ANNUAL REVIEWS

Annual Review of Genetics

The Arms Race Between KRAB–Zinc Finger Proteins and Endogenous Retroelements and Its Impact on Mammals

Melania Bruno, Mohamed Mahgoub, and Todd S. Macfarlan

Eunice Kennedy Shriver National Institute of Child Health and Human Development, The National Institutes of Health, Bethesda, Maryland 20892, USA; email: todd.macfarlan@nih.gov

Annu. Rev. Genet. 2019. 53:393–416

First published as a Review in Advance on September 13, 2019

The *Annual Review of Genetics* is online at genet.annualreviews.org

https://doi.org/10.1146/annurev-genet-112618-043717

Keywords

KRAB-ZFPs, endogenous retroelements, EREs, evolutionary arms race, genomic imprinting

Abstract

Nearly half of the human genome consists of endogenous retroelements (EREs) and their genetic remnants, a small fraction of which carry the potential to propagate in the host genome, posing a threat to genome integrity and cell/organismal survival. The largest family of transcription factors in tetrapods, the Krüppel-associated box domain zinc finger proteins (KRAB-ZFPs), binds to specific EREs and represses their transcription. Since their first appearance over 400 million years ago, KRAB-ZFPs have undergone dramatic expansion and diversification in mammals, correlating with the invasions of new EREs. In this article we review our current understanding of the structure, function, and evolution of KRAB-ZFPs and discuss growing evidence that the arms race between KRAB-ZFPs and the EREs they target is a major driving force for the evolution of new traits in mammals, often accompanied by domestication of EREs themselves.

INTRODUCTION

Transposable element (TE): genetic element than can mobilize and integrate in different positions of the genome

Endogenous retroelement (ERE): TE that mobilizes through an RNA intermediate in a copy-and-paste manner; also called a retrotransposon

Long terminal repeat (LTR): present at 5′ and 3′ ends of ERVs; contains regulatory sequences including promoter, enhancer, and polyA signal

Long interspersed nuclear element (LINE): non-LTR-ERE highly represented in mammals; LINE-1 (L1s) are still active in humans and mice

Short interspersed nuclear element (SINE): highly abundant nonautonomous, noncoding, and non-LTR-ERE

Endogenous retrovirus (ERV): LTR-containing retrotransposons representing the genetic remnants of past retroviral infections that integrated into the host germline genome

The generation of genetic diversity and the natural selection of beneficial traits are essential for the survival of life in hostile environments. Genetic diversity is generated by several mutational processes, including DNA damage, replication errors, gene duplication, and chromosome recombination, among others, but transposable elements (TEs) are unique in their ability to mutate genomes as a by-product of their selfish lifestyle. TEs account for an astounding fraction of mammalian genomes. For example, while only 1–2% of both human and mouse genomes encode for proteins (60a, 121a, 123a), nearly half consists of TEs and their remains, although the precise amount is difficult to estimate because ancient TEs continue to decay by genetic drift and young active TEs vary among individuals (2, 64).

TEs can be divided into two major classes: the DNA transposons, which account for approximately 3% of the human genome and mobilize via a cut-and-paste mechanism, and the retrotransposons or endogenous retroelements (EREs), which mobilize through an RNA intermediate in a copy-and-paste fashion, thus having the potential to integrate multiple copies at different locations within the genome (34). EREs can be further divided into non-long terminal repeat (non-LTR) retrotransposons, including the long interspersed nuclear elements (LINEs or L1s) and short interspersed nuclear elements (SINEs), and LTR retrotransposons, often termed endogenous retroviruses (ERVs). ERVs are the proviral DNA remnants of ancient retrovirus infections that integrated into the host germline genome. Once integrated as proviruses in germline cells, ERVs are vertically transmitted from parent to offspring just like host genes and therefore accumulate mutations at the same rate as host DNA, unlike exogenous retroviruses, which can rapidly mutate. Owing to the continuous infection of the germline by retroviruses and by the ongoing mutation of ERVs, LINEs, and SINEs via mutational drift and recombination, EREs can be arranged into distinct clades, the classification of which has been a major ongoing effort (51, 72, 110, 124). As a result of the continuous appearance of new EREs and their genetic drift, host genomes contain a unique repertoire of EREs that tend to be more similar in more closely related species, often facilitating the determination of phylogenetic relationships between species (47).

Functional EREs (those capable of retrotransposition) carry both protein-coding sequences and regulatory elements that ensure the transcription, processing, translation, reverse transcription, and integration of ERE sequences. In response to the threat posed by functional EREs, host organisms have evolved numerous strategies to defend themselves at nearly every step of their life cycle. These defense mechanisms have likely been critical not only to allow for EREs to be tolerated and therefore spread throughout all domains of the eukaryotic tree of life, but also to enable the domestication of ERE sequences for the benefit of the host. There are numerous recent examples of host domestication of ERE protein-coding genes like *Gag* and *Env* (28, 91), and of gene regulatory elements like enhancers, insulators, alternative promoters, splice acceptor/donor sites, and transcription termination sites (20, 34, 115). These findings raise an important question: How is the domestication of EREs by the host made possible given the immediate dangers that active mutagenic EREs pose?

In this article we discuss how a family of rapidly evolving transcription factors, the Krüppel-associated box domain zinc finger proteins (KRAB-ZFPs), likely facilitated the domestication of ERE regulatory sequences by serving the immediate needs of the host to transcriptionally silence active EREs. KRAB-ZFPs, which constitute the largest family of transcription factors in vertebrates, have expanded and diversified in parallel with the rise of novel EREs, distinguishing themselves as one of the most rapidly evolving families of proteins (111, 114, 117, 123). Here, we first discuss the protein domains and structures that define the KRAB-ZFP family. Next we discuss their unique genomic organization, which reflects and likely facilitates their rapid evolution. We then highlight the experimental and evolutionary evidence that supports the hypothesis that

this KRAB-ZFP expansion is being driven primarily by EREs. We also discuss individual cases of young KRAB-ZFPs that repress EREs and ancient KRAB-ZFP genes that evolved to play critical roles in mammalian development. Finally, we highlight the future challenges that will need to be overcome to more fully explore the exciting biology of KRAB-ZFP genes.

STRUCTURE OF KRAB–ZINC FINGER PROTEINS

KRAB-ZFPs are characterized by two main structural features: an N-terminal KRAB domain and an array of a variable number of tandem C2H2 zinc fingers at the C terminus.

C2H2 Zinc Fingers

The C2H2 zinc finger is an ancient protein domain found in green plants, fungi, and metazoans (see the sidebar titled The Discoveries and Origins of Zinc Finger Genes and the KRAB Domain) (83). C2H2 zinc finger proteins most commonly bind to DNA, but they also bind to RNA and proteins (14, 17). C2H2 zinc fingers are generally 28–30 amino acids long and share high structural similarity. The C2H2 zinc finger motif is named for two cysteines and two histidines that coordinate the binding of a zinc ion, adopting a structural conformation with two β strands and one α helix (ββα fold) (**Figure 1*a*,*b***). The DNA binding specificity of C2H2 zinc fingers resides in the α-helix amino acids, which contact the major groove of DNA, allowing consecutive fingers to wrap around the nucleic acid double helix. Typically, the α-helix amino acids at the positions −1, +2, +3, and +6 make base contacts with a DNA stretch of three consecutive nucleotides, plus one nucleotide on the opposite strand. These amino acids are thus termed the fingerprint amino acids (93, 130) (**Figure 1*a*,*b***).

Several computational tools have been developed to predict zinc finger protein DNA binding specificity, taking advantage of large data sets examining the binding specificity of individual fingers (containing unique fingerprint amino acids) in the context of multifingered proteins (79, 84). The possibility to predict and also modify zinc finger binding specificity has also been exploited to create molecular tools in which the C2H2 zinc finger domain can be engineered to bind desired DNA sequences and fused to effector modules, allowing the tethering of endonucleases, recombinases, or transposases to genomic regions of interest for genome engineering (55, 57).

Krüppel: *Drosophila* protein representing one of the first examples of a protein with tandem C2H2 zinc fingers

Krüppel-associated box domain (KRAB): protein domain associated primarily with potent transcriptional repression via the association with the corepressor KAP1

Krüppel-associated box domain zinc finger protein (KRAB-ZFP): family of proteins with an N-terminal KRAB domain and a C-terminal C2H2 zinc finger array

C2H2 zinc finger: protein domain in which two cysteines and two histidines coordinate the binding of a zinc ion

THE DISCOVERIES AND ORIGINS OF ZINC FINGER GENES AND THE KRAB DOMAIN

Zinc finger genes can be traced to the common ancestor of green plants, fungi, and animals. In plants and fungi, the diversity of zinc finger DNA binding activity is restricted to a handful of DNA triplets, but in animals there is a wide diversity of DNA binding activity such that zinc fingers can bind to every possible triplet of DNA nucleotides (83). The first C2H2 zinc finger proteins identified were the TFIIIA and Krüppel transcription factors in *Xenopus* and *Drosophila*, respectively (78, 107). The KRAB domain was initially identified as a recurrent motif encoded by the 5′ exons of a cluster of Krüppel-like zinc finger genes found on human chromosome 19 and was thus termed the Krüppel-associated box domain (10). The fully intact KRAB domain can be traced to the coelacanth *Prdm9* gene (42) (also called *Meisetz*), but a partial KRAB domain called the KRAB interior (KRI) motif can be traced to the sea urchin *Meisetz* gene itself (12). Neither of these domains bind to the corepressor KRAB-associated protein 1 (KAP1), suggesting that mutations to KRAB were necessary to enable KAP1 interaction. In coelacanth, the KRAB domain is most commonly found on the same exon as the zinc finger array, but in most species it is found on an earlier separate exon.

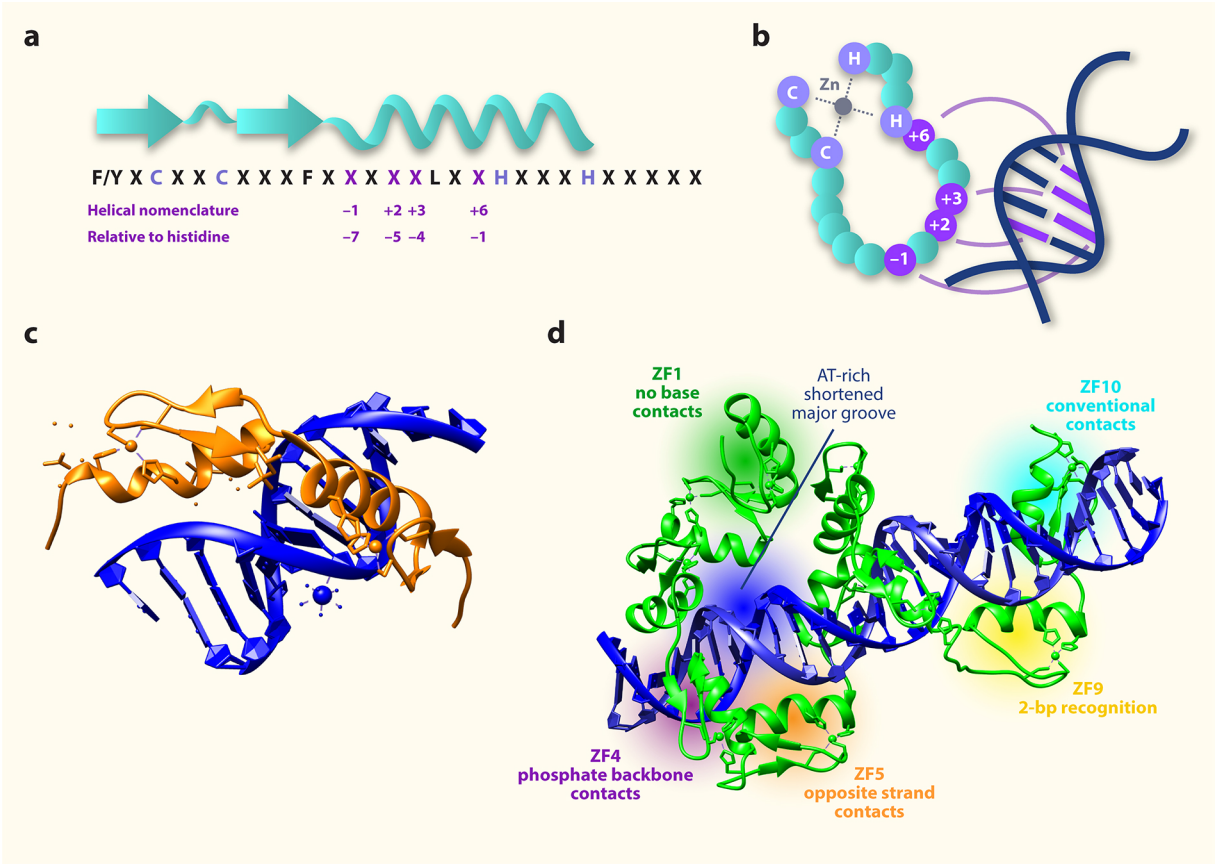


Figure 1

The structure of C2H2 zinc fingers. (*a*) Schematic representation of the two β-strands and the α-helix that form the C2H2 zinc finger, with important structural and fingerprint residues highlighted. (*b*) Schematic representation of the zinc ion bound by the two cysteines and histidines and of the −1, +2, +3, and +6 fingerprint amino acids binding to DNA nucleotides. (*c*) Structure of Zfp57 zinc fingers binding to methylated DNA (PDB ID 4GZN) (66). (*d*) Structure of the tandem zinc finger array of Zfp568 binding to its target DNA stretch (PDB ID 5V3M) (92). Abbreviation: ZF, zinc finger.

