trono@epfl.ch)

 D.T., 0000-0002-3383-0401

Box 1. How transposable elements can impact host genomes

Owing to their mobile nature and genetic constitution, TEs can perturb their genomic environment. They often bear promoters, enhancers, suppressors, insulators, splice sites or transcriptional stop signals. Accordingly, they can disrupt genes (via alternative splicing, truncation or insertion of new exons) or modify their expression (via promoter, enhancer or repressor effects). Owing to their highly repetitive nature, TEs also underlie recombination events that can lead to deletions, duplications, rearrangements or translocations. Finally, they can alter genome architecture via insulator sequences or by nucleating short- and long-range chromatin interactions, or they can provide entirely novel open reading frames (reviewed by Friedli and Trono, 2015; Rebollo et al., 2012b; Warren et al., 2015).



responsible for disease. Fewer than one out of 10,000 human TEs is still capable of transposition (Hancks and Kazazian, 2016), but a far greater proportion can alter gene expression.

Pathologies associated with new TE insertions or other types of deregulation include cancers, hemophilia, muscular dystrophy and other congenital or acquired human diseases (reviewed by Ayarpadikannan et al., 2015; Hancks and Kazazian, 2012, 2016; Mager and Stoye, 2015). Most TE-associated human disorders are related to non-LTR retrotransposons. For example, a known cause of breast cancer is the insertion of a primate-specific Alu SINE into the *BRCA1* and *BRCA2* genes (Miki et al., 1996; Puget et al., 1999). Cases of hemophilia A and B are also associated with insertional mutations of LINE-1 or Alu elements into genes that encode coagulation factors (Kazazian et al., 1988; Li et al., 2001). LTR retrotransposons have also been associated with some diseases, especially cancer. For instance, endogenous retrovirus (ERV) transcripts are upregulated in some tumors and there are reports of LTRs driving oncogene expression in human lymphomas (Lamprecht et al., 2010; Romanish et al., 2010; Babaian and Mager, 2016; Babaian et al., 2016). In mice, many LTR elements are transposition proficient, and ERVs related to mouse mammary tumor virus (MMTV) and mouse leukemia virus (MLV) can cause

cancer via activation of proto-oncogenes (Rosenberg and Jolicoeur, 1997). Finally, the expression of ERV proteins can be detrimental to the host and might be associated with autoimmune diseases such as systemic lupus erythematosus in mice and multiple sclerosis in humans (Baudino et al., 2010; Antony et al., 2011).

However, it is on an evolutionary scale that the impact of TEs is best appreciated. TEs endow the genomes of their host species with binding sites for transcription factors, which can then contribute to species-restricted phenotypes (reviewed by Thompson et al., 2016). For instance, mammals generally produce amylase in the pancreas, yet primates can release this enzyme in saliva too, owing to the insertion upstream of the amylase coding sequence of a HERV-E LTR driving expression in the salivary glands (Samuelson et al., 1996; Ting et al., 1992). Many other cases of LTR promoter exaptation have been documented, generally resulting in new or altered tissue-specific gene expression (Cohen et al., 2009; Stavenhagen and Robins, 1988; Rebollo et al., 2012a). Examples of TE-based species-specific enhancers also exist, and in mammals include MER130 elements acting as neocortex-specific units (Notwell et al., 2015), a SINE integrant functioning as a distal enhancer of *Fgf8* in the diencephalon (Nakanishi et al., 2012), RLTR13D5 ERVs co-opted as placenta-specific enhancers (Chuong et al., 2013), and the MER41-mediated dispersion of interferon-responsive elements in primates (Chuong et al., 2016). Retroelements can also contribute to embryonic stem cell (ESC) regulatory networks; many binding sites for pluripotency factors [such as Oct4 (Pou5f1) and Nanog] reside within primate- or human-specific ERVs in the human genome (Bourque et al., 2008). In addition, 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; Wang et al., 2014; Goke et al., 2015; Kunarso et al., 2010). TEs are also frequently bound by p53 (Trp53), with more than one-third of the genomic targets of this tumor suppressor overlapping with primate-specific ERVs (Wang et al., 2007), hence at locations not found for instance in the mouse genome. Finally, ERV-derived proteins can themselves be sources of genetic diversity, as illustrated in placental mammals in which formation of the syncytiotrophoblast, a placenta layer with extensive cellular fusion, is mediated by ERV envelope-derived syncytins (Mi et al., 2000; Dupressoir et al., 2009, 2011). Interestingly, across mammals, these proteins derive from the *env* gene of distinct groups of ERVs, indicating convergent evolution with multiple and independent events of ERV co-option (Lavialle et al., 2013). Together, these findings highlight the huge impact that TEs can have on the evolution and biology of complex organisms.

The domain structure of KRAB zinc finger proteins

KRAB-ZFPs are characterized by the presence of a KRAB domain and an array of C_2H_2 zinc fingers (Fig. 1). The KRAB domain encompasses approximately 75 amino acids and is often split into two modules: the A-box, which is primarily responsible for repressive activity, and the B-box, which is thought to potentiate KRAB-A effectiveness (Bellefroid et al., 1991; Mannini et al., 2006; Witzgall et al., 1994). The repressor activity of KRAB-ZFPs stems from the KRAB domain-mediated recruitment of KAP1 [KRAB-associated protein 1; also known as TRIM28 (tripartite motif protein 28), Tif1 β or KRIP-1] (Friedman et al., 1996), a scaffold protein that recruits mediators of heterochromatin formation (Iyengar and Farnham, 2011). The C-terminal C_2H_2 ZNF arrays of KRAB-ZFPs are tandem repeats of the $CX_2-4CX_{12}HX_{2-6}H$ motif (where X is any amino acid) interspaced by seven residue-long linkers (Iuchi, 2001). Human KRAB-ZFPs can

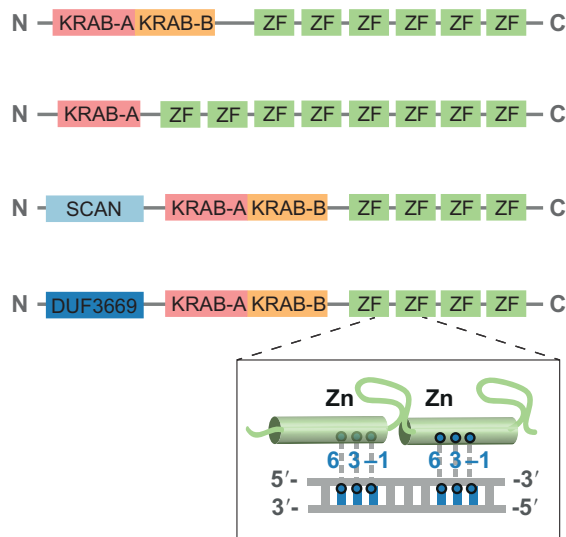


Fig. 1. The domain structure of KRAB-ZFPs. Four different examples of KRAB-ZFPs are depicted. All KRAB-ZFPs contain a KRAB domain and an array of zinc fingers with DNA-binding potential, the sequence specificity of which is dictated mainly by three amino acids within each zinc finger (at positions 6, 3 and -1). The KRAB domain can be split into two modules: the A-box (KRAB-A), which is primarily responsible for repressive activity, and the B-box (KRAB-B), which is thought to potentiate KRAB-A. Some KRAB-ZFPs also contain a SCAN or a DUF3669 domain. DUF3669, domain of unknown function 3669; KRAB, Krüppel-associated box; SCAN, SRE-ZBP, CTFin51, AW-1, and Number 18 cDNA; ZF, zinc finger.

harbor anywhere between two and more than 40 ZNFs, with the average number being 12 (Urrutia, 2003). Each zinc finger can theoretically interact with three nucleotides of the primary DNA strand (via amino acids at positions -1, 3 and 6 of the C_2H_2 helix), with some contacts being established with the secondary strand (via amino acid 2) (Emerson and Thomas, 2009; Elrod-Erickson et al., 1998). KRAB-ZFP genes display signs of strong positive selection at the positions encoding for the DNA-contacting amino acid residues, consistent with the idea that their products interact with DNA targets that themselves are capable of rapid evolution, such as TEs or viruses (Emerson and Thomas, 2009). Furthermore, although the length of KRAB-ZFPs (given the many ZNFs present in each array) should allow for a very high degree of specificity in the recognition of long DNA targets, it has been noted that KRAB-ZFP binding motifs are usually shorter than predicted. This suggests that different ZNFs in a KRAB-ZFP recognize different DNA motifs, as is the case for the C_2H_2 ZNF protein CTCF (Nakahashi et al., 2013), or that ZNFs not involved in contacting DNA could engage in other types of interactions, for instance with RNA or proteins (Najafabadi et al., 2015; Imbeault et al., 2017; Schmitges et al., 2016).

