

Published in final edited form as:

*Cold Spring Harb Symp Quant Biol.* 2015 ; 80: 281–288. doi:10.1101/sqb.2015.80.027573.

## Transposable elements, polydactyl proteins and the genesis of human-specific transcription networks

Didier Trono

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

### Abstract

Transposable elements (TEs) may account for up to two-thirds of the human genome, and as genomic threats they are subjected to epigenetic control mechanisms engaged from the earliest stages of embryonic development. We previously determined that an important component of this process is the sequence-specific recognition of TEs by KRAB-containing zinc finger proteins (KRAB-ZFPs), a large family of tetrapod-restricted transcription factors that act by recruiting inducers of heterochromatin formation and DNA methylation. We further demonstrated that KRAB-ZFPs and their cofactor KAP1 exert a marked influence on the transcription dynamics of embryonic stem cells via their docking of repressor complexes at TE-contained regulatory sequences. It is generally held that, beyond this early embryonic period, TEs become permanently silenced, and that the evolutionary selection of KRAB-ZFPs and other TE controllers is the result of a simple evolutionary arms race between the host and these genetics invaders. Here, I discuss recent evidence that invalidates this dual assumption, and instead suggests that KRAB-ZFPs are the instruments of a massive enterprise of TE domestication, whereby transposon-based regulatory sequences and their cellular ligands establish species-specific transcription regulation networks that influence multiple aspects of human development and physiology.

---

It has been some 65 years since Barbara McClintock, studying the genetics of maize at the Cold Spring Harbor Laboratory, discovered mobile genetic elements and proposed, based on formidably intuitive observations, that they were implicated in the control of genes (McClintock 1950). Her description of some of her findings and of what she foresaw as their broad implications can be found in two chapters of the Symposium in Quantitative Biology series, a 1951 article entitled “Chromosome organization and genic expression” (McClintock 1951) and its 1956 follow-up “Controlling elements and the gene” (McClintock 1956). The reading of these and other work published by Barbara McClintock during these years is at times arduous but always extraordinarily humbling, were it only considering that her discoveries were made when the notion of gene still was largely conceptual, long before molecular biology revolutionized the field of genetics, and that her proposals faced the skepticism of a scientific establishment convinced that genetic information was passed essentially unchanged from generation to generation. Specialists did

find virtue in her findings, but many viewed her so-called controlling elements as oddities of unclear general significance. It would take another thirty years for her work, which contributed so fundamentally to laying the foundations of modern epigenetics, to be finally rewarded by the 1983 Medicine Nobel Prize.

By then, it had been recognized that transposable elements (TEs) are present in all organisms from bacteria to humans, and that they constitute essential motors of evolution. Nevertheless, some of the most brilliant minds of the time still considered that the bulk of TE sequences found in the genome of higher species was junk or selfish DNA (Orgel and Crick 1980). The sequencing of the human genome at the beginning of the 21<sup>st</sup> century emphasized the magnitude of the issue at stake, by revealing that 50% of the human DNA can be readily identified as derived from TEs, in fact an underestimate of their true contribution to our genetic make up since over time many elements degenerate beyond recognition through the accumulation of mutations (Lander et al. 2001; Venter et al. 2001).

Transposable elements, or transposons, are phylogenetically and biologically related to viruses, hence my proposal to designate the sum of all transposons present in the genome of an organism as its endovirome. Our latest census reveals that the human endovirome counts approximately 4 million individual TE integrants, including about 500,000 DNA transposons, 650,000 endogenous retroviruses (ERVs) or LTR-retrotransposons (barely 10,000 of which are not truncated or internally recombined to yield so-called solo LTRs), 950,000 LINES (long interspersed intra-nuclear elements, only 1% of which are full-length or near full-length), 1.8 million SINEs (short interspersed intra-nuclear elements, which include some 1.2 million Alu repeats), and around 5,500 SVAs (SINE-VNTR-Alu, a hominoid-specific retroelement). A recent description of the molecular features, replication strategies and genomic impact of these mobile elements can be found elsewhere (Friedli and Trono 2015). Suffice it to mention here that i) TEs can disrupt genes, provide novel coding activities, exert a wide range of transcriptional influences and, owing to their repetitive nature, create ground for recombination events leading to genomic deletions and duplications, yet ii) only a very small minority of TEs present in the human genome (about 100 LINES, 1,000 SINEs and 50 SVAs) are still transposition-competent, accounting for one new germ line integrant in 20 to 50 human births, and none is capable of horizontal transfer.

As genomic threats, TEs are targeted by powerful RNA- and protein-based restriction mechanisms (reviewed in (Rowe and Trono 2011; Schlesinger and Goff 2014)). It is commonly viewed that these restriction mechanisms are selected as the result of an evolutionary arms race between host and TEs, and that they result in the permanent silencing of these genetic invaders through epigenetic modifications instated from the earliest stages of embryogenesis. Here, I will discuss novel evidence that invalidates this model, and instead indicates that the evolutionary selection of a major class of human TE controllers, the KRAB-containing zinc finger proteins (KRAB-ZFPs), is the instrument of a massive enterprise of domestication, whereby TE-residing regulatory sequences and their cellular ligands establish species-specific transcriptional networks likely at play in all aspects of human biology.

## KRAB-ZFPs are early embryonic repressors of transposable elements

Canonical KRAB-ZFPs contain a C-terminal array of zinc fingers with sequence-specific DNA binding ability, and an N-terminal KRAB (Krüppel-associated box) domain that can recruit KAP1 (KRAB-associated protein 1), which serves as a scaffold for a heterochromatin complex comprising histone methyltransferase, histone deacetylase, nucleosome remodeling and DNA methyltransferase activities (Iyengar and Farnham 2011) (Fig. 1). The *KRAB-ZFP* gene family emerged in early tetrapods about 400 million years ago, and expanded by gene and segment duplication to count some 350 members in both mouse and modern human, often grouped in gene clusters containing mixes of functional and degenerated paralogs (Huntley et al. 2006). During this process, *KRAB-ZFP* genes were subjected to strong positive selection, accumulating mutations at codons predicted to encode DNA-contacting residues within their zinc fingers (Emerson and Thomas 2009). As a result, only about a fifth of human and murine KRAB-ZFPs are orthologous based on their zinc fingerprint, that is, are predicted to recognize conserved genomic targets (Liu et al. 2014).

We first suspected that the KRAB/KAP1 system was involved in the control of transposable elements when we observed that it could trigger *de novo* promoter methylation during mouse early embryogenesis (Wiznerowicz et al. 2007). DNA methylation was indeed known to represent an important repressive mechanism for TEs, which are selectively targeted by this modification during the waves of genome reprogramming that otherwise wash off this and other epigenetic marks, right after fertilization and in primordial germ cells (Walsh et al. 1998; Bourc'his and Bestor 2004). Subsequent results confirmed our suspicion with the demonstration that, in embryonic carcinoma cells, KAP1 and the murine KRAB-ZFP ZFP809 induce the silencing of murine leukemia virus (an exogenous retrovirus closely related to some mouse TEs) (Wolf and Goff 2009), that KAP1 depletion in murine or human embryonic stem cells (ESCs) activates many transposable elements (Rowe et al. 2010; Turelli et al. 2014), and that KAP1 and ZFP57, a conserved member of the family, are responsible for the maintenance of DNA methylation at imprinting control regions (ICRs), which similar to TEs are selectively spared from loss of this mark during early embryonic reprogramming (Quenneville et al. 2011). Also supporting a major role for the *KRAB-ZFP* gene family in the control of TEs, the dynamics of its evolutionary expansion has coincided with waves of retroviral invasion in the germ line of higher species (Thomas and Schneider 2011). Furthermore, examining a couple of KRAB-ZFP/TE target pairs revealed the parallel evolution of these factors and TE mutants escaping their inhibition (Jacobs et al. 2014). This strongly suggested that the evolution of *KRAB-ZFPs* genes simply reflected an arms race, where repressors with new DNA binding specificity were selected in response to genome invasion by previously unseen exogenous retroviruses or mutating endogenous transposons. A series of recent evidence invalidates this model, and instead suggests that KRAB-ZFPs collectively mediate a broad exaptation of the endovirome for the benefit of the host.