However, several factors have precluded precision genome engineering by C2H2 zinc fingers, likely preventing their more widespread use for synthetic biology. First, the amino acids adjacent to fingerprint amino acids and the linkers between fingers can modify binding affinity (19). Second, since some C2H2 zinc fingers can sense DNA methylation (65) (**Figure 1*c***), the epigenetic signature of the target sequence also contributes to the binding specificity. Third, zinc fingers in tandem can influence each other's binding, so that the final DNA stretch bound by a zinc finger array is not always the consecutive sum of the target sequence of each individual zinc finger. This is best exemplified by the discrepancies between predicted DNA target sequences and bona fide target sequences determined experimentally (42, 82, 84, 106). The unpredictability of zinc finger binding specificity is especially relevant for the Krüppel-like zinc finger family, the largest and most diverse family of C2H2 zinc finger–containing proteins, which contains arrays of zinc fingers ranging from 3 to 40 tandem zinc fingers, with an average of 12 fingers in humans (120). Structural studies aimed at visualizing the mode of binding of tandem C2H2 zinc fingers to their target DNA have demonstrated that not all fingers in the array contribute to sequence-specific

binding. Only some fingers make conventional fingerprint/base pair contacts with the DNA major groove, while others make DNA backbone interactions or intramolecular interactions with other zinc fingers (92, 94). A good example of these noncanonical interactions can be found with the interaction of several fingers of Zfp568 with its target sequence at the *Igf2* gene, which were likely induced by an AT-rich stretch in the target sequence that altered the width of the major groove, indicating that zinc finger arrays are flexible in response to DNA shape (92) (**Figure 1*d***).

The KRAB Domain

The KRAB domain consists of 50–75 amino acids and can be divided into separate subdomains or boxes: the KRAB A box, which can also be found alone, and several auxiliary boxes denoted B, BL, b, and C, which, when present, are encoded by a separate exon (10, 40, 67, 74). The KRAB A box folds into two α helices, with the rest of the domain flexible and unstructured (PDB ID 1V65). The KRAB domain, and in particular its A box, displays strong transcriptional repressive properties (73, 96, 125). The KRAB domain does not have intrinsic enzymatic function, but it works as a transcriptional repressor by tethering the KRAB-associated protein 1 (KAP1) to the DNA sequences recognized by the zinc finger array (35, 56, 80) (**Figure 2**).

Although the precise composition, stoichiometry, and structure of the KRAB/KAP1 corepressor complex has yet to be determined, several biochemical studies have provided some insights into its molecular assembly. KAP1, also known as tripartite interaction motif-containing 28 (TRIM28) or transcriptional intermediary factor 1β (TIF1β), was originally shown to homo-oligomerize and strongly bind to the KRAB domain with 3:1 stoichiometry (95). It acts as a cofactor and scaffold protein for the recruitment of the H3K9 histone methyltransferase SET domain bifurcated histone lysine methyltransferase 1 (SETDB1), the heterochromatin protein 1 (HP1), and chromatin remodelers with histone deacetylase activity (87, 105, 108, 109). KAP1 binds to the KRAB domain via an N-terminal RING, B-box, and coiled-coil domain (RBCC)/tripartite interaction motif (TRIM) domain (35) and in turn serves as a scaffold protein for the recruitment of HP1 by a central PXVXL pentapeptide region (87, 105). A recent study found that KAP1 forms an antiparallel dimer with functional asymmetry, with HP1 occupying only one of the two potential binding sites on the KAP1 dimer (33). KAP1 also displays auto-E3 SUMO ligase activity exerted by a special conformation of its plant homeodomain (PHD) finger and bromodomain (43). KAP1 can thus SUMOylate itself and this event is required for additional recruitment of the nucleosome remodeling and histone deacetylation (NuRD) complex (109) and of SETDB1 (108), both of which help modify the chromatin environment and repress transcription. KAP1 is also implicated in establishment of de novo DNA methylation, at least in embryonic stem cells (119, 126), promoting irreversible gene silencing and imprinting (133). More recent studies have demonstrated that KAP1 may play a more important role in maintaining DNA methylation by protecting methylated CpGs from genome-wide demethylation by Tet proteins (22, 27).

KRAB-associated protein 1 (KAP1): KRAB domain–binding protein that recruits histone modifiers and chromatin remodelers to promote heterochromatin formation; also called TRIM28

SET domain bifurcated histone lysine methyltransferase 1 (SETDB1): protein recruited by KAP1; trimethylates lysine 9 of histone H3 (H3K9me3)

Auxiliary Domains

A small number of ancient KRAB-ZFPs contain the SCAN (SRE-ZBP, CTfin51, AW-1 and Number18 cDNA) or DUF3669 (Domain of Unknown Function 3669) domain at their N termini (21, 42). The SCAN domain is derived from the C-terminal portion of the Gag capsid protein from the Gmr1-like family of Gypsy/Ty3-like retrotransposons (31). Since capsid proteins multimerize to form core retroviral and retrotransposon capsids, it has been speculated that the SCAN domain may have been used to target host zinc finger proteins to retroelement capsids, although this has not yet been tested. The SCAN domain can form dimers, suggesting it may promote homodimerization or heterodimerization between SCAN domain–containing KRAB-ZFPs (44).

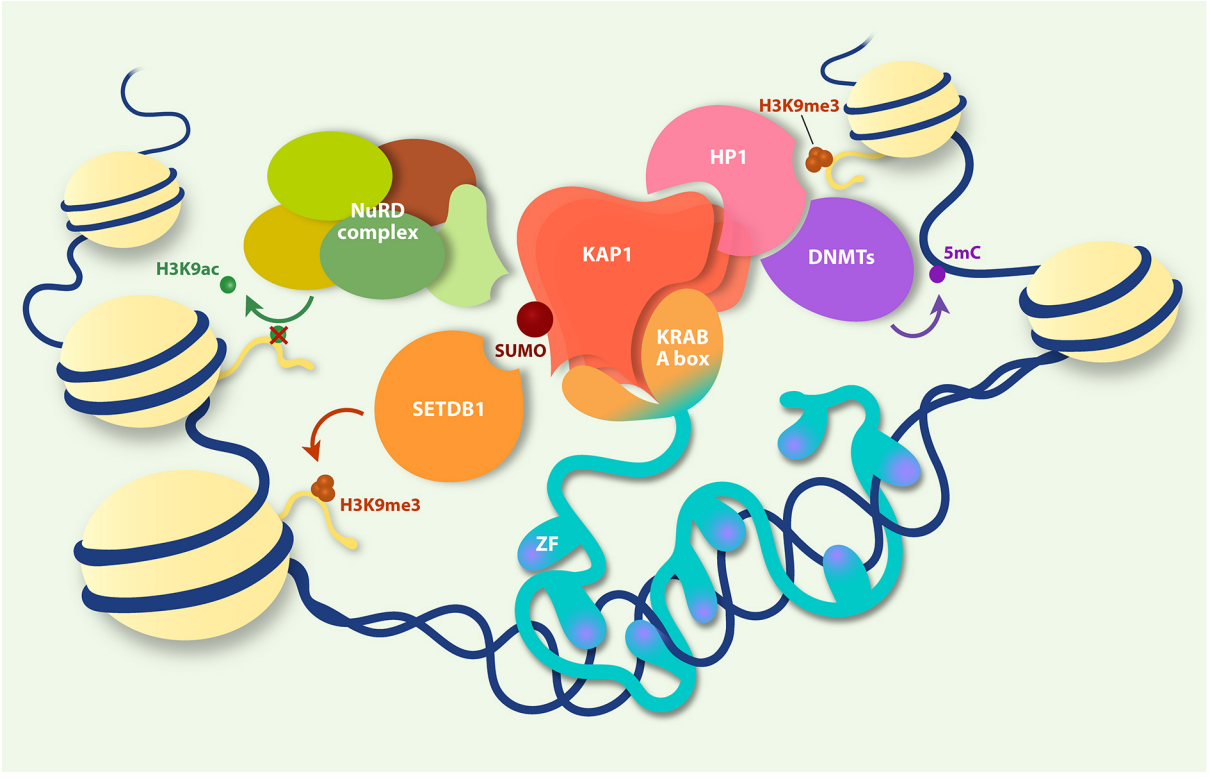
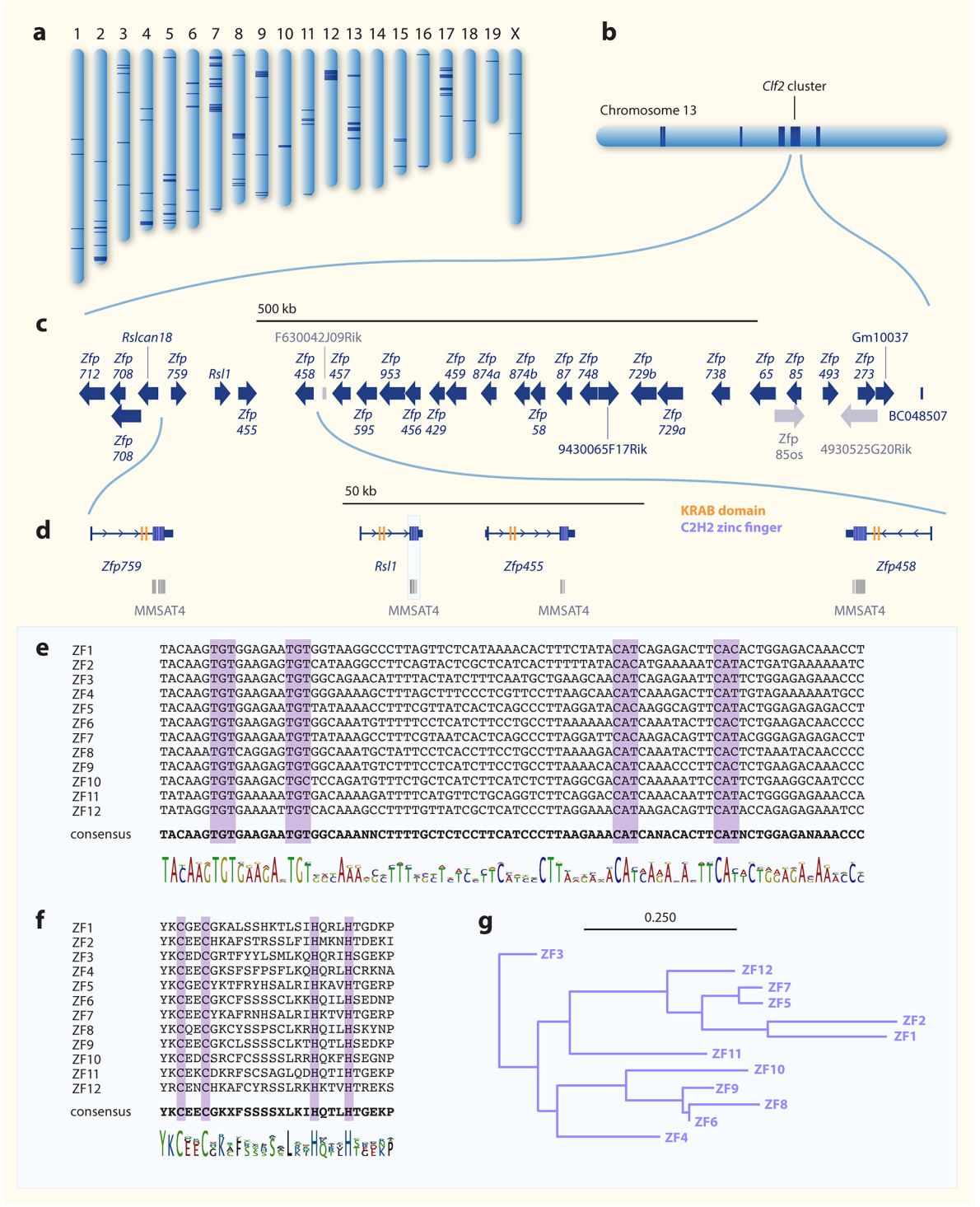