Some highly conserved KRAB-ZFPs contain additional elements in their N terminus, such as SCAN or DUF3669 domains. The vertebrate-specific SCAN domain can mediate oligomerization notably with other SCAN-containing proteins (Honer et al., 2001), whereas the function of the DUF3669 domain remains largely unknown, as indicated by its acronym (domain of unknown function).

Genomic targets of human KRAB zinc finger proteins

The genomic targets of a large fraction of human KRAB-ZFPs have been characterized in recent studies using chromatin immunoprecipitation followed by deep sequencing (ChIP-seq) and tagged proteins overexpressed in 293T cells as bait (Najafabadi et al., 2015; Imbeault et al., 2017; Schmitges et al., 2016). This type of

analysis does not allow one to conclude which genomic locus or loci are bound by a KRAB-ZFP in the physiological setting of a particular cell, as its recruitment stands to be differentially affected by the presence of other DNA-binding proteins, by the state of the chromatin and perhaps by levels of DNA methylation at potential target loci. However, these studies have revealed the type of genetic elements targeted by these KRAB-ZFPs, and in most cases have delineated a consensus binding sequence. These studies have also determined that a great majority of human KRAB-ZFPs associates with at least one subfamily of TEs, most of them retrotransposons. Some KRAB-ZFPs can bind to sequences in different TE families (e.g. HERVs and LINEs). Conversely, many TE subfamilies are recognized by several KRAB-ZFPs, which most often target clearly distinct regions of their integrants, as has been observed for ERVKs (endogenous retrovirus K) in mouse and for HERVs and LINE-1s (L1s) in humans (Imbeault et al., 2017; Ecco et al., 2016). Interestingly, in humans, it was observed that the age of the elements influences their pattern of KRAB-ZFP recruitment. For instance, the primate-specific LINE-1 (L1PA) L1PA4s, which are approximately 20 million years old (myo), are recognized by many KRAB-ZFPs. In contrast, most human-specific LINE-1s (L1Hs) are devoid of binding sites for factors recruited near the L1 promoter, such as ZNF93 (~20 myo), ZNF649 (~105 myo), ZNF765 (~7 myo) and ZNF141 (~43 myo). However, 3' binders such as ZNF382, ZNF84 (both ~105 myo) and ZNF429 (~29 myo) bind to a significant fraction of all L1PA integrants, from the ~40 myo L1PA16 to the youngest L1Hs (Imbeault et al., 2017). Thus, binding of KRAB-ZFPs to TEs is both combinatorial and evolutive.

It should be noted that about a third of tested human KRAB-ZFPs do not associate significantly with TEs and are instead found at other types of genomic targets such as promoters, simple repeats and poly-zinc finger protein genes. Many promoter-binding KRAB-ZFPs are ancient and evolutionarily conserved, and contain SCAN or DUF3669 domains; most do not recruit KAP1 and, as yet, are of unknown function (Imbeault et al., 2017; Schmitges et al., 2016). Some associate with wide arrays of promoters, for instance ZNF202, which binds in the vicinity of several thousand transcriptional start sites (TSSs). Others bind to promoters in a combinatorial fashion, such as the DUF3669-containing ZNF282 and ZNF398. Other KRAB-ZFPs, such as ZNF274 and ZNF75D, associate with the 3' region of poly-zinc finger protein genes, where they recognize conserved and partly overlapping motifs within the proximal part of ZNF-encoding sequences (Imbeault et al., 2017; Fritze et al., 2010).

The biological functions of KRAB zinc finger proteins

KRAB-ZFPs influence a variety of biological events. Many of their roles involve KAP1, which binds to a sizeable fraction of human and murine KRAB-ZFPs (Schmitges et al., 2016). KAP1 acts as a scaffold for a silencing complex that comprises the histone methyltransferase SETDB1 (also known as ESET) (Schultz et al., 2002), the nucleosome remodeling and deacetylation (NuRD) complex (Schultz et al., 2001), heterochromatin protein 1 (HP1) (Nielsen et al., 1999; Sripathy et al., 2006) and DNA methyltransferases (Quenneville et al., 2012) (Fig. 2). Accordingly, many KRAB-ZFPs act as transcriptional repressors via the KAP1-nucleated induction of heterochromatin and, in early embryonic cells, via DNA methylation (Wolf and Goff, 2009; Quenneville et al., 2012; Rowe et al., 2013; Jacobs et al., 2014; Najafabadi et al., 2015; Ecco et al., 2016; Schmitges et al., 2016; Imbeault et al., 2017). However, not all KRAB-ZFPs bind KAP1, and the interactome of more ancient human family members, notably those endowed with SCAN or

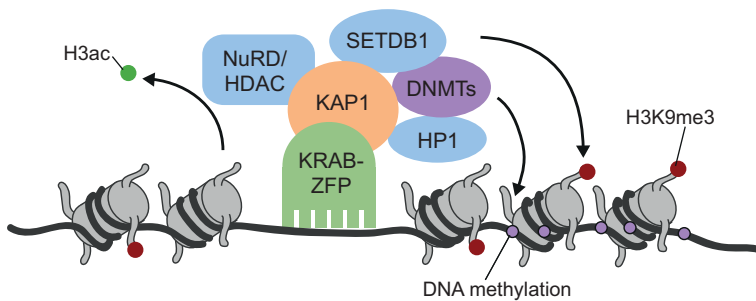


Fig. 2. The KRAB-ZFP/KAP1 repressor complex. KRAB-ZFPs (green) bind to DNA via their zinc fingers and recruit KAP1 (orange) via their KRAB domain. KAP1 then assembles a repressor complex, leading to heterochromatin formation, i.e. via histone methylation (H3K9me3), DNA methylation and histone deacetylation (H3ac), and transcriptional silencing. DNMT, DNA methyltransferase; H3ac, acetylated histone H3; H3K9me3, histone H3 trimethylated at Lys9; HDAC, histone deacetylase; HP1, heterochromatin protein 1; KAP1, Krüppel-associated box (KRAB)-associated protein 1; KRAB-ZFP, KRAB-zinc finger protein; NuRD, nucleosome remodeling deacetylase complex; SETDB1, SET domain bifurcated 1.

DUF3669 domains, reveals associations with other types of proteins, including transcriptional activators (Schmitges et al., 2016). Below, we provide an overview of the key biological functions that have been identified for KRAB-ZFPs during development.

Heterochromatin induction in early development and TE control

The best-characterized function of KRAB-ZFPs is the locus-specific induction of heterochromatin during early embryogenesis via the KRAB-mediated recruitment of KAP1, as first suggested by the discovery that KRAB could trigger promoter methylation if tethered to DNA during the first few days of mouse development (Wiznerowicz et al., 2007). At imprinting control regions, where a methylated hexanucleotide is recognized in mouse and human by ZFP57, this results in the trans-generational preservation of imprinting (Quenneville et al., 2011; Li et al., 2008; Strogantsev et al., 2015). At sequences derived from TEs, this allows for the taming of transcriptional influences that would otherwise hamper early development, from zygotic genome activation to the establishment and normal differentiation of pluripotent stem cells (Rowe et al., 2010; Matsui et al., 2010; Rowe et al., 2013; Turelli et al., 2014; Macfarlan et al., 2012). KRAB-ZFPs display exquisitely regulated patterns of expression during the first few days of embryogenesis, both in humans and mice, mirroring the tightly orchestrated transcription of TE-containing loci during this period (Corsinotti et al., 2013; Theunissen et al., 2016; Macfarlan et al., 2012; Fort et al., 2014; Gifford et al., 2013; Goke et al., 2015; Grow et al., 2015; Kunarso et al., 2010; Xue et al., 2013; Yan et al., 2013). The removal of KAP1 or its partner histone methyltransferase SETDB1 in murine or human ESCs activates the expression of multiple TEs (Matsui et al., 2010; Rowe et al., 2010; Turelli et al., 2014).