## TEs and their KRAB-ZFP controllers are orchestrators of early embryogenesis

Depleting KAP1 or selected KRAB-ZFPs in murine or human embryonic stem cells results not only in activating a large range of TEs, but also in producing widespread transcriptional perturbations, some of which result from the impact of unleashed TE-based promoters or enhancers on nearby genes (Rowe et al. 2013; Turelli et al. 2014) (Fig. 2). Growing evidence also indicates that, under physiological conditions, TEs and their regulators normally partake in regulating early embryonic development. First, a large fraction of binding sites for transcription factors crucial during this period, i.e. Oct4, Nanog and CTCF, reside in the human endovirome (Bourque et al. 2008; Kunarso et al. 2010). Second, in the mouse, 2-cell stage-specific genes are transcribed from the long terminal repeat (LTR) of MERV-L, a murine endogenous retrovirus (Macfarlan et al. 2011). Third, HERV-H, a primate-restricted endogenous retrovirus, appears important for hESC pluripotency (Santoni et al. 2012; Lu et al. 2014), and our recent data indicate that its expression is regulated by a KRAB-ZFP. Fourth, *in silico* analyses of the transcriptome of early human embryos reveal the stage-restricted expression of specific subsets of TEs at all stages from 2-cell embryos to blastocysts, with production not only of TE-derived transcripts but also of mRNAs resulting from their splicing into exons of adjacent genes (Goke et al. 2015), and this is paralleled by highly stage-specific patterns of expression of KRAB-ZFP genes (Fig. 3).

## TE-binding KRAB-ZFPs are widely expressed in human adult cells

The KRAB/KAP1-mediated recognition of TEs in embryonic stem cells typically shuts down their transcription by triggering the deposition of repressive histone marks such as H3K9me3 (histone 3 trimethylated on lysine 9), which is followed by their DNA methylation through the recruitment of *de novo* DNA methyltransferases (Matsui et al. 2010; Rowe et al. 2010; Quenneville et al. 2012; Rowe et al. 2013; Turelli et al. 2014) (Fig. 2). DNA methylation is considered as a permanent silencing mark, since it is perpetuated during cell division by so-called maintenance DNA methyltransferases. It is also held as the main mechanism of TE control in adult tissues (Yoder et al. 1997). Accordingly, while the widespread expression of TE-binding KRAB-ZFPs and the secondary recruitment of KAP1 at these loci were expected in early embryos, it seemed that both phenomena should become dispensable at later stages of development, after TE methylation had been instated. We made two observations that challenged this view. First, using a combination of our own and the FANTOM Consortium data (de Hoon et al. 2015), we determined that many *KRAB-ZFP* genes are transcribed in human adult tissues, collectively displaying highly cell-specific patterns of expression. Second, we found that a significant fraction of TEs bound by KAP1 in human ESC still bore the master regulator in peripheral blood CD4<sup>+</sup> T-lymphocytes (Turelli et al. 2014). Nevertheless, we could not exclude that most KRAB-ZFPs expressed in adult cells recognized genomic loci not associated with TEs. To solve this issue, we performed a large-scale analysis of the genomic targets of human KRAB-ZFPs using ChIP-Exo, a high-precision chromatin immunoprecipitation/DNA sequencing method. The results revealed that a vast majority of these DNA-binding proteins, including many of those expressed in human differentiated cells, primarily recognize sequences contained within

transposable elements, confirming a trend previously suggested by a smaller-scale study (Najafabadi et al. 2015).

## The transcriptome provides highly specific signatures in both embryonic and adult human cells

These data strongly suggested that, in adult human cells, many TEs conserve their transcriptional potential and that the latter is modulated by the differential expression of KRAB-ZFPs. To test this hypothesis, we first optimized the experimental and *in silico* components of established RNA sequencing-based methodology to facilitate the analysis of the transcriptome, that is, the sum of TE-derived transcripts present in a cell, individually assigned to their source loci. In all cell types examined, ranging from naive to primed hESC, cord blood CD34<sup>+</sup> hematopoietic stem cells to circulating CD4<sup>+</sup> T lymphocytes, and primary hepatocytes to ESC-derived neurons, we detected transcripts originating from tens of thousands of TE integrants. TEs were expressed irrespective of their presence within or outside of genes, and whether in the sense or antisense orientation. Members from all families of mobile genetic elements were transcribed in all cells, including DNA transposons, HERVs, LINEs, SINEs and SVAs. A very robust level of reproducibility was obtained when comparing different donors or biological samples, yet transcriptomes provided unique signatures, with patterns of TE expression that were specific of the lineage, differentiation stage and activation status of the cells. For instance, the transcriptomes of resting, IL7-treated and CD3/CD28-activated T-lymphocytes were readily distinguishable, and the *in vitro* differentiation of ESC to neurons could be closely monitored through highly dynamic changes in TE expression patterns that were induced during this process.

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

The finding that specific sets of both KRAB-ZFPs and transposons were expressed in adult cells strongly suggested that the poly-zinc finger proteins kept controlling the transcriptional potential of their TE targets well beyond the early embryonic period. We first obtained evidence supporting this hypothesis in the mouse, where using a target-centered functional screen we identified two closely related KRAB-ZFP paralogs binding partly overlapping but distinguishable subsets of ERV-K. We verified that both factors repressed these endogenous retroviruses and secondarily regulated the expression of neighboring genes in murine embryonic stem cells. In addition, we determined that these KRAB-ZFPs were individually expressed in some differentiated mouse cells, where they and KAP1 regulated both their TE targets and genes situated nearby. This process stemmed from KRAB/KAP1-induced histone modifications, and appeared independent from the DNA methylation status of these loci. Also in support of the KRAB/KAP1-mediated control of TEs in somatic tissues, we found that the TE-targeting recruitment of KAP1 in human CD4<sup>+</sup> T-lymphocytes is modulated by the activation status of these cells, with KAP1 knockdown leading to loss of repressive histone marks at these loci. Although this did not systematically trigger the up-regulation of the underlying TE, some of the observed secondary transcriptional changes could be assigned to promoter or enhancer effects exerted on nearby genes by de-repressed elements.

This notably led to the aberrant activation of genes normally expressed in other tissues, strongly suggesting that the KRAB/KAP1 control of TEs is important to maintain lineage identity. Finally, changes in TE expression patterns when hESC were differentiated into neurons was matched by alterations in KAP1 recruitment and deposition of repressive or activating chromatin marks at these loci and, at least in selected cases, by changes in expression levels of the cognate KRAB-ZFP. Furthermore, a correlation between the deregulation of genes and that of nearby TEs could be established in this system as well.