Figure 2

KRAB-ZFPs bind to their target DNA via the zinc finger array and recruit the corepressor KAP1. KAP1 in turn serves as scaffold for HP1 binding and DNMTs. In addition, KAP1 auto-SUMOylation recruits the H3K9 histone methyltransferase SETDB1 and the chromatin remodeling and histone deacetylase complex NuRD, favoring the deposition of transcriptional repressive marks. Abbreviations: 5mC, 5-methylcytosine; DNMTs, DNA methyltransferases; HP1, heterochromatin protein 1; KAP1, KRAB-associated protein 1; NuRD, nucleosome remodeling and deacetylase (complex); SETDB1, SET domain bifurcated histone lysine methyltransferase 1; ZF, zinc finger.

KRAB–ZINC FINGER PROTEIN GENE ORGANIZATION

KRAB-ZFP genes are repetitive on multiple scales (**Figure 3**). Each KRAB-ZFP gene consists of an earlier exon (or exons) encoding the KRAB A box and accessory KRAB domains and a final 3′ exon encoding the tandem zinc finger array. The zinc finger array can be classified as a coding minisatellite sequence [called MMSAT4 in the mouse genome (53)] owing to its inherent repetitiveness (**Figure 3*d–f***). KRAB-ZFP genes are then arranged primarily in clusters consisting almost exclusively of KRAB-ZFP gene family members. In humans, these clusters are predominantly found on chromosome 19 (9, 10, 40, 68), whereas in mice they are spread throughout many chromosomes, including large clusters on Chromosomes 2, 4, 10, 13, and 17, among others (53) (**Figure 3*a***). The higher degree of sequence similarity among KRAB domains within a cluster and the presence of sometimes large segmental duplication tracks consisting of one or multiple KRAB-ZFP genes suggest that tandem in situ duplication is the primary driver for the creation of new family members (30, 40) (**Figure 3*g***; also see the sidebar titled How Many KRAB–Zinc Finger Protein Genes Are There?). However, phylogenetic analyses have also revealed that unrelated



(Caption appears on following page)

Figure 3 (*Figure appears on preceding page*)

Repetitiveness of KRAB-ZFP genes at different scales. (*a*) KRAB-ZFP gene distribution on mouse chromosomes. (*b*) Detail of mouse Chromosome 13, with the *Clf2* cluster highlighted. (*c*) The *Clf2* cluster has a high density of KRAB-ZFP genes, with high genic repetitiveness, within a small genomic region. (*d*) The 3′ exon of KRAB-ZFP genes represents an additional level of repetitiveness, with each zinc finger array composed of multiple copies of the zinc finger satellite MMSAT4. (*e*) Example of the DNA sequence repetitiveness within the 3′ exon of *Rsl1* gene; the triplets encoding for the conserved structural cysteines and histidines are highlighted. (*f*) Comparison of the amino acidic sequence of Rsl1 zinc fingers. (*g*) Maximum-likelihood phylogenetic tree of Rsl1 zinc fingers generated by comparison of their DNA coding sequence, showing how the zinc finger array is likely the result of duplication and diversification of distinct fingers.

zinc finger genes can be physically intermixed, indicating that distributed duplication events and recombination have also shaped the diversification of KRAB-ZFP genes.

Recent de novo assemblies of several mouse strains have further revealed dramatic differences in KRAB-ZFP gene content within several mouse clusters, suggesting that very recent duplication and deletion events have further diversified KRAB-ZFP gene content throughout Muroidea (64). Whereas duplication events are a necessary prerequisite for amplification and functional diversification of KRAB-ZFP genes, natural selection has also played an important role. KRAB-ZFP gene paralogs display evidence of positive selection within their zinc finger domains, with selective pressure on the zinc fingerprint amino acids that likely confers changes to DNA binding specificity (30). However, numerous KRAB-ZFPs display evidence of purifying selection of fingerprint amino acids, allowing the identification of putative DNA binding functional orthologs between species (42). Thus, the KRAB-ZFP gene repertoire in a given species consists of many more recently evolved KRAB-ZFPs with rapidly evolving DNA-binding domains that may be unique to that species or shared only with more closely related species, and a smaller number of ancestral KRAB-ZFPs that have been maintained by purifying selection (**Figure 4*a*,*b***). As we describe in more detail in the next section, the younger KRAB-ZFP genes tend to bind primarily to ERE sequences and play a role in their transcriptional silencing, whereas the more ancient KRAB-ZFPs, particularly those that arose in mammals and have been maintained by purifying selection, play important roles in mammal-specific biological processes such as genomic imprinting.

One intriguing question that remains is whether there is something unique about KRAB-ZFP genes and their chromatin environment that facilitates (or perhaps restricts) their evolutionary turnover. KRAB-ZFP gene clusters display high KAP1 genomic occupancy, as determined by chromatin immunoprecipitation sequencing (ChIP-seq) experiments, raising the intriguing hypothesis that there might be an internal regulatory network controlling KRAB-ZFP genes (45). Indeed, KRAB-ZFP genes are also decorated with a unique chromatin signature at their 3′ exon encoding the minisatellite zinc finger array, which displays both the H3K36me3 mark associated

HOW MANY KRAB–ZINC FINGER PROTEIN GENES ARE THERE?

The estimated number of C2H2 zinc finger proteins has steadily grown with increased genome sequencing and assembly (8), reaching more than 700 C2H2 zinc finger–containing proteins (of which 423 are KRAB domain containing) in humans, making them the single largest protein family by domain (40). More recent surveys of KRAB-ZFP genes in the common ancestors of coelacanth, lungfish, and tetrapods found between 200 and 400 genes in most species analyzed, with a smaller number in birds (8 on average), approximately 600 in mice, up to 800 in opossums, and 1,200 in the domesticated guinea pig (42, 53). A significant fraction of KRAB-ZFP genes are pseudogenes. However, these are just estimates, because KRAB-ZFP gene clusters often contain gaps in the assembly and can vary from strain to strain (64) and perhaps from individual to individual.

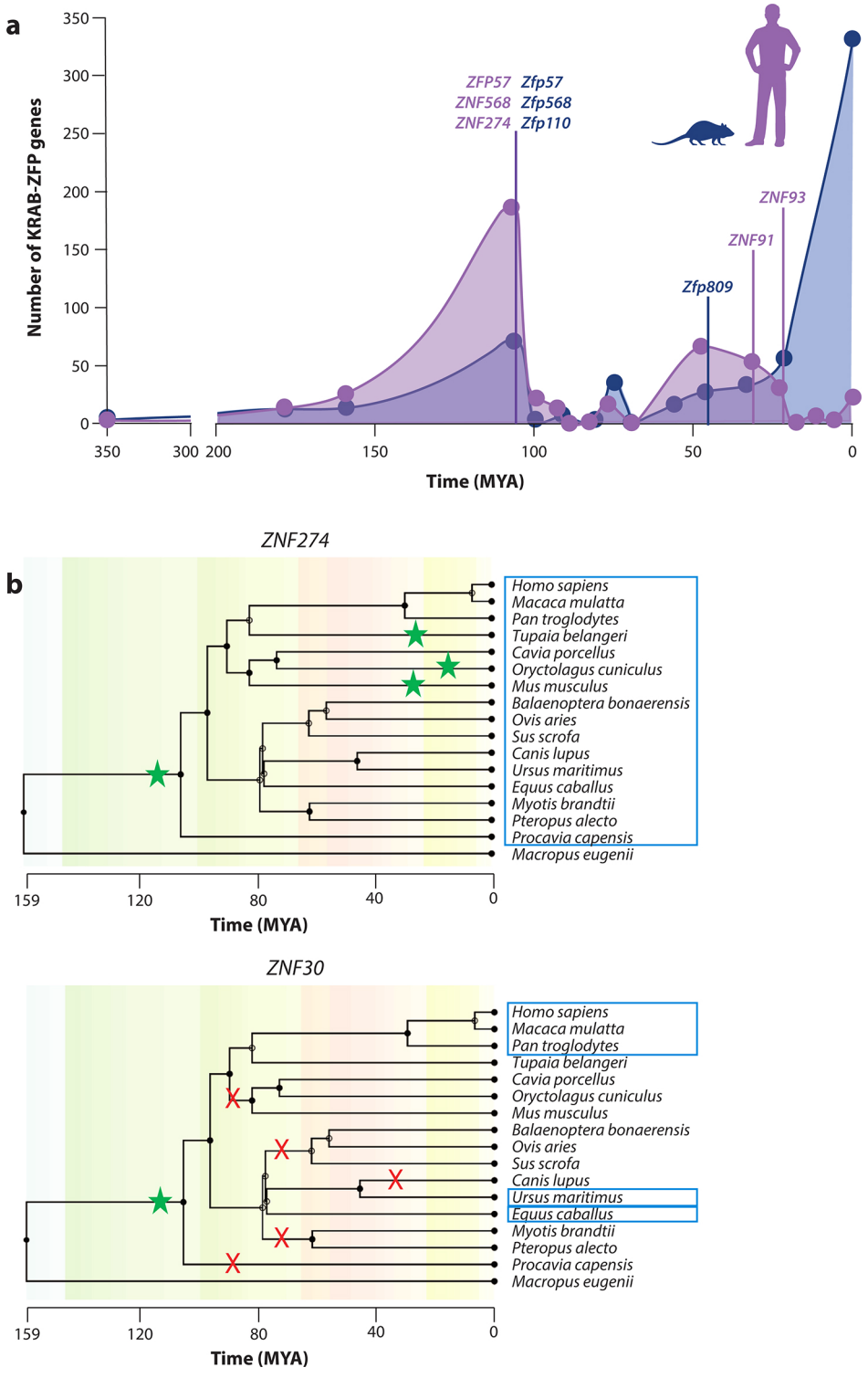


Figure 4

Life and death of KRAB-ZFP genes. (*a*) Histogram of the approximate ages of all existent KRAB-ZFP genes in mice and humans. The approximate age of each KRAB-ZFP gene was determined first by identifying orthologs based on zinc fingerprint alignments as described in Reference 42, and then by estimating evolutionary distances between species using the TimeTree database. A noticeable burst of new KRAB-ZFP genes occurred prior to the mammalian radiation over 105 MYA, with separate bursts along the human tree prior to the split from New World monkeys (68 MYA). In the mouse lineage a recent burst of KRAB-ZFP genes since the split with rats has dramatically increased the number of KRAB-ZFP genes. (*b*) Example trees showing the appearance and maintenance of two example KRAB-ZFP genes, *ZNF274* and *ZNF30*, along many branches. A green star indicates the appearance of the gene; a red cross indicates gene loss. Abbreviation: MYA, million years ago.

with actively transcribed gene bodies and the H3K9me3 mark typically associated with transcriptional repression (37). The H3K9me3 mark at the 3′ exon of many KRAB-ZFP genes is likely deposited by the direct recruitment of KAP1/SETDB1 by one or more KRAB-ZFPs, including ZNF274 and ZNF75D (36, 42, 121), suggesting that KRAB-ZFP genes are autoregulated (121). Despite these findings, little is known about the three-dimensional genomic structure of KRAB-ZFP clusters owing to the difficulty in correctly mapping to such repetitive regions. Nonetheless, there is some evidence that KRAB-ZFP gene clusters are compartmentalized in specific nuclear regions in close proximity to the nucleolar chromatin (101), which is further supported by studies linking ZNF274 to the nucleolus (132). Furthermore, ZNF274, together with the ZNF143 protein (3), is a TAD boundary-associated genomic element (39), and ZNF143 in conjunction with THAP11 and HCF1 is enriched at the promoters of KRAB-ZFP genes, among other regions (86). In sum, these data indicate that a unique machinery regulates the chromatin state of KRAB-ZFP gene clusters, which may further contribute to their rapid evolution.

