ZNF91 deletion in human embryonic stem cells leads to ectopic activation of SVA retrotransposons and up-regulation of KRAB zinc finger gene clusters

Nina L. Haring,[1] Elisabeth J. van Bree,[1] Whitney S. Jordaan,[1] Judith R.E. Roels,[1] Gonzalo Congrains Sotomayor,[1] Tiziana M. Hey,[1] Fred T.G. White,[1] Marc D. Galland,[1] Marten P. Smidt,[1] and Frank M.J. Jacobs[1,2]

[1]Evolutionary Neurogenomics, Swammerdam Institute for Life Sciences, University of Amsterdam, 1098 XH Amsterdam, The Netherlands; [2]Amsterdam Neuroscience, Complex Trait Genetics, University of Amsterdam, 1098 XH Amsterdam, The Netherlands

Transposable element (TE) invasions have shaped vertebrate genomes over the course of evolution. They have contributed an extra layer of species-specific gene regulation by providing novel transcription factor binding sites. In humans, SINE-VNTR-*Alu* (SVA) elements are one of three still active TE families; approximately 2800 SVA insertions exist in the human genome, half of which are human-specific. TEs are often silenced by KRAB zinc finger (KZNF) proteins recruiting corepressor proteins that establish a repressive chromatin state. A number of KZNFs have been reported to bind SVAs, but their individual contribution to repressing SVAs and their roles in suppressing SVA-mediated gene-regulatory effects remains elusive. We analyzed the genome-wide binding profile for ZNF91 in human cells and found that ZNF91 interacts with the VNTR region of SVAs. Through CRISPR-Cas9-mediated deletion of *ZNF91* in human embryonic stem cells, we established that loss of *ZNF91* results in increased transcriptional activity of SVAs. In contrast, SVA activation was not observed upon genetic deletion of the *ZNF611* gene encoding another strong SVA interactor. Epigenetic profiling confirmed the loss of SVA repression in the absence of *ZNF91* and revealed that mainly evolutionary young SVAs gain gene activation-associated epigenetic modifications. Genes close to activated SVAs showed a mild up-regulation, indicating SVAs adopt properties of *cis*-regulatory elements in the absence of repression. Notably, genome-wide derepression of SVAs elicited the communal up-regulation of KZNFs that reside in KZNF clusters. This phenomenon may provide new insights into the potential mechanisms used by the host genome to sense and counteract TE invasions.

[Supplemental material is available for this article.]

Over the course of evolution, the human genome has been invaded by transposable elements (TEs), which still remain in our genome today as repetitive DNA sequences (Cordaux and Batzer 2009). The majority of TE families have lost the capacity to copy–paste through accumulation of mutations or active repression by the host. In humans, only elements belonging to *Alu*, LINE-1, and SVA families are still capable of retrotransposition (Kazazian et al. 1988; Batzer et al. 1991; Brouha et al. 2003; Ostertag et al. 2003; Wang et al. 2005). Despite and owing to the genome's efforts to repress TE activity, TE invasions have been an important source of evolutionary genomic innovation by creating genomic variability and introducing new gene-regulatory properties. Throughout evolution, waves of TE insertions added a species-specific layer of gene regulation through the introduction of novel transcription factor (TF) binding sites important for chromatin remodeling (Sundaram et al. 2014), immunity (Chuong et al. 2016), brain development (Notwell et al. 2015), liver functioning (Trizzino et al. 2017, 2018), and stem cell pluripotency (Kunarso et al. 2010; Pontis et al. 2019). Profiling of enhancer-associated histone marks (H3K27ac and H3K4me1) in liver cells derived from six primate species revealed that young TEs account for most of the newly evolved *cis*-regulatory elements that drive species-specific gene expression (Trizzino et al. 2017). In addition, novel TE insertions can affect gene expression negatively through the induction of local heterochromatin mediated by KRAB zinc finger (KZNF) proteins and cofactors (Lippman et al. 2004; Rebollo et al. 2011; Jacobs et al. 2014). Thus, we can conclude that TE insertions can have a long-lasting impact on gene regulation, which continues many millions of years after the spreading of novel TE insertions has been halted. Most TE families are silenced through binding of members of the KZNF family of which members coevolved to recognize specific TE families (Wolf and Goff 2009; Thomas and Schneider 2011; Jacobs et al. 2014; Najafabadi et al. 2015; Wolf et al. 2015; Schmitges et al. 2016; Imbeault et al. 2017; Seah et al. 2019) and recruit KAP1 (Nielsen et al. 1999; Sripathy et al. 2006) and repressive epigenetic modifiers (Schultz et al. 2001, 2002; Turelli et al. 2014). The KZNF family expanded through segmental duplications along the course of evolution, and most of its members reside in clusters (Huntley et al. 2006; Thomas and Schneider 2011). How KZNFs sense the invasion of new TEs and mediate a response to control their spread remains elusive. Recently, it has been shown that derepression of TEs leads to up-regulation of KZNFs, potentially induced through an innate immune response (Kauzlaric et al. 2017; Tao et al. 2018; Tie et al. 2018). The mechanism behind this and the consequences

of the up-regulation of KZNFs remain to be established. In addition, it is becoming increasingly clear that the function of KZNFs is not limited to TE repression. Several other roles have been attributed to specific KZNFs, such as autoregulation by binding to itself and/or other members of the KZNF family (Frietze et al. 2010; Yan et al. 2017), gene regulation (Oleksiewicz et al. 2017; Yang et al. 2017; Chen et al. 2019; Farmiloe et al. 2020), and maintenance of DNA methylation at imprinting control regions (Li et al. 2008; Takahashi et al. 2015, 2019; Riso et al. 2016).