## Conclusions

Our results indicate that transposable elements and their KRAB-ZFP controllers establish transcriptional regulatory networks detectable in all cells examined so far, whether embryonic or adult, multipotent or differentiated. These networks influence the expression of cellular genes, and as a result condition processes as diverse as stem cell pluripotency, maintenance of adult cell identity or response to extracellular signals (Fig. 4). Therefore, the evolutionary selection of KRAB-ZFP genes is not a simple arms race aimed at silencing TEs but rather a mediator of their domestication, with KRAB-ZFPs and associated complexes modulating the transcriptional impact of TE-residing sequences that are co-opted to regulate the expression of cellular genes.

These data corroborate our observation that, in human ESC, the KRAB/KAP1 system controls TEs long after they have lost their transposition potential through mutations. For instance, KAP1 represses an evolutionary discrete subset of LINES, predicted to have entered the human genome between 25 and 7 million years ago, none of which is still retrotransposition-competent (Castro-Diaz et al. 2014). In contrast, younger LINES are repressed by largely self-inflicted RNA interference leading to DNA methylation. This suggests a model whereby newly emerged TEs are first tamed by ancestral, RNA-based mechanisms, which probably suffice to limit their propagation. Meanwhile, expansion of the *KRAB-ZFP* gene family continuously provides a pool of proteins with new DNA binding specificity. At some point, some of these proteins recognize specific TE-contained motifs, sometimes consolidating their RNA-based transposition blockade if it has not yet been lost by mutation, but also and more systematically giving the host the opportunity to exploit these TE sequences as platforms for transcriptional regulation (Fig. 5).

The contribution of TEs and their KRAB/KAP1 controlling system to the regulation of gene expression in likely all adult human tissues is consistent with data obtained through the conditional knockout of *KAP1* in the mouse, which revealed that the master regulator partakes in processes as diverse as the management of behavioral stress, the differentiation of erythroid precursors, the maturation and activation of B- and T-lymphocytes, or the metabolism of hormones and xenobiotics in the liver (Jakobsson et al. 2008; Bojkowska et al. 2012; Chikuma et al. 2012; Santoni de Sio et al. 2012a; Santoni de Sio et al. 2012b; Barde et al. 2013). Even though KAP1 accomplishes some functions independently from its interactions with KRAB-ZFPs and TEs (Iyengar and Farnham 2011; Singh et al. 2015), these data, together with the recent observation that some ERVs are upregulated in murine B-lymphocytes deleted for SETDB1, the histone methyltransferase recruited by KAP1



(Collins et al. 2015), indicate the persistence of a dynamic control of the transcriptional potential of TEs well beyond the early embryonic period.

TEs and their KRAB-ZFP controllers likely affect gene expression in ways that largely remain to be deciphered. *Cis*-acting influences such as promoter, enhancer or repressor effects affecting the expression of nearby genes have been abundantly documented in ES or differentiated cells, either under physiological conditions or upon KRAB-ZFP, KAP1 or SETDB1 depletion (Faulkner et al. 2009; Rebollo et al. 2011; Rebollo et al. 2012; Rowe et al. 2013; Turelli et al. 2014; Collins et al. 2015). We anticipate the transcriptional impact of the KRAB-ZFP-mediated tuning of TEs to be far more diversified. First, the exploration of longer-range influences has so far only been very superficial, because of the difficulty of interpreting assays suited for their detection within the context of repetitive sequences, for instance chromosomal conformation capture studies. Second, our ongoing large-scale analysis of the protein interactome of human KRAB-ZFPs reveals that these factors can associate with a variety of complexes involved in either transcriptional or post-transcriptional regulation, not just with KAP1 and heterochromatin-inducing activities. Accordingly, independent of their own regulatory potential, TEs could simply serve as platforms for the docking of these KRAB-ZFP-tethered functional units. Finally, it is likely that TEs, which are increasingly recognized as a major source of regulatory RNAs (Kapusta et al. 2013; Ha et al. 2014), also exert a variety of influences in *trans*, the identification of which is technically challenging. Nevertheless, the suggestion that an HERVH-produced long noncoding RNA might contribute to hESC pluripotency (Lu et al. 2014) and our finding that the KRAB/KAP1 system regulates erythropoiesis by controlling the expression of micro-RNAs targeting mitophagy genes (Barde et al. 2013) are illustrative of this general mode of action.

An important implication of our results is that the endovirome and its KRAB-ZFP controllers confer a high degree of species-specificity to the conduct of probably all biological processes relevant to human development and physiology (Fig. 6). Indeed, a large fraction of the recognizable mobile elements in our genome are unique to humans or close relatives, both in sequence and genomic location. Correspondingly, only a small fraction of human KRAB-ZFPs has orthologues outside of the primate lineage (Nowick et al. 2010; Liu et al. 2014). Therefore, while animal models such as the mouse are valid for delineating general principles and studying canonical aspects of gene regulation, many *cis*- and *trans*-regulatory features likely essential to human biology can only be studied in the cognate system, whether in tissue culture or *in vivo*. The apparent contribution of HERV-H, a primate-specific endogenous retrovirus, to human ESC pluripotency (Lu et al. 2014), and the role of MERV-L, one of its murine-restricted counterparts, as driver of the expression of zygotic activation genes in the mouse (Macfarlan et al. 2011), are good illustrations of this fundamental principle.

Our results also indicate that mutations at critical positions of the endovirome could translate into pathological phenotypes. This calls for techniques suitable to explore not only sequencer-friendly regions of the genome but also the more challenging environment of repetitive sequences.

KRAB-ZFPs and their TE docking platforms are emerging from these studies as likely essential factors in the speciation of all higher vertebrates, including humans. Interestingly, the central nervous system stands out as the organ where the widest range of KRAB-ZFPs is expressed, and recently evolved KRAB-ZFPs have been noted to contribute prominently to differences in gene expression between the brains of chimpanzees and humans (Nowick et al. 2009). Furthermore, LINE and SVA transcriptional activity can be detected in cultured human neural stem cells and somatic retrotransposition events have been documented in the human brain (Muotri et al. 2005; Coufal et al. 2009; Baillie et al. 2011; Evrony et al. 2015). It has been suggested that the resulting *de novo* LINE integrants contribute to neuronal diversification and brain plasticity (Richardson et al. 2014). However, considering that these events are rare, that TE integration is random, and that most LINE integrants are 5' truncated hence devoid of transcriptional activity, it seems unlikely that such a phenomenon is of major significance, and even less that it underlays what cannot be an anarchic response to external stimuli. Instead, it appears more plausible that it is the transcriptional response of the endovirome and its KRAB-ZFP ligands to afferences received by the brain, which translates into epigenetic changes that can durably modify the biological behavior of individual neurons, hence contribute to neuronal cell memory and diversification.

## Acknowledgments

I am grateful to the past and present members of my group, as well as the many other researchers whose discoveries help shape our view on the role of the endovirome and its KRAB-ZFP controllers in human biology. I thank Evelyne Ruchi and Alexandra Iouranova for help with the illustrations. Work in my laboratory is supported by the Swiss National Science Foundation, the European Research Council, the European Union 7th Framework Program (INGENIUM Network for Initial Training), the Gebert-Rüf Foundation and the Rothschild Caesarea Foundation.