A number of KRAB-ZFPs have been implicated in controlling TE repression in ESCs (Wolf et al., 2015a). ZFP809, a murine-specific KRAB-ZFP, was demonstrated early on to silence exogenous MLV in embryonic carcinoma cells through recognition of the provirus primer binding site-coding sequence (Wolf and Goff, 2007, 2009). Curiously, depletion of ZFP809 in mice leads to de-repression of MLV-related ERVs in adult tissues but not in ESCs (Wolf et al., 2015b). Although functional data on the role of individual human KRAB-ZFPs during this period are still missing, it is noteworthy that HERVH (human endogenous retrovirus H) integrants, which appear to play an important role in human ESC pluripotency, are recognized by several KRAB-ZFPs, the levels of which change as these cells switch from a naïve to a primed pluripotent state (Theunissen et al., 2016). Other KRAB-ZFPs controlling TEs in ESCs are ZNF91 and ZNF93, which respectively repress SVAs and LINE-1 (Jacobs et al., 2014), and the murine paralogs ZFP932 and Gm15446, which regulate ERVKs (Ecco et al., 2016). It is now established that, by controlling TEs, the KRAB-ZFP/KAP1 complex ensures the transcriptional homeostasis and normal differentiation of ESCs. Upon KAP1 or KRAB-ZFP depletion in ESCs, repressive chromatin

marks at TEs are replaced by active histone modifications typically found on enhancers, and nearby genes can become activated (Rowe et al., 2013; Jacobs et al., 2014; Turelli et al., 2014; Ecco et al., 2016).

Until recently, it was generally believed that most TEs are irreversibly silenced during these early stages of embryonic development, alleviating the need for subsequent sequence-specific control, including by the KRAB-ZFP/KAP1 system (Maksakova et al., 2008; Walsh et al., 1998). However, recent evidence suggests otherwise. First, deep transcriptome analyses indicate that some TE loci can be transcriptionally active in adult tissues, providing alternative promoters or fulfilling other regulatory functions (Faulkner et al., 2009; Belancio et al., 2010). Second, in mature T lymphocytes, a significant fraction of TEs bound by KAP1 in human ESCs still carries the co-repressor (Turelli et al., 2014). Third, KAP1 deletion in neuronal progenitors activates some endogenous retroelements (Fasching et al., 2015), and selected ERVs are similarly induced in murine B lymphocytes or mouse embryonic fibroblasts (MEFs) depleted for SETDB1 (Collins et al., 2015; Wolf et al., 2015b). Correspondingly, human KRAB-ZFPs display extensive and cell-specific patterns of expression in all adult tissues examined (Imbeault et al., 2017; Liu et al., 2014). Furthermore, the mouse-specific KRAB-ZFPs ZFP932 and Gm15446 are also involved in controlling their TE targets in somatic tissues, where they modulate the TE-mediated regulation of neighboring genes *in vivo* (Ecco et al., 2016). More broadly, by comparing KRAB-ZFP-binding sites with the ENCODE database, a significant overlap between the TE targets of a number of human KRAB-ZFPs and the binding regions of other transcription factors such as YY1, CEBPZ, GATA3, FOXA1 and STAT1 was observed (Imbeault et al., 2017). Finally, by examining the chromatin state of KRAB-ZFP-bound TEs in a subset of these tissues, it was noted that a significant fraction display cell-specific enrichment of activation marks instead of those associated with repressive heterochromatin. Moreover, in these cases, nearby genes were on average expressed at higher levels, consistent with KRAB-ZFP-controlled, TE-based enhancer effects on these genes (Imbeault et al., 2017). Considering the limited scope of this type of analysis, which can detect neither long-range effects nor *trans*-acting influences by TE-derived regulatory RNAs, and the fact that chromatin data were available only for a few cell types, it is likely that the KRAB-ZFP-mediated control of TEs in fact impacts the physiology of a range of developing and adult tissues.

KRAB-ZFPs in cell differentiation

As discussed above, many KRAB-ZFPs are expressed in ESCs and early progenitors (Corsinotti et al., 2013), where they engage together with KAP1 in repressing TEs. However a number of KRAB-ZFPs can influence other aspects of development, although no evidence for interaction with TEs has been demonstrated so far in these cases. In the mouse, for example, ZFP689, ZFP13 and KAP1

play an important role in erythropoiesis by regulating an miRNA cascade that governs mitophagy in red cell precursors (Barde et al., 2013). In humans, there is evidence that ZNF589, ZNF268 and ZNF300 influence hematopoietic differentiation (Venturini et al., 2016; Zeng et al., 2012; Xu et al., 2010). KAP1 is also important for B- and T-cell development and homeostasis (Santoni de Sio, 2014). Studies in human and mouse showed that KAP1 depletion in these cells leads to differentiation and metabolic defects (Santoni de Sio et al., 2012a,b; Chikuma et al., 2012). The KRAB-ZFPs responsible for these phenotypes, however, have not yet been identified, but many are specifically expressed in these tissues (Liu et al., 2014; Imbeault et al., 2017). Other events influenced by KRAB-ZFPs include osteogenesis (Jheon et al., 2001), mammary gland development (Oliver et al., 2012) and the formation of extra-embryonic tissues (Shibata and Garcia-Garcia, 2011; Shibata et al., 2011).

KRAB-ZFPs and metabolism

A number of KRAB-ZFPs have been implicated in cellular and organismal metabolic pathways. For example, ZFP69 was reported to mediate liver fat accumulation and mild insulin resistance in mice (Chung et al., 2015). In human cells, ZNF224 is associated with glycolysis and oxidative metabolism (Iacobazzi et al., 2009; Lupo et al., 2011). The mechanism of action of these KRAB-ZFPs are not all defined, but they most likely act via KAP1 as ZNF224, for instance, was shown to interact with the co-repressor (Medugno et al., 2005). Furthermore, KAP1 plays important roles in the liver: liver-specific KAP1 knockout leads to male-restricted hepatic carcinogenesis and perturbs the metabolism of hormones and antibiotics in the liver (Bojkowska et al., 2012). In mice, the KRAB-ZFPs RSL1 and RSL2 are involved in sexually dimorphic gene expression, also in the liver, repressing male-specific hepatic genes such as members of the cytochrome P450 (Cyp) families, which are important for the metabolism of xenobiotics (Krebs et al., 2003). These dimorphic cytochrome P450 genes are also upregulated in KAP1 knockout livers (Bojkowska et al., 2012), suggesting that RSL1 and RSL2 act via KAP1 in this context. Interestingly, it has been reported that the control of one RSL1 target, the gene encoding the sex-limited protein (SLP; also known as C4A), seems to occur via binding to an ancient endogenous retrovirus (Stavenhagen and Robins, 1988; Krebs et al., 2012).

Other examples of KRAB-ZFPs implicated in metabolism include ZNF255, an isoform of ZNF224, which interacts with a Wilms' tumor 1 (WT1) protein isoform that has affinity for RNA and has been implicated in transcript processing, suggesting a role for this KRAB-ZFP in RNA maturation and post-transcriptional control (Florio et al., 2010). Similarly, ZNF74 binds RNA and is tightly associated with the nuclear matrix, suggesting a role for this protein in RNA metabolism (Grondin et al., 1996).

The evolutionary path of KRAB-ZFPs

A survey of more than 200 vertebrate genomes reveals that KRAB-ZFP genes first appeared some 420 million years ago in a common ancestor of coelacanths, lungfish and tetrapods (Imbeault et al., 2017). The genomes of all analyzed modern species derived from this ancestor, except for birds, contain several hundreds of KRAB-ZFP genes. Interestingly, all 300 or so KRAB-ZFP genes found in coelacanths seem to be mono-exonic, whereas in all other species the KRAB and zinc finger domains are most often encoded by separate exons. This suggests that ancestral KRAB-ZFPs were mono-exonic, and that switching to a multi-exonic configuration perhaps facilitated the reshuffling of zinc finger arrays and the

independent evolution of the KRAB domain, paving the way to its coupling in some proteins to SCAN or DUF3669 domains, and to the emergence of non-canonical KRAB units not functionally linked to KAP1 recruitment (Schmitges et al., 2016; Itokawa et al., 2009; Murphy et al., 2016).

To trace putative DNA-binding orthologs, the ZNF fingerprints of KRAB-ZFPs (i.e. the series of amino acid triplets within their ZNF arrays predicted to dictate their DNA-binding specificity) have been compared (Liu et al., 2014; Imbeault et al., 2017). These analyses delineated clusters that are specific for most taxonomic orders and also allowed for the identification of KRAB-ZFPs restricted to each species (Imbeault et al., 2017). Interestingly, no ZNF fingerprint ortholog of coelacanth KRAB-ZFPs is found in any other species, suggesting that the genomic targets of these proteins are unique to this organism and possibly untested close relatives, consistent with the existence of species-restricted TEs. Many species- and class-specific KRAB-ZFPs can similarly be detected in most analyzed genomes, indicating ongoing amplification and turnover of the family with regular addition of new members (Imbeault et al., 2017; Huntley et al., 2006; Thomas and Schneider, 2011). A recent examination of the mouse genome identified about twice as many KRAB-ZFP genes as had been previously annotated as either KRAB-ZFP genes or pseudogenes, notably by assigning an entity formerly considered as a large group of satellite repeats to this family (Kauzlaric et al., 2017). It also highlighted the cluster-based organization of these genes and their distribution throughout the genome, with signs of recombination, translocation, duplication and seeding of new sites by retrotransposition of KRAB-ZFP genes. Finally, it provided evidence that closely related paralogs have evolved through both the genetic drifting and shifting of sequences encoding for zinc finger arrays; that is, with adjacent KRAB-ZFPs differing by either point mutations at DNA-contacting residues of a few ZNFs or substitutions of entire blocks of these motifs (Kauzlaric et al., 2017).