Murine leukemia virus (MuLV): retrovirus that can infect murine cells and promote cancer; endogenized in some mouse strains (class I ERV)

Primer binding site (PBS): structured element in retroviral genomes requiring binding by host tRNAs; serves as a primer for reverse transcription

Alu: primate-specific non-LTR-ERE belonging to the SINE family; highly active and represented in the human genome

BIOLOGICAL FUNCTIONS OF KRAB–ZINC FINGER PROTEINS

Control of Endogenous Retroelements

Several lines of experimental evidence demonstrate that EREs are the primary target sequences of most KRAB-ZFPs and are therefore the primary forces driving their amplification, diversification, and selection. First, KAP1 was biochemically purified as the essential factor required for the proviral silencing of murine leukemia viruses (MuLVs) in mouse embryonic stem cells (mESCs), and subsequently it was found that Zfp809, a mouse-specific KRAB-ZFP, was responsible for this targeting (127, 128). Second, genetic deletion of the universal KRAB-ZFP corepressors KAP1 or SETDB1 causes widespread activation of EREs in both cultured mESCs and mouse embryos (76, 103). Third, *Zfp809* deletion in mice fails to initiate the repression of a specific subset of LTR retrotransposons called VL30, which are related to MuLV by their use of the same Zfp809 target sequence, the primer binding site for proline transfer (t)RNA (PBS-pro) (129). Fourth, in human ESCs, about three-quarters of KAP1-binding sites reside within EREs, and KAP1 activity is essential for establishing the chromatin-repressive marks by SETDB1, which often coincide with DNA methylation (119). Fifth, large-scale ChIP-seq analyses of epitope-tagged KRAB-ZFPs in human cell lines have revealed that nearly two-thirds of human KRAB-ZFPs show preferential binding enrichment to specific families of EREs (42, 84). Those that bind to EREs (with a higher percentage of their peak regions overlapping EREs) tend to be younger, evolutionarily speaking, whereas older KRAB-ZFPs have a lower fraction of their binding sites overlapping EREs. It is nonetheless tempting to speculate that even these more ancient KRAB-ZFPs emerged in response to EREs, but these ERE targets have eroded beyond recognition as demonstrated by repeat masking software algorithms, leaving only small regulatory elements in their wake. Nevertheless, in sum these data provide compelling evidence that most KRAB-ZFPs bind and silence EREs.

To complement the experimental evidence, several evolutionary analyses strongly support the hypothesis that KRAB-ZFPs evolved to counteract EREs. For example, the number of C2H2 zinc finger genes correlates with the number of ERVs within each species of tetrapod (114). To illustrate this point, the mouse genome displays an impressive cohort of mouse-specific KRAB-ZFPs that likely counteracted a similarly large number of murine-specific active class II ERVs (72) (**Figure 4*a***). Furthermore, many KRAB-ZFPs share a similar evolutionary age as their target EREs, suggesting they emerge in response to the ERE infiltration (84). But the rapid appearance of new KRAB-ZFPs against invading EREs can also be counteracted by mutation of the target EREs within the KRAB-ZFP-binding site. This mutational tug-of-war between KRAB-ZFPs and EREs has been studied in great detail for SINE–variable number tandem repeat (VNTR)–Alu

Table 1 KRAB-ZFP loci identified as genetic modifiers

Phenotype	ERVs responsible	KRAB-ZFP responsible	KRAB-ZFP cluster location	Reference(s)
Dactylaplasia	MusD/ETn	Unknown (*Mdac* locus)	Chr 13	50
Cleft lip/palate	IAP	Unknown (*Clf2* locus)	Chr 13	49, 97
Lupus	NEERV	SNERV1 and SNERV2	Chr 13	116
Histone H3K4me3 *trans* QTLs in testes	NA	Unknown	Chr 4, 7, 12, and 13	4
Suppressor of *Escherichia coli* containing transgene expression	NA	Ssm1b	Chr 4	101a

Abbreviations: ERV, endogenous retrovirus; IAP, intracisternal A type particle; NA, not applicable; NEERV, non-ecotropic ERV; QTL, quantitative trait locus.

(SVA) elements and a subset of L1 retrotransposons, the L1PA elements, which in primates are repressed by the primate-specific ZNF91 and ZNF93, respectively (46). In particular, the gene encoding ZFP91 accumulated several mutations that enabled its binding to SVA elements, while the L1PA3 element subfamily has lost the binding site for ZFP93, thus evading its control (46). An additional KRAB-ZFP, ZNF649, also targeted L1PA family members, which then escaped silencing by the slow accumulation of point mutations in the ZNF649-binding site (32).

Despite the presence of KRAB-ZFPs that suppress EREs, EREs continue to propagate in mammalian genomes. In humans, ERVs are no longer active because of genetic drift, but LINEs and SINEs have retained the ability to retrotranspose, generating diversity in the soma and germline, contributing to phenotypic diversity, and in some cases causing human disease (54). In mice, ERVs, LINEs, and SINEs are still actively retrotransposing and therefore polymorphic in wild and inbred strains (85). Although most ERE variants that cause gene expression changes appear to be rapidly purged by purifying selection, ERE insertions have generated alleles and phenotypes in mice that can be suppressed in different strains by KRAB-ZFPs (**Table 1**). One example is the cleft lip palate phenotype caused by an intracisternal A type particle (IAP) element insertion downstream of the *Wnt9b* gene (49). In the A/WySn strain, this IAP element is transcribed, disrupting the expression of *Wnt9b*. However, in the C57Bl/6 strain, the IAP element is methylated, blocking the transcription of the IAP element and the disruption of *Wnt9b* expression. The locus responsible for this effect is *Clf2*, which maps to a cluster of KRAB-ZFP genes on Chr 13, suggesting that a specific KRAB-ZFP within the cluster likely recognizes this specific IAP element (97) (**Figure 3*b*,*c***). A second example relates to an allele of the *Fgf8* gene, which contains an insertion of a MusD element that causes a polydactyl phenotype in permissive strains (50). In these strains, the MusD element is expressed, which suppresses *Fgf8* expression in developing limb buds. However, in repressive strains, the MusD element is methylated and silenced, and the repressive factor, designated Mdac, again maps to a region of Chr 13 that contains numerous KRAB-ZFPs (50). Intriguingly, the same *Clf2* cluster of KRAB-ZFPs includes the Regulator of Sex-Limitation proteins 1 and 2 (Rsl1 and Rsl2), which in mouse are responsible for controlling sexually dimorphic liver gene expression by repression of male-specific liver genes (59) (**Figure 3*c***). Mouse females mutant for Rsl proteins express male-specific liver proteins, slightly affecting reproductive fitness (60). KRAB-ZFP genes within the Chr 13 cluster are similar to each other but are highly variable within different mouse strains, indicating this cluster is highly dynamic (58).

Despite the large body of evidence that most KRAB-ZFPs target EREs for transcriptional silencing via chromatin-based repressive mechanisms, it is much less clear whether loss of KRAB-ZFPs unleashes ERE activity and whether this increased activity would have dire consequences for the host. Deletion of the KRAB-ZFP corepressors KAP1 or SETDB1 leads to reactivation

Intracisternal A type particle (IAP): class II ERV; active in various mouse strains

MusD: class II ERV family element, which together with IAP elements constitutes the most active LTR-retrotransposons in mice; also called ETnERV

Table 2 KRAB-ZFP mouse mutant phenotypes

KRAB-ZFP (human gene/mouse gene)	Evolutionary clade	Mouse mutant phenotypes	Target genes/regions	Reference(s)
ZFP57/Zfp57	Eutheria	Partial neonatal lethal, loss of imprinting	Imprint control regions (via methylated TGCCGC)	62
ZNF568/Zfp568	Eutheria	Gastrulation lethal, partially rescued by *Igf2* deletion	Placental specific promoter of *Igf2* (*Igf2-P0*)	36a, 110a, 131
ZNF445/Zfp445	Theria	Embryo lethal (incomplete penetrance), synthetic lethality and loss of imprinting when combined with *Zfp57* mutant	Imprint control regions	112
ZNF274/Zfp110 (also called NRIF)	Eutheria	Midgestation lethal	3′ exon of zinc finger genes	18, 36
Zfp809 (no human ortholog)	Rodentia	No phenotype, VL30pro LTR elements reactivated in adult tissues	PBS-pro of several families of ERVs including VL30	128, 129

Abbreviations: ERV, endogenous retrovirus; LTR, long terminal repeat; NRIF, neurotrophin receptor interacting factor; PBS-pro, primer binding site for proline tRNA.

of EREs in mESCs and embryos and eventual cell/organismal death, but it is unclear whether ERE activation causes the lethal effect, as KAP1 and SETDB1 have functions independent of their recruitment by KRAB-ZFPs (76, 103). Thus, KRAB-ZFP loss-of-function studies of model organisms are needed. One good example is in *Zfp809* mutants (**Table 2**). Loss of *Zfp809* leads to widespread reactivation of the nonautonomous retrotransposon VL30 in all somatic tissues analyzed, but the knockout mice are born at Mendelian ratios and display no obvious phenotypes (129). In this example, several additional EREs are bound by Zfp809, whose expression is unaffected by *Zfp809* deletion, despite the fact that they lose H3K9me3 at the Zfp809-targeted PBS-pro within the ERVs. These ERVs, however, continue to maintain KAP1 recruitment and H3K9me3 at other positions along the targeted ERV genome, suggesting that redundant KRAB-ZFPs likely target these ERV families. Indeed, there are numerous examples of EREs that are bound throughout the length of the viral genome by multiple KRAB-ZFPs in humans (32, 42, 46). This indicates that it may be necessary to make deletions of several KRAB-ZFPs before phenotypes are discernible.

A second example is the recent characterization of two KRAB-ZFPs, SNERV1 and SNERV2 (suppressor of NEERV 1 and 2), which have been lost because of a natural deletion in several strains of laboratory mice, including C56BL/6N, NZB, and 129, but which are present in the C57BL/6J strain (116). In C57BL/6N mice, non-ecotropic endogenous retroviruses (NEERVs) envelope mRNA and gp70 protein, an autoantigen associated with lupus, and are highly activated in B cells and T cells. This activation is phenocopied upon deletion of SNERV1 and SNERV2 in C57BL/6J mice. Importantly, SNERV1 directly binds along with KAP1 to NEERV loci at a region overlapping the PBS-gln, suggesting that it directly suppresses NEERV. Furthermore, the NZB and 129 strains, which display elevated NEERV levels, fail to complement the *Snerv1/Snerv2* double mutant NEERV derepression phenotype. SNERV1 and SNERV2 also map to the previously described lupus susceptibility loci *Sgp3* and *Gv1*. Thus, the loss of these KRAB-ZFPs likely underlies the lupus phenotype in these mouse strains.