## References

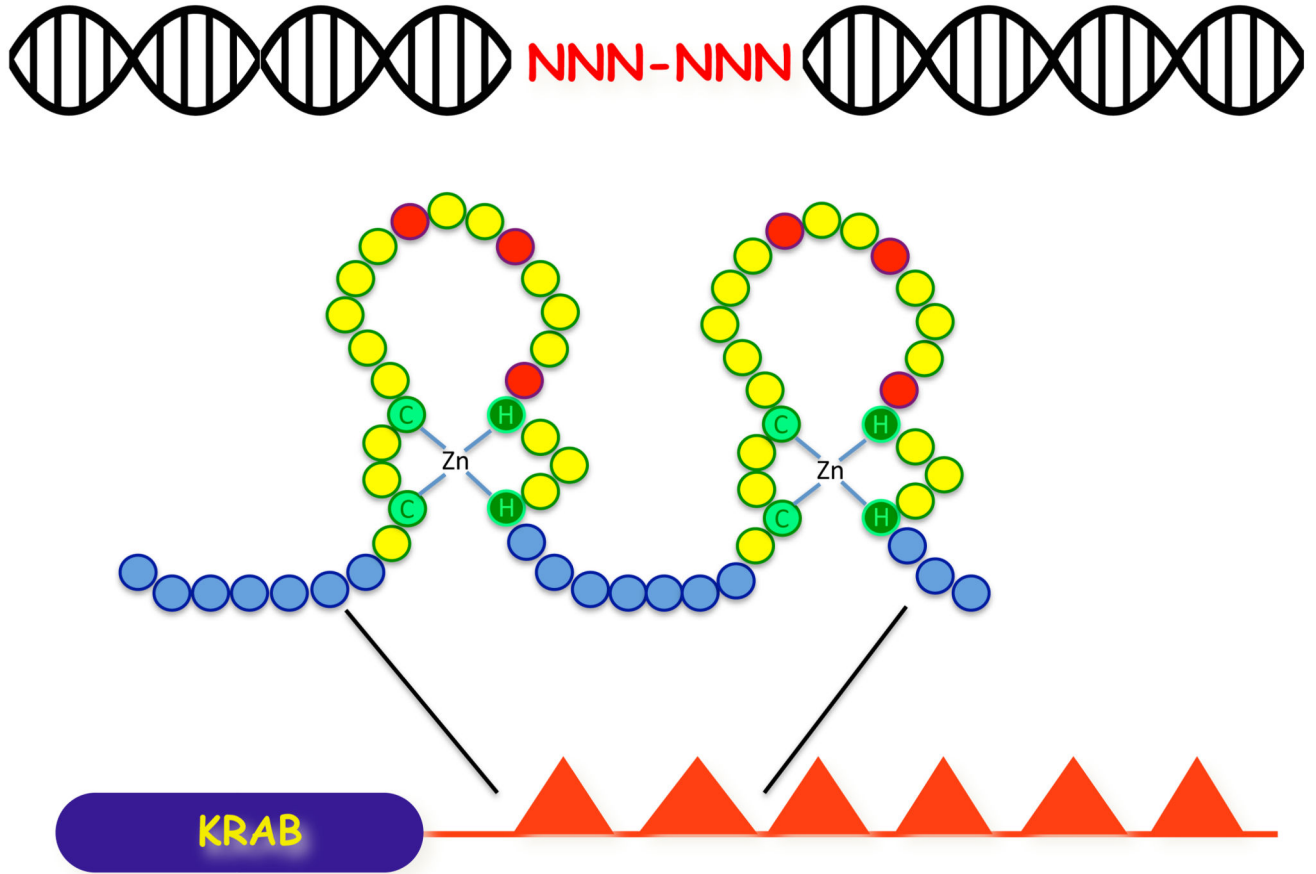
- Baillie JK, Barnett MW, Upton KR, Gerhardt DJ, Richmond TA, De Sapio F, Brennan PM, Rizzu P, Smith S, Fell M, et al. Somatic retrotransposition alters the genetic landscape of the human brain. *Nature*. 2011; 479:534–537. [PubMed: 22037309]
- Barde I, Rauwel B, Marin-Florez RM, Corsinotti A, Laurenti E, Verp S, Offner S, Marquis J, Kapopoulou A, Vanicek J, et al. A KRAB/KAP1-miRNA Cascade Regulates Erythropoiesis Through Stage-Specific Control of Mitophagy. *Science*. 2013
- Bojkowska K, Aloisio F, Cassano M, Kapopoulou A, Santoni de Sio F, Zangger N, Offner S, Cartoni C, Thomas C, Quenneville S, et al. Liver-specific ablation of Kruppel-associated box-associated protein 1 in mice leads to male-predominant hepatosteatosis and development of liver adenoma. *Hepatology*. 2012; 56:1279–1290. [PubMed: 22684873]
- Bourc'his D, Bestor TH. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature*. 2004; 431:96–99. [PubMed: 15318244]
- Bourque G, Leong B, Vega VB, Chen X, Lee YL, Srinivasan KG, Chew JL, Ruan Y, Wei CL, Ng HH, et al. Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome research*. 2008; 18:1752–1762. [PubMed: 18682548]
- Castro-Diaz N, Ecco G, Coluccio A, Kapopoulou A, Yazdanpanah B, Friedli M, Duc J, Jang SM, Turelli P, Trono D. Evolutionally dynamic L1 regulation in embryonic stem cells. *Genes & development*. 2014; 28:1397–1409. [PubMed: 24939876]
- Chikuma S, Suita N, Okazaki IM, Shibayama S, Honjo T. TRIM28 prevents autoinflammatory T cell development in vivo. *Nature immunology*. 2012; 13:596–603. [PubMed: 22544392]
- Collins PL, Kyle KE, Egawa T, Shinkai Y, Oltz EM. 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*. 2015; 112:8367–8372. [PubMed: 26100872]