It has been noted that the invasion by new families of endogenous retroviruses coincided with the appearance of novel KRAB-ZFP duplicates in primates (Thomas and Schneider, 2011). The guinea pig, opossum and, to a lesser extent, mouse genomes display an unusually high number of species-specific paralogs. The mouse genome is known to harbor a significant fraction of retrotransposition-competent TEs, including ERVs and LINEs (Kazazian, 2004; DeBerardinis et al., 1998), supporting a model whereby new TE variants contribute to fix recently emerged KRAB-ZFP paralogs. Conversely, a few KRAB-ZFPs that are highly conserved in other mammals have been lost in primates, whereas some human KRAB-ZFP pseudogenes have functional orthologs in closely related species, indicating divergence in the selective pressures responsible for their maintenance (Imbeault et al., 2017). Emergence of a paralog acting as a functional substitute, rather than extinction of the corresponding TE targets, probably accounts for most of these occurrences.

It is remarkable that all examined bird genomes stand out for their very low content of KRAB-ZFP genes, no more than ten in most of them. Interestingly, avian genomes are significantly smaller than those of other amniotes (Organ et al., 2007; Wallis et al., 2004), and a much smaller fraction of the chicken and zebrafish genomes can be readily attributed to TEs, compared with most other tetrapods (15% versus 40-50% on average) (Chalopin et al., 2015). This suggests that TE burden and activity contribute to the maintenance of a pool of functional KRAB-ZFPs. Alternatively, it is tempting to hypothesize that birds, when they emerged from theropod dinosaurs, evolved another TE control system that exhibits many of

the same functional properties as KRAB-ZFPs, rendering the latter dispensable.

TE/KRAB-ZFP co-evolution: both an arms race and domestication

A wealth of data indicates that KRAB-ZFPs and TEs have co-evolved. This has led to proposition of the ‘arms race’ model, which states that competition between KRAB-ZFPs and TEs, with KRAB-ZFPs continuously trying to suppress invasion by rapidly mutating TEs, drives their selection. However, more recent data suggest that this model is too simplistic, and additionally point towards a ‘domestication’ model in which at least some KRAB-ZFPs help the host co-opt TEs for its benefit.

The arms race model

Several lines of evidence indicate that TEs have served as an important motor for the selection of KRAB-ZFP genes. During evolution, KRAB-ZFP genes underwent strong positive selection at positions encoding amino acids predicted to determine the DNA-binding specificity of their products (Emerson and Thomas, 2009; Liu et al., 2014). Furthermore, KRAB-ZFP paralogs exhibit not only significant differences in ZNF fingerprints, but also differential expression and splicing patterns across tissues, consistent with the acquisition of new functions following gene duplication events (Nowick et al., 2010; Kauzlaric et al., 2017). An analysis of data from the 1000 Genomes Project revealed that human KRAB-ZFP genes harboring non-synonymous single nucleotide polymorphisms in sequences encoding their predictive DNA-contacting residues are generally expressed at lower levels, are evolutionarily younger, and seem to be less evolutionarily constrained than those without such polymorphisms, suggesting that they are on their way to becoming pseudogenes (Kapopoulou et al., 2016).

Most importantly, both KRAB-ZFPs and TEs underwent parallel waves of expansion in the genomes of tetrapods (Thomas and Schneider, 2011). Moreover, in human ESCs, a dynamic regulation model of LINE elements by KRAB-ZFP/KAP1 can be documented, whereby the expression of newly emerged LINE-1 families is initially repressed by small-RNA-induced DNA methylation, before KAP1-mediated repression takes over through the selection of KRAB-ZFPs sequentially capable of recognizing these TEs, until these are ultimately deprived of any activity by mutations (Castro-Diaz et al., 2014).

Together, these findings have led to the ‘arms race’ model (Fig. 3A), which asserts that dynamic competition between TEs and KRAB-ZFPs drives their co-evolution, with TEs that are controlled by a KRAB-ZFP mutating away to escape repression while the pool of KRAB-ZFP genes evolves proteins with novel zinc finger arrays, which get fixed once they can recognize the renegade TE (Imbeault and Trono, 2014). This model is best exemplified by the primate-specific L1PA subfamily of LINE elements, as these TEs, devoid of an extracellular phase, display a linear evolutionary path, each new subfamily deriving from the one previously expanded in the genome of its host species and ancestors. Indeed, compelling evidence for the arms race model stems from the characterization of ZNF93 and its binding to L1PA elements – in particular the loss (via deletion) of the ZNF93 recognition site in newer L1PA subfamilies (Jacobs et al., 2014). Additional support comes from the recent identification (Imbeault et al., 2017) of TE targets in a large set of human KRAB-ZFPs, which reveals the sequential recruitment at the 5' ends of primate-specific L1 elements of not only ZNF93 but also ZNF141, ZNF649 and ZNF765, with zinc finger mutations accumulating coincidentally with the appearance of new L1PA

subfamilies, and loss of binding sites for all of these KRAB-ZFPs in the newest human-specific LINE-1. This study could also retrace specific mutation events in the binding motifs of KRAB-ZFPs that correlated with loss of binding in the youngest elements, generally subtler than the 129-bp deletion event that led to escape from ZNF93 (Imbeault et al., 2017).

With ERVs, the situation is more complicated, as these TEs are endogenized following waves of genomic invasion originating from external sources, with potential iterations precluding firm dating. Nonetheless, in mice, the KRAB-ZFP paralogs ZFP932 and Gm15446 regulate overlapping but distinct sets of ERVKs with both proteins binding to the 3' end of members from the same families of retroelements, but with different preferences (Ecco et al., 2016). For instance, whereas ZFP932 and Gm15446 are similarly enriched at RLTR44-int, IAP-d-int and MMERVK10D3_I-int elements, Gm15446 is more frequently found at MMERVK10C-int, IAPEY-int and IAPEY3-int. Further analyses suggest that ZFP932 appeared first and that Gm15446 arose secondarily by duplication, with subsequent accumulation of mutations leading to a partial shift in target range (Kauzlaric et al., 2017).

The domestication model

More recently, evidence has emerged to suggest that a host-invader arms race cannot have been the sole motor of the evolutionary selection of KRAB-ZFP genes. First, the recognition of many TEs by several KRAB-ZFPs would constitute a major obstacle to mutational escape if these factors were all simultaneously engaged in their repression. Second, LINE-1 integrants controlled by KAP1 in human ESCs are between ~7 and 25 million years of age, and have long lost all transposition potential (Castro-Diaz et al., 2014), as have all HERVs, including the tens of thousands of integrants still controlled by KAP1; therefore, the conservation of KRAB-ZFP-binding sites in these elements does not arise from the need to suppress their replication. Third, it appears that numerous TEs kept spreading or even started invading the human ancestral genome long after KRAB-ZFPs capable of recognizing their sequence had emerged. For instance ZNF649, which like ZNF93 binds the L1PA promoter and exhibits a very similar expression pattern, dates back to the time of mammalian radiation, some 60 million years before either ZNF93 or any of its target L1PA subfamilies appeared. In addition, recent data suggest that enrichment for certain KRAB-ZFPs is positively selected on some TEs, as for ZNF382 and ZNF84 on L1Hs, most integrants of which are recognized by these proteins (Imbeault et al., 2017). It could be hypothesized that these TEs have evolved to bind KRAB-ZFPs in order to be able to spread in the germ line, where these proteins might not be produced, and be subsequently controlled by their action in differentiated tissues, which would minimize negative selection. However, we note that most of the corresponding KRAB-ZFPs do not recruit KAP1 (P.-Y. Helleboid and D.T., unpublished), and are therefore not predicted to act as repressors. In addition, most KRAB-ZFPs exhibit highly sophisticated patterns of expression, exhibiting tissue- and lineage-specificity and being influenced by the differentiation and activation states of the cell, indicating that their interactions with, and hence their influences on, their TE targets are highly regulated. Collectively, these findings strongly suggest a ‘domestication model’, in which KRAB-ZFPs, rather than just blocking the transposition potential of TEs, participate in their domestication (Fig. 3B). It is noteworthy that data revealing highly tissue-specific expression patterns, individualized sets of post-translational modifications and very distinct protein interactomes for many human KRAB-ZFPs (P.-Y. Helleboid and D.T., unpublished)