Non-ecotropic endogenous retrovirus (NEERV): ERV not restricted to a single or a small group of species

On the basis of the example of SNERV1 and SNERV2, it is tempting to speculate that the broader KRAB-ZFP family may serve an important role in preventing ERE-based activation of the innate immune system. Several reports have suggested that EREs and their activity could contribute to autoimmunity via the activation of nucleic acid sensors (for a review of this topic, see

Reference 122). Furthermore, epigenetic therapies that show promise as anticancer agents (including histone methyltransferase and demethylase inhibitors and DNA demethylating agents) work via activation of the innate and adaptive immune system. This occurs in part through activation of EREs, a process termed viral mimicry (48). Thus, KRAB-ZFPs likely play an important role in recognizing EREs for transcriptional silencing, such that adaptive and innate immune sensors are not inappropriately triggered.

A Cellular Immune System Against Exogenous Viruses?

LTR-containing EREs are the endogenous counterpart of exogenous viruses that managed to integrate into the germline of the host cell genome. Since KRAB-ZFPs are dynamic and efficient repressors of EREs, it is reasonable to suspect they might also repress exogenous viruses or perhaps even be naturally selected to bind exogenous viruses. This could allow the subsequent endogenization of such viruses and could represent a cellular adaptive immune system whose memory is encoded in the fingerprint amino acids of the zinc finger array (see sidebar titled Fighting the Exogenous and Endogenous Viral Threat: Comparisons of KRAB–Zinc Finger Protein Defense with Small RNA–Mediated Silencing Pathways). However, the evidence to date in support of this hypothesis is limited. KRAB-ZFPs can indeed bind and regulate expression of episomal viral DNA (6), and they have been engineered to repress HIV retroviral replication by targeting its LTR (102). One study also identified ZNF10, ZNF566, ZNF333, ZNF561, and ZNF324 as potential suppressors of HIV LTR activity using viral reporters and knockdown studies, but the binding sites were not precisely mapped biochemically, suggesting that the effects could be indirect (88). Similar lines of evidence using KRAB-ZFP knockdowns imply that the KRAB-ZFPs SZF1 and ZNF557 may play a role in Epstein–Barr virus latency, although again the binding sites of the KRAB-ZFPs were not determined, suggesting that the effects could be indirect (61).

The best evidence that KRAB-ZFPs restrict exogenous retroviruses comes from Zfp809. Zfp809 represses infectious MuLV transcription after integration into mESCs by binding to the PBS-pro sequence, and it can block transcription from DNA constructs of human T-cell

FIGHTING THE EXOGENOUS AND ENDOGENOUS VIRAL THREAT: COMPARISONS OF KRAB–ZINC FINGER PROTEIN DEFENSE WITH SMALL RNA–MEDIATED SILENCING PATHWAYS

Viruses represent a threat to the survival of all cells, from bacteria to multicellular organisms, so it is not surprising that organisms within each kingdom of life developed molecular pathways that function as an adaptive immune system. The key component to these systems is a means to learn from previous viral infections such that the host can promptly fight off similar future infections. In prokaryotes and archaea, the clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) proteins system represents an elegant adaptive defense mechanism to foreign nucleic acids, in which this memory is encoded by the actual nucleic acid sequences of the viral invaders that serve to produce guide RNAs targeting the Cas nucleases to cleave newly invading viral DNA (11, 75). Quite similarly, the metazoan germline PIWI-interacting RNAs (piRNAs) system utilizes transcripts originating from clusters of EREs, in order to either tether the cytoplasmic PIWI proteins to the target ERE transcripts for cleavage or target the nuclear PIWI proteins to the nascent EREs' RNAs and suppress their further transcription, promoting both H3K9 and DNA methylation (90). The vertebrate-restricted KRAB-ZFPs represent the first cellular adaptive immune system based on proteins that recognize the proviral DNA, in which the memory of retroviral infection is encoded in the fingerprint amino acids.

lymphotropic virus-1 (HTLV-1) in a PBS-pro-dependent manner (128). The targeting of the PBS-pro would seem to be a clever strategy for the host, as the PBS must base-pair with a cellular tRNA to prime reverse transcription, a necessary step in the viral life cycle. Thus, the virus cannot easily escape Zfp809 binding by mutation without effects on replicative fitness. However, Zfp809 is not the only KRAB-ZFP that targets a PBS. The recently identified KRAB-ZFP SNERV1 targets the LTR of a NEERV in a region that overlaps the PBS for glutamine (PBS-gln) (116). It is tempting to speculate that mice lacking SNERV1 would also be more sensitive to exogenous retrovirus infection (in retroviruses that utilize the PBS-gln), although this has not been directly tested.

Some viruses may have also exploited KRAB-ZFPs and/or KAP1 to precisely tune their transcriptional output. One such example is the Epstein–Barr virus, in which an enhancer element of its lytic origin of replication is recognized and bound by ZBRK1 (ZNF350), KAP1, and BRCA1, having a positive influence on viral replication efficiency (63). In addition, KAP1 can both positively and negatively regulate HIV activity independently of KRAB-ZFPs. On the one hand, KAP1 promotes viral transcription by promoter-proximal LTR tethering of the positive transcription elongation factor b (P-TEFb), promoting transcriptional elongation upon activation cues (77). On the other hand, HIV uses KAP1-dependent SUMOylation of CDK9 to promote latency and evade surveillance systems (69).

From Genome Defenders to Punctilious Transcription Factors

The activation of EREs can be harmful to cells, potentially inducing extensive DNA damage and inflammatory immune responses, and thus organisms have a need to suppress EREs. However, fine-tuned control of regulatory sequences present in retrotransposons can add more complexity to gene regulatory networks, potentially benefitting the host in a changing environment (34). Thus, it is not surprising that a large fraction of transcription factor–binding sites reside within EREs, and that ERE-derived regulatory sequences have rewired gene expression networks in many tissues in humans and other mammals (for recent reviews of this topic, see 20, 115). These findings lead to the question whether the presence of KRAB-ZFPs actually facilitates the domestication of particular ERE-derived regulatory sequences. It is clear that deletion of KAP1 and individual KRAB-ZFPs results in the unleashing of ERE-derived enhancer and promoter activity that affects the expression of nearby genes (104, 129). However, it is less clear whether KRAB-ZFPs are permanent clamps on ERE activity or are perhaps regulated in a tissue- or stage-specific manner to allow EREs to serve as dynamic enhancers. Several lines of evidence suggest that the latter scenario is likely. KRAB-ZFPs and retrotransposons display dynamic expression patterns throughout development and in differentiating cells and adult tissues, suggesting that they might have cell-type-specific regulatory functions (26, 29, 40, 70). Furthermore, KAP1 genomic occupancy, which directly reflects the KRAB-ZFP pool expressed in that tissue, also differs between different cell types. There are also clear examples in which KRAB-ZFP-binding sites within EREs are under purifying selection (42). Additionally, many KRAB-ZFPs remain in the genome long after their target EREs accumulated mutations that limit their ability to retrotranspose, suggesting they have adopted new functions. This is also reflected in the finding that older KRAB-ZFPs are much more likely to occupy gene promoter sequences than EREs (42). The best recent example comes from studies of TE-born enhancers, such as those found within SVA elements, that are suppressed by KRAB-ZFPs in primed human ESCs but are active during embryonic genome activation and in several adult tissues including the brain (98). In the next section, we highlight how several members of the KRAB-ZFP family, which may have originated as ERE silencing factors in our mammalian ancestors, became embedded as factors responsible for genomic imprinting.

Genomic Imprinting in Mammals Requires KRAB–Zinc Finger Proteins

Genomic imprinting is an epigenetic process whereby alleles of a gene are expressed in a parent-of-origin-specific manner. Genomic imprinting has been demonstrated in plants, fungi, and animals, but in the mammalian lineage, its roots can be traced back to the common ancestors of marsupials and eutherian mammals. Genomic imprinting is thought to have arisen as a result of the intragenomic evolutionary conflict between maternally and paternally inherited alleles brought about by development of an embryo in utero (for a recent review of the process and evolutionary theory of imprinting, see Reference 118). In general, maternally imprinted genes tend to be genes involved in promoting growth, like the insulin-like growth factor 2 (*Igf2*) gene, a key growth promoter in embryo development, whereas paternally imprinted genes tend to be growth restricting, like the insulin-like growth factor 2 receptor (*Igf2r*), which serves as a sink for IGF2 peptide. The key mechanistic basis of imprinting is the specific epigenetic marking of alleles at imprint control regions (ICRs) in either the male (for paternally imprinted genes) or the female (for maternally imprinted genes) germline. These imprints must then be maintained in the offspring during the period of extensive epigenetic reprogramming in early development. At mammalian imprinted genes, the ICRs are differentially methylated on each allele and are therefore termed differentially methylated regions. The KRAB-ZFP ZFP57 is essential for maintaining genomic imprinting by binding specifically to a methylated hexanucleotide DNA sequence found at ICRs and tethering KAP1 to these regions (65, 66, 100) (**Table 2**). This in turn recruits histone and DNA methyltransferases essential for establishing methylation of the ICRs (133). While *Zfp57* deletion in zygotes leads to partial perinatal and neonatal lethality in mice, homozygous maternal mutation leads to embryonic lethality (62). Mutations of human *ZFP57* have been linked instead to transient neonatal diabetes (13, 71). The incomplete penetrance of the phenotypes linked with *Zfp57* loss-of-function suggested that additional proteins might also regulate imprinting. Indeed, ZNF445 along with ZFP57 interacts with ICRs and is essential for establishing and maintaining proper imprints in both mouse and human, although its precise binding activity has not been determined (112) (**Table 2**). Whereas ZFP57 can be traced to eutherian mammals, ZNF445 can be traced further back to marsupials, suggesting that it was likely the initial factor responsible for the evolution of imprinting. It is tempting to speculate that ZFP57 and ZNF445 may have initially evolved to silence EREs and that EREs themselves may have evolved into ICRs. Indeed, the expression of EREs in the germline is directly coupled to the methylation of ICRs (16).

The *Igf2* gene is a highly conserved growth factor in chordates. In eutherian mammals *Igf2* is imprinted in a manner dependent on ZFP57 and ZNF445. However, these are not the only KRAB-ZFPs that evolved to regulate *Igf2*. In mammals ZFP568 can also trace its roots to eutherian mammals (92), and it binds with high selectivity to a sequence motif upstream of a placental promoter of the *Igf2* gene (designated *Igf2-P0* in mice) (131). Loss of *Zfp568* in mice leads to gastrulation failure, and this lethal phenotype is associated with premature activation of the Igf2-P0 promoter and Igf2 peptide in early embryonic cells (**Table 2**). Despite the fact that Zfp568 binds with high affinity to approximately 100 sites in the genome, the *Igf2-P0*-binding site is one of only a few that are highly conserved in mammals, and deletion of *Igf2* alone can restore the viability of *Zfp568* knockout embryos, suggesting that this specific binding site is the only critical binding site for survival through embryogenesis. Thus, the appearance of *Zfp568* likely facilitated the domestication of the *Igf2-P0* promoter sequence for use in the placenta, by providing a means to suppress its activity, which would otherwise be toxic, in early development. Although not essential for viability in the mouse, the *Igf2-P0* promoter accounts for expression of *Igf2* specifically in the labyrinthine trophoblast of the placenta, and deletion of *Igf2-P0* transcript leads to placental growth restriction, fetal growth restriction, and alteration in placental weight ratios, suggesting it plays an important role in regulating maternal supply and fetal demand for nutrients in mammals (23).