- Coufal NG, Garcia-Perez JL, Peng GE, Yeo GW, Mu Y, Lovci MT, Morell M, O'Shea KS, Moran JV, Gage FH. L1 retrotransposition in human neural progenitor cells. *Nature*. 2009; 460:1127–1131. [PubMed: 19657334]
- de Hoon M, Shin JW, Carninci P. Paradigm shifts in genomics through the FANTOM projects. *Mammalian genome : official journal of the International Mammalian Genome Society*. 2015
- Emerson RO, Thomas JH. Adaptive evolution in zinc finger transcription factors. *PLoS genetics*. 2009; 5:e1000325. [PubMed: 19119423]
- Evrony GD, Lee E, Mehta BK, Benjamini Y, Johnson RM, Cai X, Yang L, Haseley P, Lehmann HS, Park PJ, et al. Cell lineage analysis in human brain using endogenous retroelements. *Neuron*. 2015; 85:49–59. [PubMed: 25569347]
- Faulkner GJ, Kimura Y, Daub CO, Wani S, Plessy C, Irvine KM, Schroder K, Cloonan N, Steptoe AL, Lassmann T, et al. The regulated retrotransposon transcriptome of mammalian cells. *Nature genetics*. 2009; 41:563–571. [PubMed: 19377475]
- Friedli M, Trono D. The developmental control of transposable elements and the evolution of higher species. *Annual Review of Cellular and Developmental Biology*. 2015; 31:13.11–13.23.
- Goke J, Lu X, Chan YS, Ng HH, Ly LH, Sachs F, Szczerbinska I. Dynamic transcription of distinct classes of endogenous retroviral elements marks specific populations of early human embryonic cells. *Cell stem cell*. 2015; 16:135–141. [PubMed: 25658370]
- Ha H, Song J, Wang S, Kapusta A, Feschotte C, Chen KC, Xing J. A comprehensive analysis of piRNAs from adult human testis and their relationship with genes and mobile elements. *BMC genomics*. 2014; 15:545. [PubMed: 24981367]
- Huntley S, Baggott DM, Hamilton AT, Tran-Gyamfi M, Yang S, Kim J, Gordon L, Branscomb E, Stubbs L. A comprehensive catalog of human KRAB-associated zinc finger genes: insights into the evolutionary history of a large family of transcriptional repressors. *Genome research*. 2006; 16:669–677. [PubMed: 16606702]
- Iyengar S, Farnham PJ. KAP1 protein: an enigmatic master regulator of the genome. *The Journal of biological chemistry*. 2011; 286:26267–26276. [PubMed: 21652716]
- Jacobs FM, Greenberg D, Nguyen N, Haeussler M, Ewing AD, Katzman S, Paten B, Salama SR, Haussler D. An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. *Nature*. 2014; 516:242–245. [PubMed: 25274305]
- Jakobsson J, Cordero MI, Bisaz R, Groner AC, Busskamp V, Bensadoun JC, Cammas F, Losson R, Mansuy IM, Sandi C, et al. KAP1-mediated epigenetic repression in the forebrain modulates behavioral vulnerability to stress. *Neuron*. 2008; 60:818–831. [PubMed: 19081377]
- Kapusta A, Kronenberg Z, Lynch VJ, Zhuo X, Ramsay L, Bourque G, Yandell M, Feschotte C. Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. *PLoS genetics*. 2013; 9:e1003470. [PubMed: 23637635]
- Kunarski G, Chia NY, Jeyakani J, Hwang C, Lu X, Chan YS, Ng HH, Bourque G. Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nature genetics*. 2010; 42:631–634. [PubMed: 20526341]
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001; 409:860–921. [PubMed: 11237011]
- Liu H, Chang LH, Sun Y, Lu X, Stubbs L. Deep vertebrate roots for mammalian zinc finger transcription factor subfamilies. *Genome biology and evolution*. 2014; 6:510–525. [PubMed: 24534434]
- Lu X, Sachs F, Ramsay L, Jacques PE, Goke J, Bourque G, Ng HH. The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. *Nature structural & molecular biology*. 2014; 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. Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A. *Genes & development*. 2011; 25:594–607. [PubMed: 21357675]
- Matsui T, Leung D, Miyashita H, Maksakova IA, Miyachi H, Kimura H, Tachibana M, Lorincz MC, Shinkai Y. Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature*. 2010; 464:927–931. [PubMed: 20164836]

- McClintock B. The origin and behavior of mutable loci in maize. *Proceedings of the National Academy of Sciences of the United States of America*. 1950; 36:344–355. [PubMed: 15430309]
- McClintock B. Chromosome organization and genic expression. *Cold Spring Harbor symposia on quantitative biology*. 1951; 16:13–47. [PubMed: 14942727]
- McClintock B. Controlling elements and the gene. *Cold Spring Harbor symposia on quantitative biology*. 1956; 21:197–216. [PubMed: 13433592]
- Muotri AR, Chu VT, Marchetto MC, Deng W, Moran JV, Gage FH. Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature*. 2005; 435:903–910. [PubMed: 15959507]
- Najafabadi HS, Mnaimneh S, Schmitges FW, Garton M, Lam KN, Yang A, Albu M, Weirauch MT, Radovani E, Kim PM, et al. C2H2 zinc finger proteins greatly expand the human regulatory lexicon. *Nature biotechnology*. 2015; 33:555–562.
- Nowick K, Gernat T, Almaas E, Stubbs L. Differences in human and chimpanzee gene expression patterns define an evolving network of transcription factors in brain. *Proceedings of the National Academy of Sciences of the United States of America*. 2009; 106:22358–22363. [PubMed: 20007773]
- Nowick K, Hamilton AT, Zhang H, Stubbs L. Rapid sequence and expression divergence suggest selection for novel function in primate-specific KRAB-ZNF genes. *Molecular biology and evolution*. 2010; 27:2606–2617. [PubMed: 20573777]
- Orgel LE, Crick FH. Selfish DNA: the ultimate parasite. *Nature*. 1980; 284:604–607. [PubMed: 7366731]
- Quenneville S, Turelli P, Bojkowska K, Raclot C, Offner S, Kapopoulou A, Trono D. The KRAB-ZFP/KAP1 system contributes to the early embryonic establishment of site-specific DNA methylation patterns maintained during development. *Cell reports*. 2012; 2:766–773. [PubMed: 23041315]
- Quenneville S, Verde G, Corsinotti A, Kapopoulou A, Jakobsson J, Offner S, Baglivo I, Pedone PV, Grimaldi G, Riccio A, et al. In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. *Molecular cell*. 2011; 44:361–372. [PubMed: 22055183]
- Rebollo R, Karimi MM, Bilenky M, Gagnier L, Miceli-Royer K, Zhang Y, Goyal P, Keane TM, Jones S, Hirst M, et al. Retrotransposon-induced heterochromatin spreading in the mouse revealed by insertional polymorphisms. *PLoS genetics*. 2011; 7:e1002301. [PubMed: 21980304]
- Rebollo R, Miceli-Royer K, Zhang Y, Farivar S, Gagnier L, Mager DL. Epigenetic interplay between mouse endogenous retroviruses and host genes. *Genome biology*. 2012; 13:R89. [PubMed: 23034137]
- Richardson SR, Morell S, Faulkner GJ. L1 retrotransposons and somatic mosaicism in the brain. *Annual review of genetics*. 2014; 48:1–27.
- Rowe HM, Jakobsson J, Mesnard D, Rougemont J, Reynard S, Aktas T, Maillard PV, Layard-Liesching H, Verp S, Marquis J, et al. KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature*. 2010; 463:237–240. [PubMed: 20075919]
- Rowe HM, Kapopoulou A, Corsinotti A, Fasching L, Macfarlan TS, Tarabay Y, Viville S, Jakobsson J, Pfaff SL, Trono D. TRIM28 repression of retrotransposon-based enhancers is necessary to preserve transcriptional dynamics in embryonic stem cells. *Genome research*. 2013; 23:452–461. [PubMed: 23233547]
- Rowe HM, Trono D. Dynamic control of endogenous retroviruses during development. *Virology*. 2011; 411:273–287. [PubMed: 21251689]
- Santoni de Sio FR, Barde I, Offner S, Kapopoulou A, Corsinotti A, Bojkowska K, Genolet R, Thomas JH, Luescher IF, Pinschewer D, et al. KAP1 regulates gene networks controlling T-cell development and responsiveness. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*. 2012a; 26:4561–4575. [PubMed: 22872677]
- Santoni de Sio FR, Massacand J, Barde I, Offner S, Corsinotti A, Kapopoulou A, Bojkowska K, Dagklis A, Fernandez M, Ghia P, et al. KAP1 regulates gene networks controlling mouse B-lymphoid cell differentiation and function. *Blood*. 2012b; 119:4675–4685. [PubMed: 22452978]
- Santoni FA, Guerra J, Luban J. HERV-H RNA is abundant in human embryonic stem cells and a precise marker for pluripotency. *Retrovirology*. 2012; 9:111. [PubMed: 23253934]