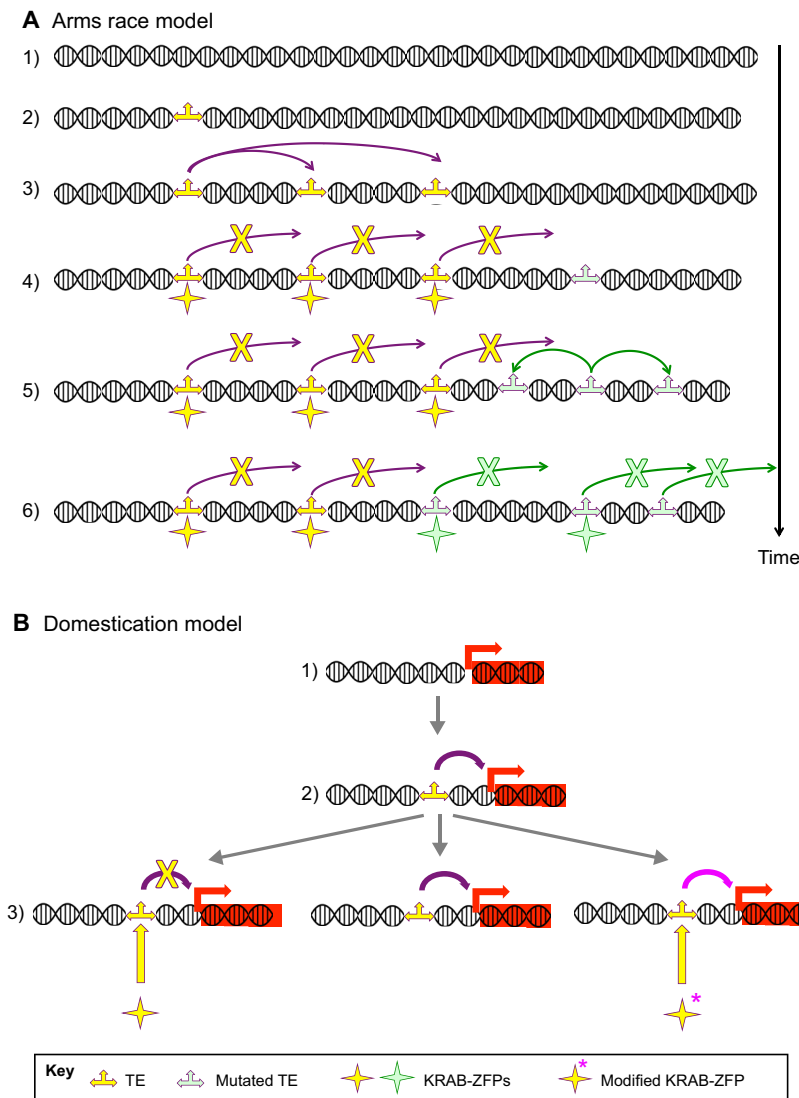


Fig. 3. Arms race and domestication. Two non-mutually exclusive models are depicted to describe interactions between TEs and KRAB-ZFPs. (A) The arms race model. A genomic locus (1) is invaded by a new TE (2), which initially spreads (3) by partly escaping control via RNA-based mechanisms and other restriction factors such as APOBEC proteins. A KRAB-ZFP gene emerges (4), the product of which (yellow star) represses this TE, preventing its further propagation. A new wave of integration is then launched by a mutant TE, for instance one generated owing to the high error rate of reverse transcriptase, that is not bound by this KRAB-ZFP and hence is immune to its control (5), until another KRAB-ZFP (green star) capable of repressing the activity of this mutant TE is produced (6). Of note, long-term KRAB-ZFP-mediated control might be necessary on some TE integrants, even after mutations have destroyed their transposition-disrupting potential, in order to keep silencing their transcription-disrupting potential. However, it is expected that, over time, mutations will also neutralize this type of TE activity, thus rendering KRAB-ZFP-mediated control unnecessary, unless it is of benefit to the host. (B) The domestication model. A gene-containing locus (1) is invaded by a TE (2), which results in modulation of its expression (for instance by acting as an enhancer). A KRAB-ZFP recognizing this TE emerges (3) and, depending on the cell lineage, differentiation stage or activation state, can either silence the TE (left) by inducing the local formation of heterochromatin, modify its regulatory activity (right), for instance if upon post-translational modification it recruits not just a repressor complex but other types of transcriptional or post-transcriptional modulators, or leave it intact (center), for instance in a cell in which this KRAB-ZFP is not expressed (thus allowing the TE-based sequence to act as a tissue-specific enhancer).

indicate that this system is likely to be far more amenable to regulation than RNA-based TE control mechanisms, which are predominantly at work in the germ line and during early embryogenesis and mostly result in permanent silencing. As such, KRAB-ZFPs might be important instruments towards a full participation of TEs in shaping transcriptional regulatory networks as imagined by Britten and Davidson some fifty years ago (Britten and Davidson, 1969).

The evolutionarily ephemeral nature of many KRAB-ZFP genes should not be taken as an argument against their involvement in the domestication of TEs. Indeed, TEs sprinkle genomes with a constant flux of *cis*-acting sequences. Although many TE integrants are likely to be of neutral impact, destined to be progressively erased by mutational drift or to be eliminated by recombination, some can exert significant influences, which provide the host with new options for regulating biological events. At times, these new integrants could render more ancient TE-based *cis*-acting sequences less essential, opening the door to their evolutionary removal. Meanwhile, the genomes of higher vertebrates generate steady supplies of new KRAB-ZFP genes. Some are paralogs subtly differing from their immediate predecessors, which they can functionally replace, leading to their evolutionary loss. Others are

more novel family members, which can either be positively selected if they fulfill a useful role, or disappear if they do not match a functionally relevant target.

A general picture integrating these various considerations thus emerges and can describe the interactions between TEs and KRAB-ZFPs. When a new TE enters a host genome, whether from an exogenous source (for ERVs) or by mutation of an endogenous predecessor, it is initially silenced via ancestral RNA-based mechanisms, such as those mediated by Piwi-interacting RNAs (piRNAs). Over time, its integrants accumulate mutations that progressively hamper their transposition potential. Meanwhile, KRAB-ZFP paralogs with novel DNA-binding specificities are generated, some of which recognize these TEs and get fixed, because they contribute to preventing the further spread of these elements and/or because they partake in their co-option for the benefit of the host, for instance by allowing the transcriptional regulatory potential of these TEs to be developmentally regulated or tissue restricted. Based on the observed evolutionary dynamics of KAP1-mediated control of LINE-1 in human ESCs, it seems that, at least in recent time and for this class of retroelements, the matching of a newly appeared TE and an inhibitory KRAB-ZFP can take more than 7 million years, as KAP1 does not repress any human-specific

LINEs (Castro-Diaz et al., 2014). Over time, KRAB-ZFP/KAP1-controlled TE integrants continue to undergo mutational drift, so that in some cases only their KRAB-ZFP-recruiting region remains to serve as a transcription regulatory platform, which could explain why we frequently find the oldest human KRAB-ZFPs at promoters, without identifiable TE signatures. KRAB-ZFPs themselves might evolve to become capable of recruiting activities distinct from KAP1-nucleated repression. Ultimately, all that might be left from the TE/repressor pair is a DNA target motif and its sequence-specific polypeptidic ligand, with no recognizable trace of their source elements.

Concluding remarks

TEs and their KRAB-ZFP controllers confer a high degree of species specificity to many biological processes relevant to the development and physiology of their hosts, including humans. Indeed, a large fraction of the human endovirome is unique to our species and its close relatives, both with regards to its sequence and the genomic distribution of its individual components. Correspondingly, many human KRAB-ZFPs are relatively recent products of our evolution (Nowick et al., 2010; Liu et al., 2014; Imbeault et al., 2017). Therefore, when studying regulatory networks in the human system or in animal models one should carefully discern general principles from species-specific layers of regulation.

Considering that species-restricted KRAB-ZFPs and TEs probably shape regulatory networks in all mammals, the high degree of similarity in the physiology of these organisms might seem surprising. However, although the dynamic partnership between TEs and their KRAB-ZFP ligands provides plenty of ground for divergence, functional evolution for most organ systems is limited by physiological and environmental constraints. For instance, even though early embryogenesis is regulated by different sets of TEs and KRAB-ZFPs in mice and humans, major deviations are difficult to introduce in this highly orchestrated process.