ZFP568, ZFP57, and ZNF445 are just three of several dozen KRAB-ZFPs that emerged in the earliest mammals (**Figure 4*a***) and that have been maintained in the genome under purifying selection of their zinc finger arrays (and specifically their fingerprint amino acids), suggesting that their DNA-binding domains are conserved. These unexplored KRAB-ZFPs provide fertile ground to identify other likely essential mammalian adaptations in addition to imprinting and *Igf2* regulation. One additional example of an essential KRAB-ZFP that emerged in early mammals is ZNF274 (**Figure 4*a*,*b***). This factor binds specifically to the 3′ end of zinc finger genes via the binding of a recurrent motif present on many, but not all, of the zinc finger satellite sequences (36). The *ZNF274* ortholog in mice, *Zfp110* [also called neurotrophin receptor interacting factor (NRIF)], is essential in C57BL/6J mice (**Table 2**), although the cause of the embryonic lethal phenotype is unknown (18). One possibility is that ZNF274 protects the repetitive zinc finger satellite from recombination, DNA damage, or both, as was indicated by the increased levels of the DNA damage marker γH2A.X in cultured human cells lacking ZNF274 (121). It is tempting to speculate that such a factor may have facilitated the expansion of the KRAB-ZFP family in early mammals in response to ERV infiltration, which may have been necessary in mammals for the evolution of placenta, as retroviral envelope captures have been linked to placentation in multiple lineages (25, 28).

KRAB–Zinc Finger Proteins and Endogenous Retroelements Shape the Meiotic Recombination Landscape

KRAB-ZFPs, including the ancestor of all KRAB-ZFPs, PR/SET domain 9 protein (PRDM9), play a major role in shaping the meiotic recombination landscape, which is likely influenced by EREs. PRDM9 is a unique KRAB-ZFP in several respects. It is expressed specifically in meiotic cells as they initiate meiosis (38), where it determines meiotic recombination hotspots in a wide array of vertebrates (5, 7, 81). Although its KRAB domain is essential for its function (41), PRDM9 does not interact with KAP1 and does not promote heterochromatin at its binding sites. In contrast, PRDM9 uses its PR/SET domain to methylate histone H3K4 and H3K36 (38, 99), creating a dual H3K4me3/H3K36me3 mark that leads to the recruitment of the meiotic double strand break (DSB) machinery. In species that utilize PRDM9 to determine meiotic DSB hotspots, the PRDM9 zinc finger array is rapidly evolving by homogenization of its zinc finger array (89, 113). This appears to be necessary to overcome the eventual loss of PRDM9-binding sites over evolutionary timescales as a result of gene conversion associated with DSB repair (24). Thus, multiple alleles of PRDM9 are usually present within a population, with older alleles eventually replaced by younger alleles that change the locations of hotspots. PRDM9-controlled hotspots in mice and humans are enriched with EREs (15, 81), suggesting that EREs play an important role in determining hotspots, leading to the hypothesis that PRDM9 itself may have evolved as a type of defense against EREs. However, activation of DSBs within EREs might also pose a threat of nonallelic homologous recombination. Thus, it is an intriguing hypothesis that the broader KRAB-ZFP family, which binds primarily to EREs to promote H3K9me3, might interfere with PRDM9 activity and therefore shape the meiotic DSB landscape. This hypothesis has been supported by two crucial lines of evidence. First, in humans, there is an enrichment of KRAB-ZFP binding motifs near PRDM9-binding sites that fail to become hotspots (1). Second, in mice, histone quantitative trait loci that control PRDM9-dependent H3K4me3 levels in testes have been mapped to several of the major KRAB-ZFP clusters that have rapidly evolved in mouse strains (4). Both lines of evidence suggest that KRAB-ZFP-initiated H3K9me3 at EREs can influence the binding and methyltransferase activity of PRDM9, likely preventing DSB formation at newly acquired EREs. A direct test of this hypothesis would require analyzing DSBs in KRAB-ZFP knockout meiotic cells, where novel ERE hotspots might be revealed.

FUTURE CHALLENGES

In the last decade, several studies aimed to extensively characterize the binding sites of many KRAB-ZFPs (42, 106), leading to the primary hypothesis that KRAB-ZFP amplification is being driven by the need to recognize and silence EREs. Additional studies have demonstrated that more ancient KRAB-ZFPs such as ZFP57, ZNF445, and ZFP568, which emerged over 100 million years ago, were essential to mammalian development and evolution. Yet there are still many open questions regarding the evolution and function of this large protein family, and several barriers that must be overcome to answer them.

One of the biggest challenges relates to the discovery and annotation of KRAB-ZFPs. We still do not know how many genes there are, nor do we know what fraction are truly expressed and encode proteins. This is because of the high repetitiveness of their coding sequences on two scales. First, the coding sequences of zinc finger arrays are minisatellites [in the mouse genome most zinc finger arrays overlap the satellite MMSAT4 in RepeatMasker (53)], making errors in their annotation likely (2). Second, KRAB-ZFP genes are present in large clusters (up to several megabases) that share high homology with each other owing to the recent segmental duplications. The rapid evolution of these genes further increases the complexity of the annotation problem. Together with EREs, KRAB-ZFP clusters exhibit heterogeneity between individuals within the same species, as shown in studies of both human and mouse (2, 53, 64).

A second challenge that must be overcome to probe the cellular and molecular functions of KRAB-ZFPs is the lack of KRAB-ZFP mutants and antibodies. The clustering and high sequence similarity of most KRAB-ZFP genes make them refractory to conventional and even CRISPR/Cas9-based gene targeting in mouse and human cell culture models. It is possible to overcome these limitations by making large genome KRAB-ZFP cluster deletions. The lack of antibodies for most members of the KRAB-ZFP family has further contributed to the difficulties in assessing their cellular functions, and this may be impossible to overcome because many KRAB-ZFPs are so similar at the amino acid level. Thus, it will likely require painstaking efforts to genetically tag or generate specific antibodies for each member.

A final series of important questions related to how KRAB-ZFPs evolve also remain unanswered. What are the molecular gears that facilitate the evolution of KRAB-ZFPs? Precisely how do KRAB-ZFPs evolve in response to EREs? Does the intrinsic repetitiveness of KRAB-ZFP genes suffice to produce new KRAB-ZFPs for natural selection to work on, or are there specific factors that facilitate KRAB-ZFP gene evolution (52)? Are there mechanisms preventing uncontrolled expansion of KRAB-ZFP genes? Addressing these questions will not only further our understanding of KRAB-ZFPs but also give us better insights into the evolution and development of species-specific gene regulatory networks.

SUMMARY POINTS

1. Tandem C2H2 zinc finger proteins are ancient DNA-binding proteins found in green plants, fungi, and metazoans. Krüppel-associated box domain zinc finger proteins (KRAB-ZFPs) arose in the common ancestor of coelacanths, birds, reptiles, and mammals and have rapidly expanded and diversified.
2. Zinc finger proteins use tandem arrays of C2H2 zinc fingers primarily to achieve high-affinity DNA binding via the association of fingerprint amino acids within each finger with three nucleotides of target DNA. However, zinc finger arrays are also flexible in their mode of binding to allow reading of methylation marks and accommodate target DNA shape alterations.

3. Most KRAB-ZFPs bind specifically to endogenous retroelements (EREs) and participate in their transcriptional silencing via the recruitment of the corepressor KAP1 and associated repressive histone-modifying enzymes such as histone deacetylases and the H3K9 histone methyltransferase SETDB1.
4. KRAB-ZFPs emerge and decay coincidentally with their target EREs, suggesting that ERE invasions and their eventual genetic drift are the primary drivers of KRAB-ZFP evolution and turnover.
5. Several ancient mammal-specific KRAB-ZFPs have been maintained by purifying selection, and these KRAB-ZFPs play important roles in unique mammalian biological phenomena, including genomic imprinting and placental gene regulation.
6. KRAB-ZFP genes are among the most rapidly evolving genes in laboratory and wild mice and are likely distinct in different human individuals. Precise annotation of KRAB-ZFP genes will likely be realized with long-range sequencing technologies that can sequence through large KRAB-ZFP gene clusters, which currently contain gaps.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We would like to acknowledge Sienna Schaeffer and Anna Dorothea Senft for providing feedback on the review. T.S.M., M.M., and M.B. are supported by grant 1ZIAHD008933-01 of the NIH Intramural Research Program through the Division of Intramural Research at the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development.

LITERATURE CITED

1. Altemose N, Noor N, Bitoun E, Tumian A, Imbeault M, et al. 2017. A map of human PRDM9 binding provides evidence for novel behaviors of PRDM9 and other zinc-finger proteins in meiosis. *eLife* 6:e28383
2. Audano PA, Sulovari A, Graves-Lindsay TA, Cantsilieris S, Sorensen M, et al. 2019. Characterizing the major structural variant alleles of the human genome. *Cell* 176:663–75.e19
3. Bailey SD, Zhang X, Desai K, Aid M, Corradin O, et al. 2015. ZNF143 provides sequence specificity to secure chromatin interactions at gene promoters. *Nat. Commun.* 2:6186
4. Baker CL, Walker M, Arat S, Ananda G, Petkova P, et al. 2019. Tissue-specific *trans* regulation of the mouse epigenome. *Genetics* 211:831–45
5. Baker Z, Schumer M, Haba Y, Bashkirova L, Holland C, et al. 2017. Repeated losses of PRDM9-directed recombination despite the conservation of PRDM9 across vertebrates. *eLife* 6:e24133
6. Barde I, Laurenti E, Verp S, Groner AC, Towne C, et al. 2009. Regulation of episomal gene expression by KRAB/KAP1-mediated histone modifications. *J. Virol.* 83:5574–80
7. Baudat F, Buard J, Grey C, Fledel-Alon A, Ober C, et al. 2010. PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Science* 327:836–40
8. Bellefroid EJ, Lecocq PJ, Benhida A, Poncelet DA, Belayew A, Martial JA. 1989. The human genome contains hundreds of genes coding for finger proteins of the Krüppel type. *DNA* 8:377–87

9. Bellefroid EJ, Marine JC, Ried T, Lecocq PJ, Riviere M, et al. 1993. Clustered organization of homologous KRAB zinc-finger genes with enhanced expression in human T lymphoid cells. *EMBO J.* 12:1363–74
10. Bellefroid EJ, Poncelet DA, Lecocq PJ, Revelant O, Martial JA. 1991. The evolutionarily conserved Krüppel-associated box domain defines a subfamily of eukaryotic multifingered proteins. *PNAS* 88:3608–12
11. Bhaya D, Davison M, Barrangou R. 2011. CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annu. Rev. Genet.* 45:273–97
12. Birtle Z, Ponting CP. 2006. Meisetz and the birth of the KRAB motif. *Bioinformatics* 22:2841–45
13. Boonen SE, Mackay DJ, Hahnemann JM, Docherty L, Gronskov K, et al. 2013. Transient neonatal diabetes, ZFP57, and hypomethylation of multiple imprinted loci: a detailed follow-up. *Diabetes Care* 36:505–12
14. Brayer KJ, Segal DJ. 2008. Keep your fingers off my DNA: protein-protein interactions mediated by C2H2 zinc finger domains. *Cell Biochem. Biophys.* 50:111–31
15. Brick K, Smagulova F, Khil P, Camerini-Otero RD, Petukhova GV. 2012. Genetic recombination is directed away from functional genomic elements in mice. *Nature* 485:642–45
16. Brind'Amour J, Kobayashi H, Richard Albert J, Shirane K, Sakashita A, et al. 2018. LTR retrotransposons transcribed in oocytes drive species-specific and heritable changes in DNA methylation. *Nat. Commun.* 9:3331
17. Brown RS. 2005. Zinc finger proteins: getting a grip on RNA. *Curr. Opin. Struct. Biol.* 15:94–98
18. Casademunt E, Carter BD, Benzel I, Frade JM, Dechant G, Barde YA. 1999. The zinc finger protein NRIF interacts with the neurotrophin receptor p75(NTR) and participates in programmed cell death. *EMBO J.* 18:6050–61
19. Choo Y, Klug A. 1993. A role in DNA binding for the linker sequences of the first three zinc fingers of TFIIIA. *Nucleic Acids Res.* 21:3341–46
20. Chuong EB, Elde NC, Feschotte C. 2017. Regulatory activities of transposable elements: from conflicts to benefits. *Nat. Rev. Genet.* 18:71–86
21. Collins T, Stone JR, Williams AJ. 2001. All in the family: the BTB/POZ, KRAB, and SCAN domains. *Mol. Cell. Biol.* 21:3609–15
22. Coluccio A, Ecco G, Duc J, Offner S, Turelli P, Trono D. 2018. Individual retrotransposon integrants are differentially controlled by KZFP/KAP1-dependent histone methylation, DNA methylation and TET-mediated hydroxymethylation in naive embryonic stem cells. *Epigenet. Chromatin* 11:7
23. Constancia M, Hemberger M, Hughes J, Dean W, Ferguson-Smith A, et al. 2002. Placental-specific IGF-II is a major modulator of placental and fetal growth. *Nature* 417:945–48
24. Coop G, Myers SR. 2007. Live hot, die young: transmission distortion in recombination hotspots. *PLOS Genet.* 3:e35
25. Cornelis G, Funk M, Vernochet C, Leal F, Tarazona OA, et al. 2017. An endogenous retroviral envelope syncytin and its cognate receptor identified in the viviparous placental *Mabuya* lizard. *PNAS* 114:E10991–1000
26. Corsinotti A, Kapopoulou A, Gubelmann C, Imbeault M, Santoni de Sio FR, et al. 2013. Global and stage specific patterns of Krüppel-associated-box zinc finger protein gene expression in murine early embryonic cells. *PLOS ONE* 8:e56721
27. Deniz Ö, de la Rica L, Cheng KCL, Spensberger D, Branco MR. 2018. SETDB1 prevents TET2-dependent activation of IAP retroelements in naïve embryonic stem cells. *Genome Biol.* 19:6
28. Dupressoir A, Lavialle C, Heidmann T. 2012. From ancestral infectious retroviruses to bona fide cellular genes: role of the captured syncytins in placentation. *Placenta* 33:663–71
29. Ecco G, Cassano M, Kauzlaric A, Duc J, Coluccio A, et al. 2016. Transposable elements and their KRAB-ZFP controllers regulate gene expression in adult tissues. *Dev. Cell* 36:611–23
30. Emerson RO, Thomas JH. 2009. Adaptive evolution in zinc finger transcription factors. *PLOS Genet.* 5:e1000325
31. Emerson RO, Thomas JH. 2011. Gypsy and the birth of the SCAN domain. *J. Virol.* 85:12043–52