- Schlesinger S, Goff SP. Retroviral Transcriptional Regulation and Embryonic Stem Cells: War and Peace. *Molecular and cellular biology*. 2014
- Singh K, Cassano M, Planet E, Sebastian S, Jang SM, Sohi G, Faralli H, Choi J, Youn HD, Dilworth FJ, et al. A KAP1 phosphorylation switch controls MyoD function during skeletal muscle differentiation. *Genes & development*. 2015; 29:513–525. [PubMed: 25737281]
- Thomas JH, Schneider S. Coevolution of retroelements and tandem zinc finger genes. *Genome research*. 2011; 21:1800–1812. [PubMed: 21784874]
- Turelli P, Castro-Diaz N, Marzetta F, Kapopoulou A, Raclot C, Duc J, Tieng V, Quenneville S, Trono D. Interplay of TRIM28 and DNA methylation in controlling human endogenous retroelements. *Genome research*. 2014; 24:1260–1270. [PubMed: 24879559]
- Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, et al. The sequence of the human genome. *Science*. 2001; 291:1304–1351. [PubMed: 11181995]
- Walsh CP, Chaillet JR, Bestor TH. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nature genetics*. 1998; 20:116–117. [PubMed: 9771701]
- Wiznerowicz M, Jakobsson J, Szulc J, Liao S, Quazzola A, Beermann F, Aebischer P, Trono D. The Kruppel-associated box repressor domain can trigger de novo promoter methylation during mouse early embryogenesis. *The Journal of biological chemistry*. 2007; 282:34535–34541. [PubMed: 17893143]
- Wolf D, Goff SP. Embryonic stem cells use ZFP809 to silence retroviral DNAs. *Nature*. 2009; 458:1201–1204. [PubMed: 19270682]
- Yoder JA, Walsh CP, Bestor TH. Cytosine methylation and the ecology of intragenomic parasites. *Trends in genetics : TIG*. 1997; 13:335–340. [PubMed: 9260521]

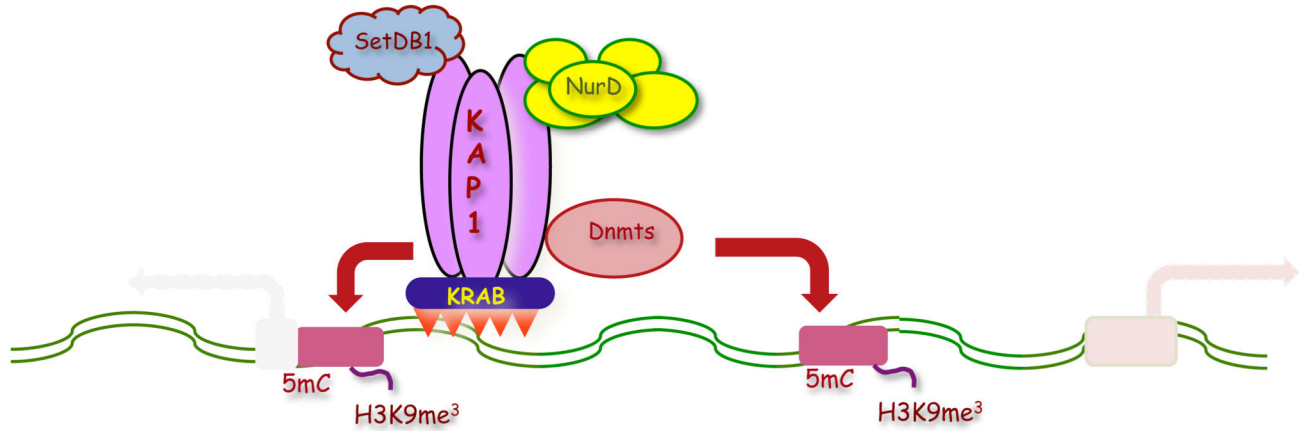


**Fig. 1. Sequence-specific DNA recognition by a prototypic KAPB-ZFP.**

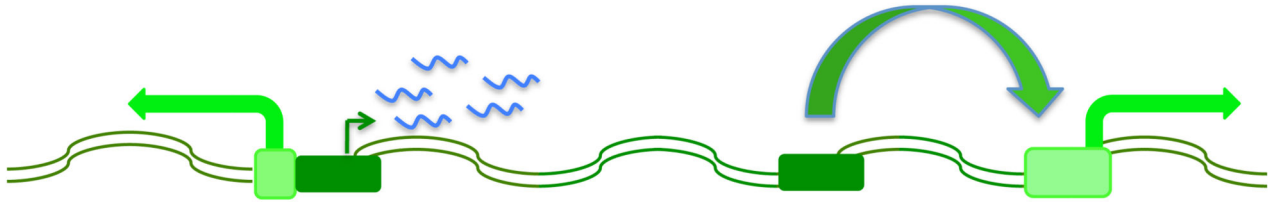
Top, DNA target, with two targeted triplers of deoxynucleotides highlighted. Middle, amino-acid representation of two zinc fingers and flanking spacers (blue balls), typically 7 residue-long. Within each ZF, amino acids represented in red (at positions conventionally designated -1, +3 and +6 within the ZF) are major determinant of DNA binding sequence specificity.

Bottom, general structure of a prototypic KRAB-ZFP, with N-terminal KAP1-recruiting domain (KRAB) and C-terminal poly-zinc finger array. Some KRAB-ZFPs harbor additional N-terminal functional domains (e.g. SCAN, BTB, DUF),