One organ that might partly escape such evolutionary canalization is the central nervous system as, at least in humans, a very wide range of cognitive and psychological phenotypes are compatible with normal life expectancy and efficient reproduction. It is thus interesting to note that remarkably elevated levels of TE activity have been recorded in the brain (Erwin et al., 2014), that a higher range of KRAB-ZFPs is expressed in the brain than in most other adult human tissues (Imbeault et al., 2017), and that KRAB-ZFPs disproportionately contribute to differences between the brain gene networks of chimpanzees and humans (Nowick et al., 2009). These observations suggest that the endovirome and its KRAB-ZFP controllers could have played an important role in the expansion of higher brain functions that were key to the emergence of modern humans. Future studies should test this hypothesis and further decipher the function and mechanisms of action of individual KRAB-ZFPs, and how, together with their TE or non-TE targets, these proteins so uniquely impact on the biology of their host species.

Acknowledgements

We thank all former and current members of the Trono Lab for discussions. We regret being unable to cite all work relevant to this primer owing to space constraints.

Competing interests

The authors declare no competing or financial interests.

Funding

The authors were funded by the Swiss National Foundation (Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung) and the European Research Council (ERC 268721 and ERC 694658).

References

- 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? *Biochim. Biophys. Acta* **1812**, 162-176.
- 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.
- Babaian, A. and Mager, D. L. (2016). Endogenous retroviral promoter exaptation in human cancer. *Mob. DNA* **7**, 24.
- Babaian, A., Romanish, M. T., Gagnier, L., Kuo, L. Y., Karimi, M. M., Steidl, C. and Mager, D. L. (2016). Onco-exaptation of an endogenous retroviral LTR drives IRF5 expression in Hodgkin lymphoma. *Oncogene* **35**, 2542-2546.
- Barde, I., Rauwel, B., Marin-Florez, R. M., Corsinotti, A., Laurenti, E., Verp, S., Offner, S., Marquis, J., Kapopoulou, A., Vanicek, J. et al. (2013). A KRAB/KAP1-miRNA cascade regulates erythropoiesis through stage-specific control of mitophagy. *Science* **340**, 350-353.
- Baudino, L., Yoshinobu, K., Morito, N., Santiago-Raber, M.-L. and Izui, S. (2010). Role of endogenous retroviruses in murine SLE. *Autoimmun. Rev.* **10**, 27-34.
- Belancio, V. P., Roy-Engel, A. M., Pochampally, R. R. and Deininger, P. (2010). Somatic expression of LINE-1 elements in human tissues. *Nucleic Acids Res.* **38**, 3909-3922.
- Bellefroid, E. J., Poncelet, D. A., Lecocq, P. J., Revelant, O. and Martial, J. A. (1991). The evolutionarily conserved Kruppel-associated box domain defines a subfamily of eukaryotic multifingered proteins. *Proc. Natl. Acad. Sci. USA* **88**, 3608-3612.
- 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* **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 Res.* **18**, 1752-1762.
- Britten, R. J. and Davidson, E. H. (1969). Gene regulation for higher cells: a theory. *Science* **165**, 349-357.
- 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 Dev.* **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 Biol. Evol.* **7**, 567-580.
- Chikuma, S., Suita, N., Okazaki, I.-M., Shibayama, S. and Honjo, T. (2012). TRIM28 prevents autoinflammatory T cell development in vivo. *Nat. Immunol.* **13**, 596-603.
- Chung, B., Stadion, M., Schulz, N., Jain, D., Scherneck, S., Joost, H.-G. and Schurmann, A. (2015). The diabetes gene Zfp69 modulates hepatic insulin sensitivity in mice. *Diabetologia* **58**, 2403-2413.
- Chuong, E. B., Rumi, M. A. K., Soares, M. J. and Baker, J. C. (2013). Endogenous retroviruses function as species-specific enhancer elements in the placenta. *Nat. Genet.* **45**, 325-329.
- Chuong, E. B., Elde, N. C. and Feschotte, C. (2016). Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science* **351**, 1083-1087.
- Chuong, E. B., Elde, N. C. and Feschotte, C. (2017). Regulatory activities of transposable elements: from conflicts to benefits. *Nat. Rev. Genet.* **18**, 71-86.
- 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. *Proc. Natl. Acad. Sci. USA* **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.
- De Koning, A. P. J., Gu, W., Castoe, T. A., Batzer, M. A. and Pollock, D. D. (2011). Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet.* **7**, e1002384.
- Deberardinis, R. J., Goodier, J. L., Ostertag, E. M. and Kazanian, H. H., Jr. (1998). Rapid amplification of a retrotransposon subfamily is evolving the mouse genome. *Nat. Genet.* **20**, 288-290.
- Doolittle, W. F. and Sapienza, C. (1980). Selfish genes, the phenotype paradigm and genome evolution. *Nature* **284**, 601-603.
- 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 placentalization of a fusogenic, endogenous retrovirus-derived, envelope gene. *Proc. Natl. Acad. Sci. USA* **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. *Proc. Natl. Acad. Sci. USA* **108**, E1164-E1173.
- Ecco, G., Cassano, M., Kauzlaric, A., Duc, J., Coluccio, A., Offner, S., Imbeault, M., Rowe, H. M., Turelli, P. and Trono, D.** (2016). Transposable elements and their KRAB-ZFP controllers regulate gene expression in adult tissues. *Dev. Cell* **36**, 611-623.
- Elrod-Erickson, M., Benson, T. E. and Pabo, C. O.** (1998). High-resolution structures of variant Zif268-DNA complexes: implications for understanding zinc finger-DNA recognition. *Structure* **6**, 451-464.
- Emerson, R. O. and Thomas, J. H.** (2009). Adaptive evolution in zinc finger transcription factors. *PLoS Genet.* **5**, e1000325.
- Erwin, J. A., Marchetto, M. C. and Gage, F. H.** (2014). Mobile DNA elements in the generation of diversity and complexity in the brain. *Nat. Rev. Neurosci.* **15**, 497-506.
- 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 Rep* **10**, 20-28.
- Faulkner, G. J., Kimura, Y., Daub, C. O., Wani, S., Plessy, C., Irvine, K. M., Schroder, K., Cloonan, N., Steptoe, A. L., Lassmann, T. et al.** (2009). The regulated retrotransposon transcriptome of mammalian cells. *Nat. Genet.* **41**, 563-571.
- Florio, F., Cesaro, E., Montano, G., Izzo, P., Miles, C. and Costanzo, P.** (2010). Biochemical and functional interaction between ZNF224 and ZNF255, two members of the Kruppel-like zinc-finger protein family and WT1 protein isoforms. *Hum. Mol. Genet.* **19**, 3544-3556.
- 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. *Nat. Genet.* **46**, 558-566.
- Friedli, M. and Trono, D.** (2015). The developmental control of transposable elements and the evolution of higher species. *Annu. Rev. Cell. Dev. Biol.* **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. III.** (1996). KAP-1, a novel corepressor for the highly conserved KRAB repression domain. *Genes Dev.* **10**, 2067-2078.
- Frietze, S., Lan, X., Jin, V. X. and Farnham, P. J.** (2010). Genomic targets of the KRAB and SCAN domain-containing zinc finger protein 263. *J. Biol. Chem.* **285**, 1393-1403.
- Garcia-Perez, J. L., Widmann, T. J. and Adams, I. R.** (2016). The impact of transposable elements on mammalian development. *Development* **143**, 4101-4114.
- Gifford, W. D., Pfaff, S. L. and Macfarlan, T. S.** (2013). Transposable elements as genetic regulatory substrates in early development. *Trends Cell Biol.* **23**, 218-226.
- Goke, J., Lu, X., Chan, Y.-S., Ng, H.-H., Ly, L.-H., Sachs, F. and Szczerbinska, I.** (2015). Dynamic transcription of distinct classes of endogenous retroviral elements marks specific populations of early human embryonic cells. *Cell Stem Cell* **16**, 135-141.
- Grondin, B., Bazinet, M. and Aubry, M.** (1996). The KRAB zinc finger gene ZNF74 encodes an RNA-binding protein tightly associated with the nuclear matrix. *J. Biol. Chem.* **271**, 15458-15467.
- Grow, E. J., Flynn, R. A., Chavez, S. L., Bayless, N. L., Wossidlo, M., Wesche, D. J., Martin, L., Ware, C. B., Blish, C. A., Chang, H. Y. et al.** (2015). Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. *Nature* **522**, 221-225.
- Hancks, D. C. and Kazazian, H. H. Jr.** (2012). Active human retrotransposons: variation and disease. *Curr. Opin. Genet. Dev.* **22**, 191-203.
- Hancks, D. C. and Kazazian, H. H. Jr.** (2016). Roles for retrotransposon insertions in human disease. *Mob. DNA* **7**, 9.
- Honer, C., Chen, P., Toth, M. J. and Schumacher, C.** (2001). Identification of SCAN dimerization domains in four gene families. *Biochim. Biophys. Acta* **1517**, 441-448.
- Hubley, R., Finn, R. D., Clements, J., Eddy, S. R., Jones, T. A., Bao, W., Smit, A. F. A. and Wheeler, T. J.** (2016). The Dfam database of repetitive DNA families. *Nucleic Acids Res.* **44**, D81-D89.
- Huntley, S., Baggott, D. M., Hamilton, A. T., Tran-Gyamfi, M., Yang, S., Kim, J., Gordon, L., Branscomb, E. and Stubbs, L.** (2006). A comprehensive catalog of human KRAB-associated zinc finger genes: insights into the evolutionary history of a large family of transcriptional repressors. *Genome Res.* **16**, 669-677.
- Iacobazzi, V., Infantino, V., Convertini, P., Voza, A., Agrimi, G. and Palmieri, F.** (2009). Transcription of the mitochondrial citrate carrier gene: identification of a silencer and its binding protein ZNF224. *Biochem. Biophys. Res. Commun.* **386**, 186-191.
- Imbeault, M. and Trono, D.** (2014). As time goes by: KRABs evolve to KAP endogenous retroelements. *Dev. Cell* **31**, 257-258.
- Imbeault, M., Helleboid, P.-Y. and Trono, D.** (2017). KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. *Nature* **543**, 550-554.
- Itokawa, Y., Yanagawa, T., Yamakawa, H., Watanabe, N., Koga, H. and Nagase, T.** (2009). KAP1-independent transcriptional repression of SCAN-KRAB-containing zinc finger proteins. *Biochem. Biophys. Res. Commun.* **388**, 689-694.
- luchi, S.** (2001). Three classes of C2H2 zinc finger proteins. *Cell. Mol. Life Sci.* **58**, 625-635.
- Iyengar, S. and Farnham, P. J.** (2011). KAP1 protein: an enigmatic master regulator of the genome. *J. Biol. Chem.* **286**, 26267-26276.
- Jacobs, F. M. J., Greenberg, D., Nguyen, N., Haussler, 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.
- Jheon, A. H., Ganss, B., Cheifetz, S. and Sodek, J.** (2001). Characterization of a novel KRAB/C2H2 zinc finger transcription factor involved in bone development. *J. Biol. Chem.* **276**, 18282-18289.
- Kapopoulou, A., Mathew, L., Wong, A., Trono, D. and Jensen, J. D.** (2016). The evolution of gene expression and binding specificity of the largest transcription factor family in primates. *Evolution* **70**, 167-180.
- Kauzlaric, A., Ecco, G., Cassano, M., Duc, J., Imbeault, M. and Trono, D.** (2017). The mouse genome displays highly dynamic populations of KRAB-zinc finger protein genes and related genetic units. *PLoS ONE* **12**, e0173746.
- Kazazian, H. H. Jr.** (2004). Mobile elements: drivers of genome evolution. *Science* **303**, 1626-1632.
- 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.