32. Fernandes JD, Haeussler M, Armstrong J, Tigyi K, Gu J, et al. 2018. KRAB zinc finger proteins coordinate across evolutionary time scales to battle retroelements. bioRxiv 429563. **https://doi.org/10.1101/429563**
33. Fonti G, Marcaida M, Bryan L, Traeger S, Kalantzi A, et al. 2019. KAP1 is an antiparallel dimer with a natively functional asymmetry. bioRxiv 553511. **https://doi.org/10.1101/553511**
34. Friedli M, Trono D. 2015. The developmental control of transposable elements and the evolution of higher species. *Annu. Rev. Cell Dev. Biol.* 31:429–51
35. Friedman JR, Fredericks WJ, Jensen DE, Speicher DW, Huang XP, et al. 1996. KAP-1, a novel corepressor for the highly conserved KRAB repression domain. *Genes Dev*. 10:2067–78
36. Frietze S, O'Geen H, Blahnik KR, Jin VX, Farnham PJ. 2010. ZNF274 recruits the histone methyltransferase SETDB1 to the 3′ ends of ZNF genes. *PLOS ONE* 5:e15082
36a. García-García MJ, Shibata M, Anderson KV. 2008. Chato, a KRAB zinc-finger protein, regulates convergent extension in the mouse embryo. *Development* 135(18):3053–62
37. Hahn MA, Wu X, Li AX, Hahn T, Pfeifer GP. 2011. Relationship between gene body DNA methylation and intragenic H3K9me3 and H3K36me3 chromatin marks. *PLOS ONE* 6:e18844
38. Hayashi K, Yoshida K, Matsui Y. 2005. A histone H3 methyltransferase controls epigenetic events required for meiotic prophase. *Nature* 438:374–78
39. Hong S, Kim D. 2017. Computational characterization of chromatin domain boundary-associated genomic elements. *Nucleic Acids Res*. 45:10403–14
40. Huntley S, Baggott DM, Hamilton AT, Tran-Gyamfi M, Yang S, et al. 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*. 16:669–77
41. Imai Y, Baudat F, Taillepierre M, Stanzione M, Toth A, de Massy B. 2017. The PRDM9 KRAB domain is required for meiosis and involved in protein interactions. *Chromosoma* 126:681–95
42. Imbeault M, Helleboid PY, Trono D. 2017. KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. *Nature* 543:550–54
43. Ivanov AV, Peng H, Yurchenko V, Yap KL, Negorev DG, et al. 2007. PHD domain-mediated E3 ligase activity directs intramolecular sumoylation of an adjacent bromodomain required for gene silencing. *Mol. Cell* 28:823–37
44. Ivanov D, Stone JR, Maki JL, Collins T, Wagner G. 2005. Mammalian SCAN domain dimer is a domain-swapped homolog of the HIV capsid C-terminal domain. *Mol. Cell* 17:137–43
45. Iyengar S, Ivanov AV, Jin VX, Rauscher FJ 3rd, Farnham PJ. 2011. Functional analysis of KAP1 genomic recruitment. *Mol. Cell. Biol.* 31:1833–47
46. Jacobs FM, Greenberg D, Nguyen N, Haeussler M, Ewing AD, et al. 2014. An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. *Nature* 516:242–45
47. Johnson WE, Coffin JM. 1999. Constructing primate phylogenies from ancient retrovirus sequences. *PNAS* 96:10254–60
48. Jones PA, Ohtani H, Chakravarthy A, De Carvalho DD. 2019. Epigenetic therapy in immune-oncology. *Nat. Rev. Cancer* 19:151–61
49. Juriloff DM, Harris MJ, Dewell SL, Brown CJ, Mager DL, et al. 2005. Investigations of the genomic region that contains the *clf1* mutation, a causal gene in multifactorial cleft lip and palate in mice. *Birth Defects Res. A Clin. Mol. Teratol.* 73:103–13
50. Kano H, Kurahashi H, Toda T. 2007. Genetically regulated epigenetic transcriptional activation of retrotransposon insertion confers mouse dactylaplasia phenotype. *PNAS* 104:19034–39
51. Kapitonov VV, Jurka J. 2008. A universal classification of eukaryotic transposable elements implemented in Repbase. *Nat. Rev. Genet.* 9:411–12
52. Kapopoulou A, Mathew L, Wong A, Trono D, Jensen JD. 2016. The evolution of gene expression and binding specificity of the largest transcription factor family in primates. *Evolution* 70:167–80
53. Kauzlaric A, Ecco G, Cassano M, Duc J, Imbeault M, Trono D. 2017. The mouse genome displays highly dynamic populations of KRAB-zinc finger protein genes and related genetic units. *PLOS ONE* 12:e0173746
54. Kazazian HH Jr., Moran JV. 2017. Mobile DNA in health and disease. *N. Engl. J. Med.* 377:361–70

55. Kim MS, Kini AG. 2017. Engineering and application of zinc finger proteins and TALEs for biomedical research. *Mol. Cells* 40:533–41
56. Kim SS, Chen YM, O'Leary E, Witzgall R, Vidal M, Bonventre JV. 1996. A novel member of the RING finger family, KRIP-1, associates with the KRAB-A transcriptional repressor domain of zinc finger proteins. *PNAS* 93:15299–304
57. Klug A. 2010. The discovery of zinc fingers and their applications in gene regulation and genome manipulation. *Annu. Rev. Biochem.* 79:213–31
58. Krebs CJ, Larkins LK, Khan SM, Robins DM. 2005. Expansion and diversification of KRAB zinc-finger genes within a cluster including *Regulator of sex-limitation 1* and *2*. *Genomics* 85:752–61
59. Krebs CJ, Larkins LK, Price R, Tullis KM, Miller RD, Robins DM. 2003. *Regulator of sex-limitation* (*Rsl*) encodes a pair of KRAB zinc-finger genes that control sexually dimorphic liver gene expression. *Genes Dev*. 17:2664–74
60. Krebs CJ, Robins DM. 2010. A pair of mouse KRAB zinc finger proteins modulates multiple indicators of female reproduction. *Biol. Reprod.* 82:662–68
60a. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al. 2001. Initial sequencing and analysis of the human genome. Nature 409(6822):860–921. Erratum. 2001. *Nature* 412:565–66
61. Li X, Burton EM, Koganti S, Zhi J, Doyle F, et al. 2018. KRAB-ZFP repressors enforce quiescence of oncogenic human herpesviruses. *J. Virol.* 92:e00298-18
62. Li X, Ito M, Zhou F, Youngson N, Zuo X, et al. 2008. A maternal-zygotic effect gene, *Zfp57*, maintains both maternal and paternal imprints. *Dev. Cell* 15:547–57
63. Liao G, Huang J, Fixman ED, Hayward SD. 2005. The Epstein-Barr virus replication protein BBLF2/3 provides an origin-tethering function through interaction with the zinc finger DNA binding protein ZBRK1 and the KAP-1 corepressor. *J. Virol.* 79:245–56
64. Lilue J, Doran AG, Fiddes IT, Abrudan M, Armstrong J, et al. 2018. Sixteen diverse laboratory mouse reference genomes define strain-specific haplotypes and novel functional loci. *Nat. Genet.* 50:1574–83
65. Liu Y, Olanrewaju YO, Zhang X, Cheng X. 2013. DNA recognition of 5-carboxylcytosine by a Zfp57 mutant at an atomic resolution of 0.97 Å. *Biochemistry* 52:9310–17
66. Liu Y, Toh H, Sasaki H, Zhang X, Cheng X. 2012. An atomic model of Zfp57 recognition of CpG methylation within a specific DNA sequence. *Genes Dev*. 26:2374–79
67. Looman C, Hellman L, Abrink M. 2004. A novel Krüppel-Associated Box identified in a panel of mammalian zinc finger proteins. *Mamm. Genome* 15:35–40
68. Lukic S, Nicolas J-C, Levine AJ. 2014. The diversity of zinc-finger genes on human chromosome 19 provides an evolutionary mechanism for defense against inherited endogenous retroviruses. *Cell Death Differ*. 21:381–87
69. Ma X, Yang T, Luo Y, Wu L, Jiang Y, et al. 2019. TRIM28 promotes HIV-1 latency by SUMOylating CDK9 and inhibiting P-TEFb. *eLife* 8:e42426
70. Macfarlan TS, Gifford WD, Driscoll S, Lettieri K, Rowe HM, et al. 2012. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* 487:57–63
71. Mackay DJ, Callaway JL, Marks SM, White HE, Acerini CL, et al. 2008. Hypomethylation of multiple imprinted loci in individuals with transient neonatal diabetes is associated with mutations in ZFP57. *Nat. Genet.* 40:949–51
72. Mager DL, Stoye JP. 2015. Mammalian endogenous retroviruses. *Microbiol. Spectr.* 3:MDNA3-0009-2014
73. Margolin JF, Friedman JR, Meyer WK, Vissing H, Thiesen HJ, Rauscher FJ 3rd. 1994. Krüppel-associated boxes are potent transcriptional repression domains. *PNAS* 91:4509–13
74. Mark C, Abrink M, Hellman L. 1999. Comparative analysis of KRAB zinc finger proteins in rodents and man: evidence for several evolutionarily distinct subfamilies of KRAB zinc finger genes. *DNA Cell Biol*. 18:381–96
75. Marraffini LA, Sontheimer EJ. 2010. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat. Rev. Genet.* 11:181–90
76. Matsui T, Leung D, Miyashita H, Maksakova IA, Miyachi H, et al. 2010. Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature* 464:927–31