## Early embryonic cell, repressed TE locus

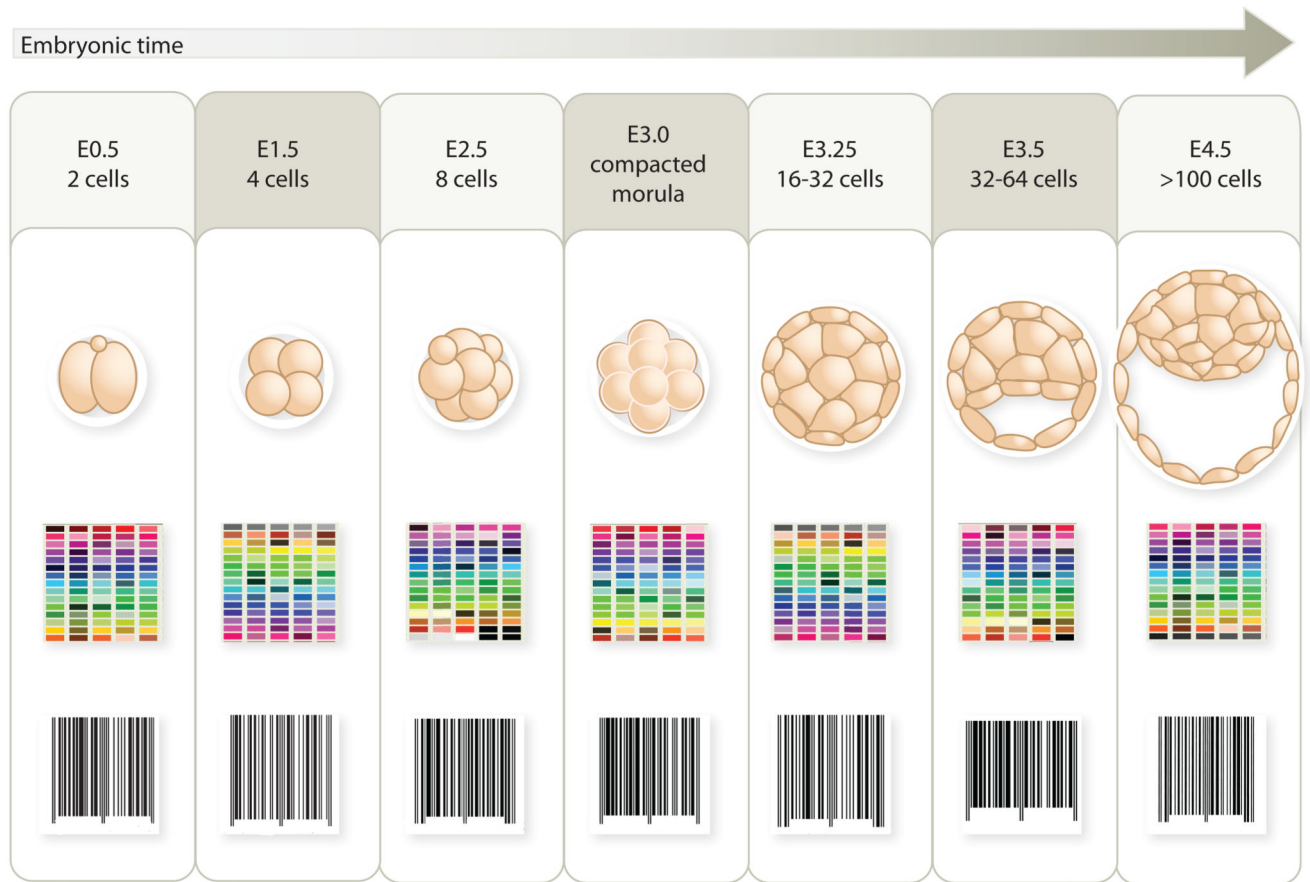


## Early embryonic cell, unrepressed TE locus



**Fig. 2. The KRAB'n'KAP system controls the transcriptional activity of TEs in pluripotent stem cells.**

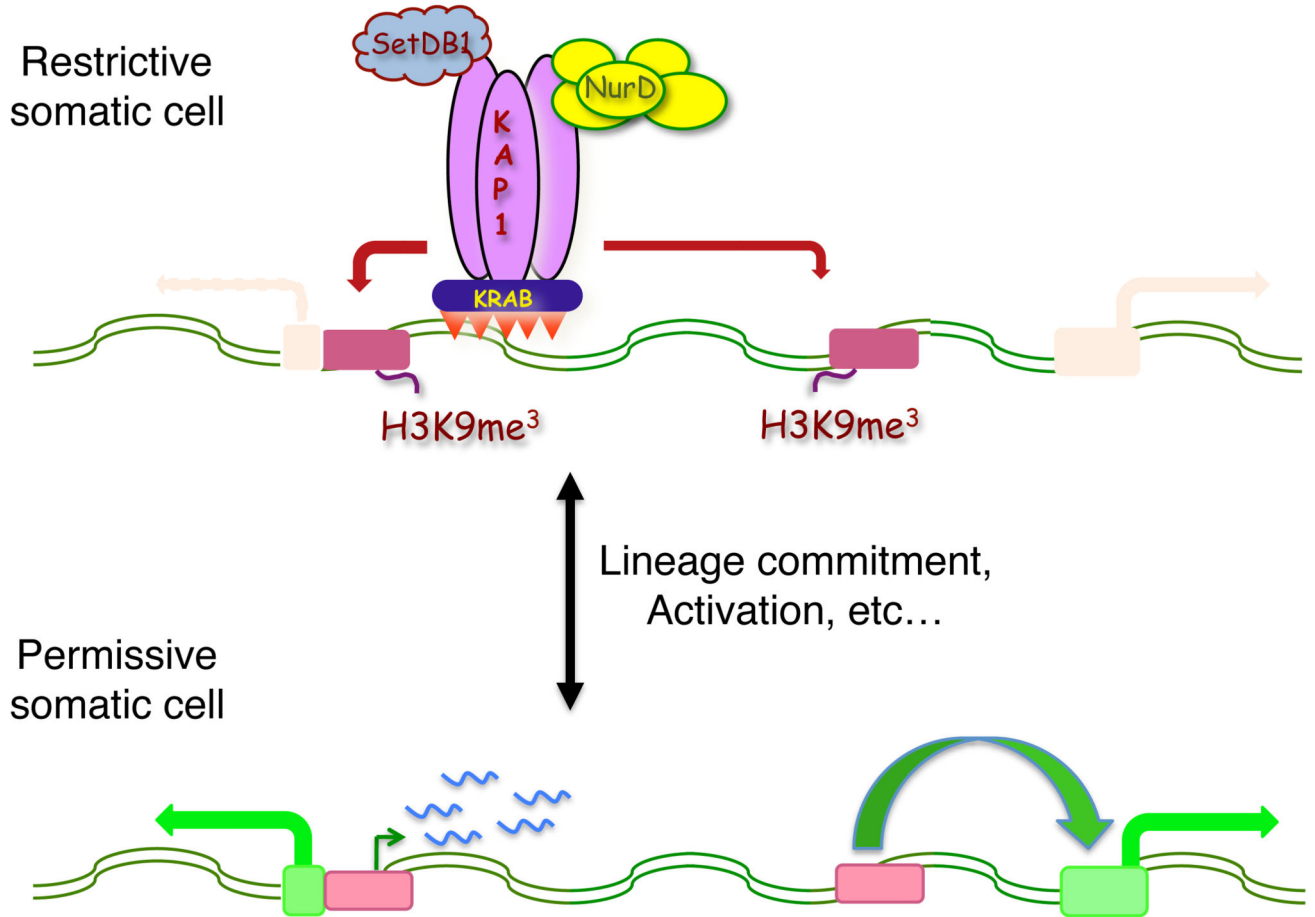
In embryonic stem cells, the sequence-specific recognition of TEs by their cognate KRAB-ZFPs results in docking KAP1 and its associated effectors at these loci, leading to repression by deposition of repressive histone marks (e.g. H3K9me<sup>3</sup>) and cytosine methylation (5<sup>mC</sup>), through the respective actions of SETDB1 and other chromatin modifiers (e.g. NurD) and of DNA methyltransferases (DNMTs). When transcriptionally active, TEs not only produce transcripts, some of which can have long-range regulatory functions, but can also stimulate the expression of nearby genes through promoter or enhancer effects.