- Krebs, C. J., Larkins, L. K., Price, R., Tullis, K. M., Miller, R. D. and Robins, D. M.** (2003). Regulator of sex-limitation (Rsl) encodes a pair of KRAB zinc-finger genes that control sexually dimorphic liver gene expression. *Genes Dev.* **17**, 2664-2674.
- Krebs, C. J., Schultz, D. C. and Robins, D. M.** (2012). The KRAB zinc finger protein RSL1 regulates sex- and tissue-specific promoter methylation and dynamic hormone-responsive chromatin configuration. *Mol. Cell. Biol.* **32**, 3732-3742.
- Kunaroo, G., Chia, N.-Y., Jeyakani, J., Hwang, C., Lu, X., Chan, Y.-S., Ng, H.-H. and Bourque, G.** (2010). Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat. Genet.* **42**, 631-634.
- Lamprecht, B., Walter, K., Kreher, S., Kumar, R., Hummel, M., Lenze, D., Köcher, K., Bouhrel, M. A., Richter, J., Soler, E. et al.** (2010). Derepression of an endogenous long terminal repeat activates the CSF1R proto-oncogene in human lymphoma. *Nat. Med.* **16**, 571-579, 1p following 579.
- 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. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **368**, 20120507.
- 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. *Hum. Mutat.* **17**, 511-519.
- Li, X., Ito, M., Zhou, F., Youngson, N., Zuo, X., Leder, P. and Ferguson-Smith, A. C.** (2008). A maternal-zygotic effect gene, Zfp57, maintains both maternal and paternal imprints. *Dev. Cell* **15**, 547-557.
- Liu, H., Chang, L.-H., Sun, Y., Lu, X. and Stubbs, L.** (2014). Deep vertebrate roots for mammalian zinc finger transcription factor subfamilies. *Genome Biol. Evol.* **6**, 510-525.
- Lupo, A., Cesaro, E., Montano, G., Izzo, P. and Costanzo, P.** (2011). ZNF224: structure and role of a multifunctional KRAB-ZFP protein. *Int. J. Biochem. Cell Biol.* **43**, 470-473.
- Lupo, A., Cesaro, E., Montano, G., Zurlo, D., Izzo, P. and Costanzo, P.** (2013). KRAB-zinc finger proteins: a repressor family displaying multiple biological functions. *Curr. Genomics* **14**, 268-278.
- 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. *Microbiol Spectr* **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. *Cell. Mol. Life Sci.* **65**, 3329-3347.
- Mannini, R., Riviaccio, V., D'auria, S., Tanfani, F., Ausili, A., Facchiano, A., Pedone, C. and Grimaldi, G.** (2006). Structure/function of KRAB repression domains: structural properties of KRAB modules inferred from hydrodynamic, circular dichroism, and FTIR spectroscopic analyses. *Proteins* **62**, 604-616.
- 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. *Proc. Natl. Acad. Sci. USA* **36**, 344-355.
- Medugno, L., Florio, F., De Cegli, R., Grosso, M., Lupo, A., Costanzo, P. and Izzo, P.** (2005). The Kruppel-like zinc-finger protein ZNF224 represses aldolase A gene transcription by interacting with the KAP-1 co-repressor protein. *Gene* **359**, 35-43.
- 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. *Nat. Genet.* **13**, 245-247.
- Murphy, K. E., Shylo, N. A., Alexander, K. A., Churchill, A. J., Copperman, C. and García-García, M. J. (2016). The transcriptional repressive activity of KRAB zinc finger proteins does not correlate with their ability to recruit TRIM28. *PLoS ONE* **11**, e0163555.
- 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. *Nat. Biotechnol.* **33**, 555-562.
- Nakahashi, H., Kwon, K.-R. K., Resch, W., Vian, L., Dose, M., Stavreva, D., Hakim, O., Pruett, N., Nelson, S., Yamane, A. et al. (2013). A genome-wide map of CTCF multivalency redefines the CTCF code. *Cell Rep.* **3**, 1678-1689.
- 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.
- 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. *EMBO J.* **18**, 6385-6395.
- 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. *Nat. Commun.* **6**, 6644.
- Nowick, K., Gernat, T., Almaas, E. and Stubbs, L. (2009). Differences in human and chimpanzee gene expression patterns define an evolving network of transcription factors in brain. *Proc. Natl. Acad. Sci. USA* **106**, 22358-22363.
- Nowick, K., Hamilton, A. T., Zhang, H. and Stubbs, L. (2010). Rapid sequence and expression divergence suggest selection for novel function in primate-specific KRAB-ZNF genes. *Mol. Biol. Evol.* **27**, 2606-2617.
- Oliver, C. H., Khaled, W. T., Frend, H., Nichols, J. and Watson, C. J. (2012). The Stat6-regulated KRAB domain zinc finger protein Zfp157 regulates the balance of lineages in mammary glands and compensates for loss of Gata-3. *Genes Dev.* **26**, 1086-1097.
- Organ, C. L., Shedlock, A. M., Meade, A., Pagel, M. and Edwards, S. V. (2007). Origin of avian genome size and structure in non-avian dinosaurs. *Nature* **446**, 180-184.
- 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. *Dev. Cell* **7**, 597-606.
- Puget, N., Sinilnikova, O. M., Stoppa-Lyonnet, D., Audouy, 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? *Am. J. Hum. Genet.* **64**, 300-302.
- Quenneville, S., Verde, G., Corsinotti, A., Kapopoulou, A., Jakobsson, J., Offner, S., Baglivo, I., Pedone, P. V., Grimaldi, G., Riccio, A. et al. (2011). In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. *Mol. Cell* **44**, 361-372.
- 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 Rep.* **2**, 766-773.
- Rebollo, R., Farivar, S. and Mager, D. L. (2012a). C-GATE - catalogue of genes affected by transposable elements. *Mob. DNA* **3**, 9.
- Rebollo, R., Romanish, M. T. and Mager, D. L. (2012b). Transposable elements: an abundant and natural source of regulatory sequences for host genes. *Annu. Rev. Genet.* **46**, 21-42.
- Romanish, M. T., Cohen, C. J. and Mager, D. L. (2010). Potential mechanisms of endogenous retroviral-mediated genomic instability in human cancer. *Semin. Cancer Biol.* **20**, 246-253.
- Rosenberg, N. and Jolicoeur, P. (1997). Retroviral pathogenesis. In *Retroviruses* (ed. J. M. Coffin, S. H. Hughes and H. E. Varmus). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Rowe, H. M., Jakobsson, J., Mesnard, D., Rougemont, J., Reynard, S., Aktas, T., Maillard, P. V., Layard-Liesching, H., Verp, S., Marquis, J. et al. (2010). KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* **463**, 237-240.
- Rowe, H. M., Kapopoulou, A., Corsinotti, A., Fasching, L., Macfarlan, T. S., Tarabay, Y., Viville, S., Jakobsson, J., Pfaff, S. L. and Trono, D. (2013). TRIM28 repression of retrotransposon-based enhancers is necessary to preserve transcriptional dynamics in embryonic stem cells. *Genome Res.* **23**, 452-461.
- Samuelson, L. C., Phillips, R. S. and Swanberg, L. J. (1996). Amylase gene structures in primates: retroposon insertions and promoter evolution. *Mol. Biol. Evol.* **13**, 767-779.
- Santoni De Sio, F. R. (2014). Kruppel-associated box (KRAB) proteins in the adaptive immune system. *Nucleus* **5**, 138-148.
- Santoni De Sio, F. R., Barde, I., Offner, S., Kapopoulou, A., Corsinotti, A., Bojkowska, K., Genolet, R., Thomas, J. H., Luescher, I. F., Pinschewer, D. et al. (2012a). KAP1 regulates gene networks controlling T-cell development and responsiveness. *FASEB J.* **26**, 4561-4575.
- Santoni De Sio, F. R., Massacand, J., Barde, I., Offner, S., Corsinotti, A., Kapopoulou, A., Bojkowska, K., Dagklis, A., Fernandez, M., Ghia, P. et al. (2012b). KAP1 regulates gene networks controlling mouse B-lymphoid cell differentiation and function. *Blood* **119**, 4675-4685.
- Schmitges, F. W., Radovani, E., Najafabadi, H. S., Barazandeh, M., Campitelli, L. F., Yin, Y., Jolma, A., Zhong, G., Guo, H., Kanagalingam, T. et al. (2016). Multiparameter functional diversity of human C2H2 zinc finger proteins. *Genome Res.* **26**, 1742-1752.
- Schultz, D. C., Friedman, J. R. and Rauscher, F. J. III. (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 Dev.* **15**, 428-443.
- Schultz, D. C., Ayyanathan, K., Negorev, D., Maul, G. G. and Rauscher, F. J. 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.
- Shibata, M. and Garcia-Garcia, M. J. (2011). The mouse KRAB zinc-finger protein CHATO is required in embryonic-derived tissues to control yolk sac and placenta morphogenesis. *Dev. Biol.* **349**, 331-341.
- Shibata, M., Blauvelt, K. E., Liem, K. F., Jr and Garcia-Garcia, M. J. (2011). TRIM28 is required by the mouse KRAB domain protein ZFP568 to control convergent extension and morphogenesis of extra-embryonic tissues. *Development* **138**, 5333-5343.
- Sripathy, S. P., Stevens, J. and Schultz, D. C. (2006). The KAP1 corepressor functions to coordinate the assembly of de novo HP1-demarcated microenvironments of heterochromatin required for KRAB zinc finger protein-mediated transcriptional repression. *Mol. Cell Biol.* **26**, 8623-8638.
- 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.
- Strogantsev, R., Krueger, F., Yamazawa, K., Shi, H., Gould, P., Goldman-Roberts, M., Mcewen, K., Sun, B., Pedersen, R. and Ferguson-Smith, A. C. (2015). Allele-specific binding of ZFP57 in the epigenetic regulation of imprinted and non-imprinted monoallelic expression. *Genome Biol.* **16**, 112.
- Theunissen, T. W., Friedli, M., He, Y., Planet, E., O'Neil, R. C., Markoulaki, S., Pontis, J., Wang, H., Iouranova, A., Imbeault, M. et al. (2016). Molecular criteria for defining the naive human pluripotent state. *Cell Stem Cell* **19**, 502-515.
- Thomas, J. H. and Schneider, S. (2011). Coevolution of retroelements and tandem zinc finger genes. *Genome Res.* **21**, 1800-1812.
- Thompson, P. J., Macfarlan, T. S. and Lorincz, M. C. (2016). Long terminal repeats: from parasitic elements to building blocks of the transcriptional regulatory repertoire. *Mol. Cell* **62**, 766-776.
- 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 Dev.* **6**, 1457-1465.
- 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 Res.* **24**, 1260-1270.
- Urrutia, R. (2003). KRAB-containing zinc-finger repressor proteins. *Genome Biol.* **4**, 231.
- Venturini, L., Stadler, M., Manukjan, G., Scherr, M., Schlegelberger, B., Steinemann, D. and Ganser, A. (2016). The stem cell zinc finger 1 (SZF1)/ZNF589 protein has a human-specific evolutionary nucleotide DNA change and acts as a regulator of cell viability in the hematopoietic system. *Exp. Hematol.* **44**, 257-268.
- Wallis, J. W., Aerts, J., Groenen, M. A. M., Crooijmans, R. P. M. A., Layman, D., Graves, T. A., Scheer, D. E., Kremitzki, C., Fedele, M. J., Mudd, N. K. et al. (2004). A physical map of the chicken genome. *Nature* **432**, 761-764.
- Walsh, C. P., Chaillet, J. R. and Bestor, T. H. (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat. Genet.* **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. *Proc. Natl. Acad. Sci. USA* **104**, 18613-18618.
- Wang, J., Xie, G., Singh, M., Ghanbarian, A. T., Rasko, T., Szvetnik, A., Cai, H., Besser, D., Prigione, A., Fuchs, N. V. et al. (2014). Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. *Nature* **516**, 405-409.
- 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 Res.* **23**, 505-531.
- Witzgall, R., O'leary, E., Leaf, A., Onaldi, D. and Bonventre, J. V. (1994). The Kruppel-associated box-A (KRAB-A) domain of zinc finger proteins mediates transcriptional repression. *Proc. Natl. Acad. Sci. USA* **91**, 4514-4518.