77. McNamara RP, Reeder JE, McMillan EA, Bacon CW, McCann JL, D'Orso I. 2016. KAP1 recruitment of the 7SK snRNP complex to promoters enables transcription elongation by RNA polymerase II. *Mol. Cell* 61:39–53
78. Miller J, McLachlan AD, Klug A. 1985. Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *EMBO J.* 4:1609–14
79. Molparia B, Goyal K, Sarkar A, Kumar S, Sundar D. 2010. ZiF-Predict: a web tool for predicting DNA-binding specificity in C2H2 zinc finger proteins. *Genom. Proteom. Bioinform.* 8:122–26
80. Moosmann P, Georgiev O, Le Douarin B, Bourquin JP, Schaffner W. 1996. Transcriptional repression by RING finger protein TIF1β that interacts with the KRAB repressor domain of KOX1. *Nucleic Acids Res.* 24:4859–67
81. Myers S, Bowden R, Tumian A, Bontrop RE, Freeman C, et al. 2010. Drive against hotspot motifs in primates implicates the *PRDM9* gene in meiotic recombination. *Science* 327:876–79
82. Najafabadi HS, Albu M, Hughes TR. 2015. Identification of C2H2-ZF binding preferences from ChIP-seq data using RCADE. *Bioinformatics* 31:2879–81
83. Najafabadi HS, Garton M, Weirauch MT, Mnaimneh S, Yang A, et al. 2017. Non-base-contacting residues enable kaleidoscopic evolution of metazoan C2H2 zinc finger DNA binding. *Genome Biol.* 18:167
84. Najafabadi HS, Mnaimneh S, Schmitges FW, Garton M, Lam KN, et al. 2015. C2H2 zinc finger proteins greatly expand the human regulatory lexicon. *Nat. Biotechnol.* 33:555–62
85. Nellaker C, Keane TM, Yalcin B, Wong K, Agam A, et al. 2012. The genomic landscape shaped by selection on transposable elements across 18 mouse strains. *Genome Biol.* 13:R45
86. Ngondo-Mbongo RP, Myslinski E, Aster JC, Carbon P. 2013. Modulation of gene expression via overlapping binding sites exerted by ZNF143, Notch1 and THAP11. *Nucleic Acids Res.* 41:4000–14
87. Nielsen AL, Ortiz JA, You J, Oulad-Abdelghani M, Khechumian R, et al. 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.* 18:6385–95
88. Nishitsuji H, Sawada L, Sugiyama R, Takaku H. 2015. ZNF10 inhibits HIV-1 LTR activity through interaction with NF-κB and Sp1 binding motifs. *FEBS Lett.* 589:2019–25
89. Oliver PL, Goodstadt L, Bayes JJ, Birtle Z, Roach KC, et al. 2009. Accelerated evolution of the *Prdm9* speciation gene across diverse metazoan taxa. *PLOS Genet.* 5:e1000753
90. Ozata DM, Gainetdinov I, Zoch A, O'Carroll D, Zamore PD. 2019. PIWI-interacting RNAs: small RNAs with big functions. *Nat. Rev. Genet.* 20:89–108
91. Pastuzyn ED, Day CE, Kearns RB, Kyrke-Smith M, Taibi AV, et al. 2018. The neuronal gene *Arc* encodes a repurposed retrotransposon Gag protein that mediates intercellular RNA transfer. *Cell* 172:275–88.e18
92. Patel A, Yang P, Tinkham M, Pradhan M, Sun MA, et al. 2018. DNA conformation induces adaptable binding by tandem zinc finger proteins. *Cell* 173:221–33.e12
93. Pavletich NP, Pabo CO. 1991. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* 252:809–17
94. Pavletich NP, Pabo CO. 1993. Crystal structure of a five-finger GLI-DNA complex: new perspectives on zinc fingers. *Science* 261:1701–7
95. Peng H, Begg GE, Harper SL, Friedman JR, Speicher DW, Rauscher FJ 3rd. 2000. Biochemical analysis of the Krüppel-associated box (KRAB) transcriptional repression domain. *J. Biol. Chem.* 275:18000–10
96. Pengue G, Calabro V, Bartoli PC, Pagliuca A, Lania L. 1994. Repression of transcriptional activity at a distance by the evolutionarily conserved KRAB domain present in a subfamily of zinc finger proteins. *Nucleic Acids Res.* 22:2908–14
97. Plamondon JA, Harris MJ, Mager DL, Gagnier L, Juriloff DM. 2011. The *clf2* gene has an epigenetic role in the multifactorial etiology of cleft lip and palate in the A/WySn mouse strain. *Birth Defects Res. A Clin. Mol. Teratol.* 91:716–27
98. Pontis J, Planet E, Offner S, Turelli P, Duc J, et al. 2019. Hominoid-specific transposable elements and KZFPs facilitate human embryonic genome activation and control transcription in naive human ESCs. *Cell Stem Cell* 24:724–35

99. Powers NR, Parvanov ED, Baker CL, Walker M, Petkov PM, Paigen K. 2016. The meiotic recombination activator PRDM9 trimethylates both H3K36 and H3K4 at recombination hotspots in vivo. *PLOS Genet*. 12:e1006146
100. Quenneville S, Verde G, Corsinotti A, Kapopoulou A, Jakobsson J, et al. 2011. In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. *Mol. Cell* 44:361–72
101. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, et al. 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159:1665–80
101a. Ratnam S, Engler P, Bozek G, Mao L, Podlutsky A, et al. 2014. Identification of Ssm1b, a novel modifier of DNA methylation, and its expression during mouse embryogenesis. Development 141(10):2024–34
102. Reynolds L, Ullman C, Moore M, Isalan M, West MJ, et al. 2003. Repression of the HIV-1 5′ LTR promoter and inhibition of HIV-1 replication by using engineered zinc-finger transcription factors. *PNAS* 100:1615–20
103. Rowe HM, Jakobsson J, Mesnard D, Rougemont J, Reynard S, et al. 2010. KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* 463:237–40
104. Rowe HM, Kapopoulou A, Corsinotti A, Fasching L, Macfarlan TS, et al. 2013. TRIM28 repression of retrotransposon-based enhancers is necessary to preserve transcriptional dynamics in embryonic stem cells. *Genome Res*. 23:452–61
105. Ryan RF, Schultz DC, Ayyanathan K, Singh PB, Friedman JR, et al. 1999. KAP-1 corepressor protein interacts and colocalizes with heterochromatic and euchromatic HP1 proteins: a potential role for Krüppel-associated box-zinc finger proteins in heterochromatin-mediated gene silencing. *Mol. Cell. Biol.* 19:4366–78
106. Schmitges FW, Radovani E, Najafabadi HS, Barazandeh M, Campitelli LF, et al. 2016. Multiparameter functional diversity of human C2H2 zinc finger proteins. *Genome Res*. 26:1742–52
107. Schuh R, Aicher W, Gaul U, Côté S, Preiss A, et al. 1986. A conserved family of nuclear proteins containing structural elements of the finger protein encoded by *Krüppel*, a Drosophila segmentation gene. *Cell* 47:1025–32
108. Schultz DC, Ayyanathan K, Negorev D, Maul GG, Rauscher FJ 3rd. 2002. SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes Dev*. 16:919–32
109. Schultz DC, Friedman JR, Rauscher FJ 3rd. 2001. Targeting histone deacetylase complexes via KRAB-zinc finger proteins: The PHD and bromodomains of KAP-1 form a cooperative unit that recruits a novel isoform of the Mi-2α subunit of NuRD. *Genes Dev*. 15:428–43
110. Seberg O, Petersen G. 2009. A unified classification system for eukaryotic transposable elements should reflect their phylogeny. *Nat. Rev. Genet.* 10:276
110a. Shibata M, García-García MJ. 2011. The mouse KRAB zinc-finger protein CHATO is required in embryonic-derived tissues to control yolk sac and placenta morphogenesis. Dev. Biol. 349(2):331–41
111. Tadepally HD, Burger G, Aubry M. 2008. Evolution of C2H2-zinc finger genes and subfamilies in mammals: species-specific duplication and loss of clusters, genes and effector domains. *BMC Evol. Biol.* 8:176
112. Takahashi N, Coluccio A, Thorball CW, Planet E, Shi H, et al. 2019. ZNF445 is a primary regulator of genomic imprinting. *Genes Dev*. 33:49–54
113. Thomas JH, Emerson RO, Shendure J. 2009. Extraordinary molecular evolution in the PRDM9 fertility gene. *PLOS ONE* 4:e8505
114. Thomas JH, Schneider S. 2011. Coevolution of retroelements and tandem zinc finger genes. *Genome Res*. 21:1800–12
115. Thompson PJ, Macfarlan TS, Lorincz MC. 2016. Long terminal repeats: from parasitic elements to building blocks of the transcriptional regulatory repertoire. *Mol. Cell* 62:766–76
116. Treger RS, Pope SD, Kong Y, Tokuyama M, Taura M, Iwasaki A. 2019. The lupus susceptibility locus *Sgp3* encodes the suppressor of endogenous retrovirus expression SNERV. *Immunity* 50:334–47.e9
117. Trono D. 2015. Transposable elements, polydactyl proteins, and the genesis of human-specific transcription networks. *Cold Spring Harb. Symp. Quant. Biol.* 80:281–88

118. Tucci V, Isles AR, Kelsey G, Ferguson-Smith AC. 2019. Genomic imprinting and physiological processes in mammals. *Cell* 176:952–65
119. Turelli P, Castro-Diaz N, Marzetta F, Kapopoulou A, Raclot C, et al. 2014. Interplay of TRIM28 and DNA methylation in controlling human endogenous retroelements. *Genome Res.* 24:1260–70
120. Urrutia R. 2003. KRAB-containing zinc-finger repressor proteins. *Genome Biol.* 4:231
121. Valle-García D, Qadeer ZA, McHugh DS, Ghiraldini FG, Chowdhury AH, et al. 2016. ATRX binds to atypical chromatin domains at the 3′ exons of zinc finger genes to preserve H3K9me3 enrichment. *Epigenetics* 11:398–414
121a. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, et al. 2001. The sequence of the human genome. Science 291(5507):1304–51
122. Volkman HE, Stetson DB. 2014. The enemy within: endogenous retroelements and autoimmune disease. *Nat. Immunol.* 15:415–22
123. Wang J, Wang J, Tian CY. 2016. Evolution of KRAB-containing zinc finger proteins and their roles in species evolution. *Yi Chuan* 38:971–78
123a. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, et al. 2002. Initial sequencing and comparative analysis of the mouse genome. Nature 420(6915):520–62
124. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, et al. 2007. A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* 8:973–82
125. Witzgall R, O'Leary E, Leaf A, Onaldi D, Bonventre JV. 1994. The Krüppel-associated box-A (KRAB-A) domain of zinc finger proteins mediates transcriptional repression. *PNAS* 91:4514–18
126. Wiznerowicz M, Jakobsson J, Szulc J, Liao S, Quazzola A, et al. 2007. The Krüppel-associated box repressor domain can trigger de novo promoter methylation during mouse early embryogenesis. *J. Biol. Chem.* 282:34535–41
127. Wolf D, Goff SP. 2007. TRIM28 mediates primer binding site-targeted silencing of murine leukemia virus in embryonic cells. *Cell* 131:46–57
128. Wolf D, Goff SP. 2009. Embryonic stem cells use ZFP809 to silence retroviral DNAs. *Nature* 458:1201–4
129. Wolf G, Yang P, Fuchtbauer AC, Fuchtbauer EM, Silva AM, et al. 2015. The KRAB zinc finger protein ZFP809 is required to initiate epigenetic silencing of endogenous retroviruses. *Genes Dev.* 29:538–54
130. Wolfe SA, Nekludova L, Pabo CO. 2000. DNA recognition by Cys2His2 zinc finger proteins. *Annu. Rev. Biophys. Biomol. Struct.* 29:183–212
131. Yang P, Wang Y, Hoang D, Tinkham M, Patel A, et al. 2017. A placental growth factor is silenced in mouse embryos by the zinc finger protein ZFP568. *Science* 356:757–59
132. Yano K, Ueki N, Oda T, Seki N, Masuho Y, Muramatsu M. 2000. Identification and characterization of human ZNF274 cDNA, which encodes a novel Krüppel-type zinc-finger protein having nucleolar targeting ability. *Genomics* 65:75–80
133. Zuo X, Sheng J, Lau HT, McDonald CM, Andrade M, et al. 2012. Zinc finger protein ZFP57 requires its co-factor to recruit DNA methyltransferases and maintains DNA methylation imprint in embryonic stem cells via its transcriptional repression domain. *J. Biol. Chem.* 287:2107–18