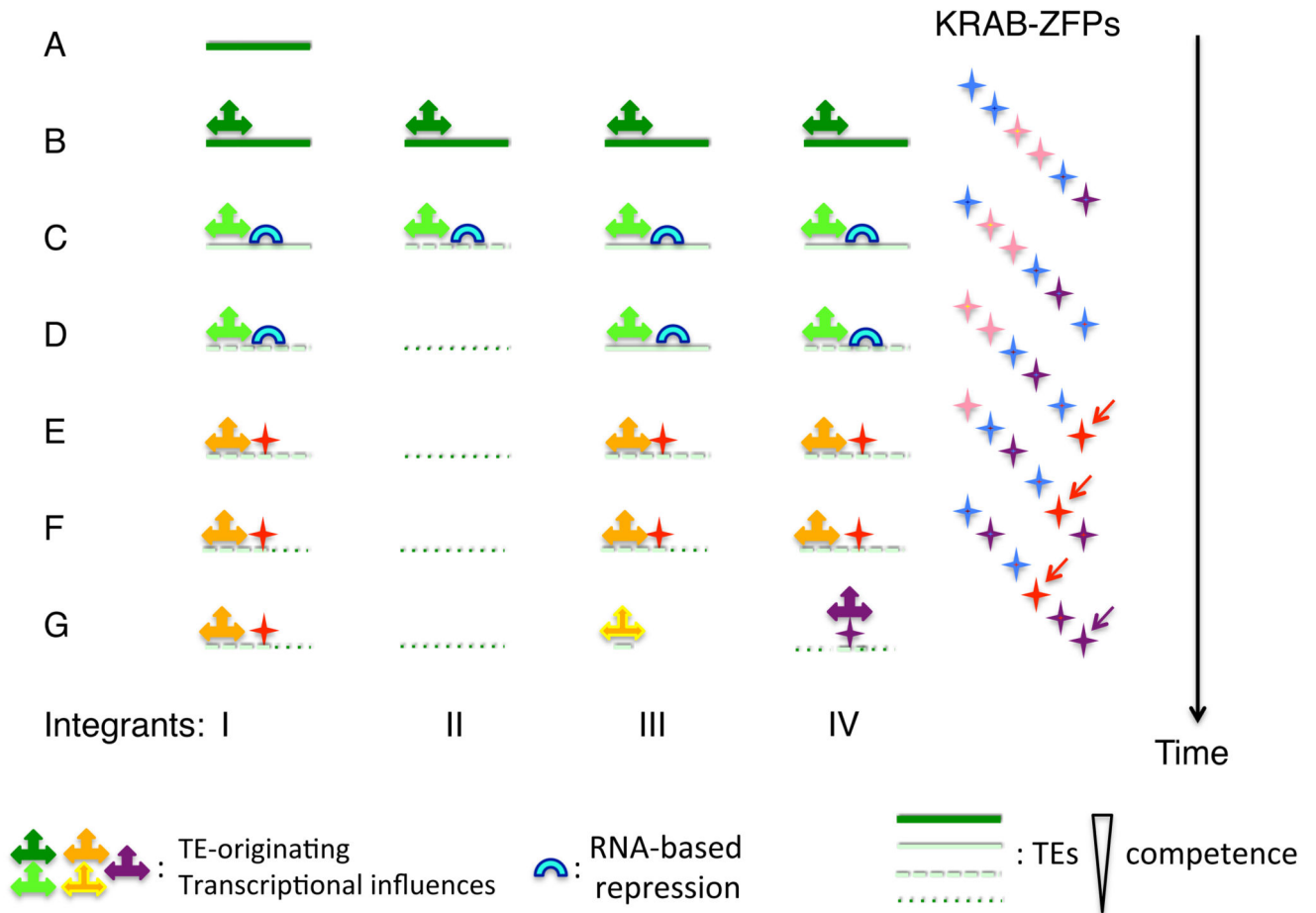
**Fig. 3. KRAB-ZFPs and their TE targets display temporally regulated patterns of expression in early embryonic human cells.**

Highly stage-differential levels of RNAs derived from both transposable elements and their KRAB-ZFP controllers define a double transcriptional barcode for each step of early embryonic human development, as schematically represented here.



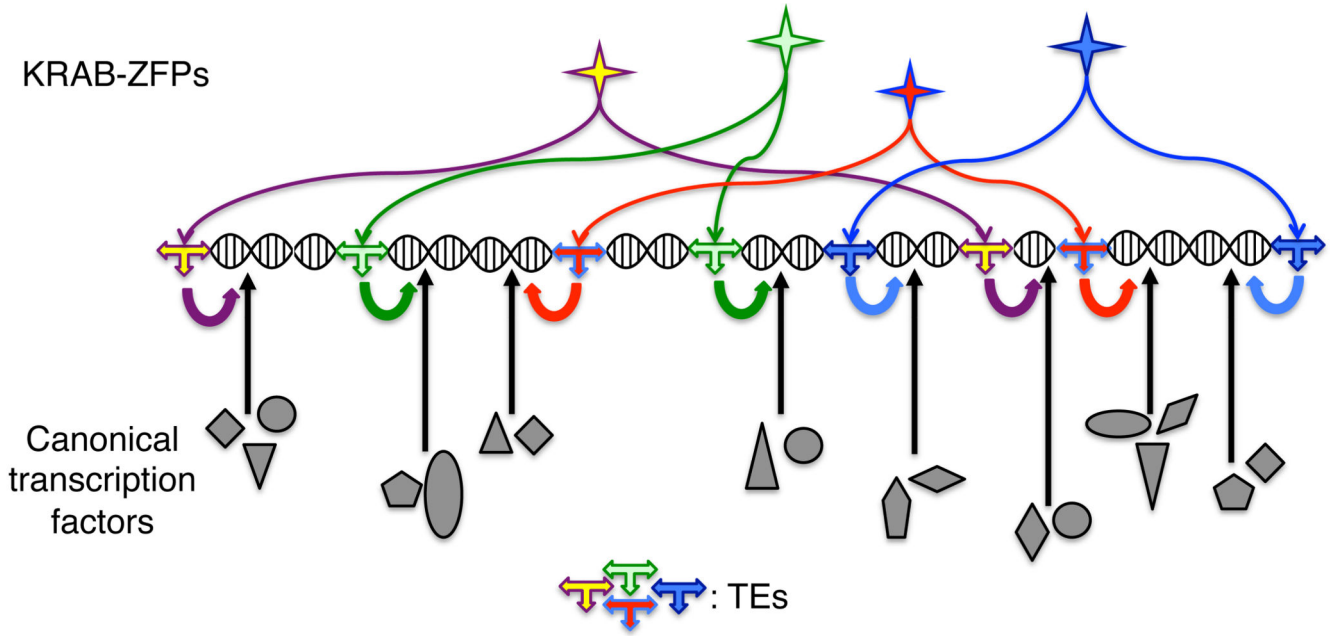


**Fig. 4. KRAB-ZFPs and their TE targets modulate gene expression in adult tissues.** In adult cells, KRAB-ZFPs also control TE transcriptional influences, although in this case histone-based modifications rather than changes in DNA methylation seem to prevail.



**Fig. 5. A brief history of TE time.**

When a new TE emerges from either exogenous or endogenous sources (A), it initially can spread (B), before being largely controlled by RNA-based restriction mechanisms such as piRNAs (C). Its individual integrants accumulate mutations that progressively inactivate their transposition potential (D). Meanwhile, the KRAB-ZFP gene family generates new paralogs, one of which can at some point recognize and control this group of TEs (E). Individual integrants continue to undergo genetic drift (F), so that in some cases only their KRAB-ZFP-recruiting transcription regulatory region is left, as part of a cellular gene promoter or enhancer (G, integrant III). The effector complexes brought about by the KRAB-ZFPs can themselves evolve over time, so that all that is ultimately left of the original TE/repressor pair is a DNA target motif and a sequence-specific polypeptidic ligand (G, integrant IV).



**Fig. 6. Of the specific-specific regulation of human biological processes by KRAB-ZFPs and their TE targets.**

TEs sprinkle the human genome with sequences largely unique to this species both in sequence and genomic location. These TE sequences are used as landing platforms for KRAB-ZFPs, which themselves display limited levels of orthology with similar proteins from other species. A triply species-restricted regulatory layer is thus laid over canonical, conserved transcription pathways. The impact of this phenomenon on speciation might be particularly pronounced in organs subjected to environmental constraints that are not overly coercive, such as the brain.