- 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. *J. Biol. Chem.* **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, G., Greenberg, D. and Macfarlan, T. S. (2015a). Spotting the enemy within: targeted silencing of foreign DNA in mammalian genomes by the Kruppel-associated box zinc finger protein family. *Mob. DNA* **6**, 17.
- Wolf, G., Yang, P., Füchtbauer, A. C., Füchtbauer, E.-M., Silva, A. M., Park, C., Wu, W., Nielsen, A. L., Pedersen, F. S. and Macfarlan, T. S. (2015b). The KRAB zinc finger protein ZFP809 is required to initiate epigenetic silencing of endogenous retroviruses. *Genes Dev.* **29**, 538-554.
- Xu, J.-H., Wang, T., Wang, X.-G., Wu, X.-P., Zhao, Z.-Z., Zhu, C.-G., Qiu, H.-L., Xue, L., Shao, H.-J., Guo, M.-X. et al. (2010). PU.1 can regulate the ZNF300 promoter in APL-derived promyelocytes HL-60. *Leuk. Res.* **34**, 1636-1646.
- Xue, Z., Huang, K., Cai, C., Cai, L., Jiang, C.-Y., Feng, Y., Liu, Z., Zeng, Q., Cheng, L., Sun, Y. E. et al. (2013). Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature* **500**, 593-597.
- Yan, L., Yang, M., Guo, H., Yang, L., Wu, J., Li, R., Liu, P., Lian, Y., Zheng, X., Yan, J. et al. (2013). Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nat. Struct. Mol. Biol.* **20**, 1131-1139.
- Zeng, Y., Wang, W., Ma, J., Wang, X., Guo, M. and Li, W. (2012). Knockdown of ZNF268, which is transcriptionally downregulated by GATA-1, promotes proliferation of K562 cells. *PLoS ONE* **7**, e29518.