In this study we focus on a primate-specific TE class, named after its composite domains: SINE-VNTR-*Alu* (SVA). There are approximately 2800 fixed SVA insertions in the human genome, half of which were inserted in our genome after the split with the last common ancestor (LCA) with chimpanzees and are therefore unique to humans (Wang et al. 2005). Several lines of evidence showed that SVA insertions harbor a strong gene-regulatory potential (Savage et al. 2013, 2014; Jacobs et al. 2014; Pontis et al. 2019) and can function both as transcriptional activators and repressors (Trizzino et al. 2017, 2018; Pontis et al. 2019). Although the primate-specific ZNF91 was identified as a strong repressor of SVA activity (Jacobs et al. 2014), it remained unclear to what extent ZNF91 binds to SVAs and if ZNF91 is sufficient for SVA repression in a human-cellular environment. Moreover, a recent study showed that a number of other KZNF proteins also bind SVAs (Imbeault et al. 2017), raising the possibility that multiple KZNFs are necessary to silence SVAs. Here, we assessed the role of ZNF91 in repressing SVAs by mapping its binding sites and creating a genetic deletion of *ZNF91* in human embryonic stem cells. The data described here elucidate the gene-regulatory role of SVAs under conditions in which epigenetic repression is lost. Furthermore, our findings strengthen previous observations regarding TE-mediated communal up-regulation of KZNF clusters (Kauzlaric et al. 2017; Tie et al. 2018; Pontis et al. 2019), which may deepen our understanding of how KZNFs act in the presence of unrepressed TEs in our genome.

Results

ZNF91 binds at the *Alu*-VNTR border of the vast majority of SVAs

To determine genome-wide binding of ZNF91, we transfected HEK293 cells with a ZNF91-GFP fusion construct and performed GFP chromatin immunoprecipitation followed by sequencing (ChIP-seq) (Supplemental Fig. S1A,B). Two independent ChIP-seq replicates revealed 8302 and 16,425 high-confidence binding sites for replicate 1 and 2, respectively (Supplemental Fig. S1C). For further analysis, 7220 ZNF91 peaks that were shared between both replicates were considered.

First, we analyzed binding sites with the strongest ZNF91 enrichment (top 10% based on MACS score) to determine to which genomic regions ZNF91 preferentially binds. This analysis revealed that 85% of the strongest enriched sites reside within SVAs, indicating that they are the primary target of ZNF91 (Fig. 1A). Genome-wide, the majority (88%) of SVAs is bound by ZNF91 (Fig. 1B), which was seen for all subclasses (Fig. 1C). Therefore, ZNF91 does not seem to have a clear preference for evolutionary old or younger SVAs. ZNF91 binding was highest at the *Alu*-VNTR border within SVAs, at the same location where KAP1 shows the highest level of binding (Fig. 1D). Fifty-nine percent of ZNF91-bound SVAs showed a second smaller ZNF91 peak at the other side of the VNTR, at the VNTR-SINE border (Fig. 1D,E). This second ZNF91 binding site was observed mainly for young SVA subclasses

(Fig. 1E). To determine the core sequence bound by ZNF91, we discovered two motifs on either side of the VNTR region within the SVA. The motifs are partly overlapping on a VNTR-derived sequence, indicating that ZNF91 predominantly binds to the VNTR region of SVAs (Fig. 1D,F). The observed high level of ZNF91 ChIP coverage between the two ZNF91 peak summits suggests that ZNF91 binds to multiple locations throughout the VNTR region (Fig. 1E; Supplemental Fig. S1D). Many of the SVAs not bound by ZNF91 are shorter or truncated, suggesting that the lack of ZNF91 binding is likely caused by the absence of that part of the SVA that contains the ZNF91 binding site (Fig. 1G). Collectively, these data support earlier findings pointing toward a major role of ZNF91 in SVA repression.

SVAs become epigenetically and transcriptionally active in the absence of ZNF91

To determine the fate of SVAs in the absence of ZNF91, we genetically deleted *ZNF91* in human embryonic stem cells (hESCs) using CRISPR-Cas9 technology. RNA-sequencing analysis of three independent clonal *ZNF91* knockout (KO) hESC lines confirmed the absence of *ZNF91* transcripts (Fig. 2A). Whereas SVAs are efficiently silenced in hESCs under normal conditions, SVAs became transcriptionally active in *ZNF91* KO cells. Most SVAs showed only modest transcriptional activation in the absence of ZNF91; differential expression analysis revealed 173 significantly up-regulated SVA transcripts in *ZNF91* KO versus wild-type (WT) cells (Fig. 2B). Analysis of the transcript structure revealed that many SVAs generate transcripts that run from start (5′) to end (3′) of the element. However, we also identified 10 SVAs that generated transcripts that extend into the downstream sequence (up to 150 kb) generating large lncRNAs (Fig. 2C; Supplemental Table S1).

To determine the epigenetic fate of SVAs in the absence of ZNF91-mediated repression, we performed chromatin immunoprecipitation for the repressive epigenetic mark H3K9me3, which is present on the majority of TEs including SVAs under normal conditions in hESCs (Fig. 3A; Supplemental Fig. S2). By ChIP-qPCR we observed an approximate threefold decrease of H3K9me3 signal on SVAs in *ZNF91* KO cells compared to WT cells (Fig. 3B). This indicates that in the absence of ZNF91, SVAs lose their epigenetic repressive marks. We next assessed whether loss of epigenetic repression in *ZNF91* KO cells is accompanied by gain of epigenetic activation by analyzing the presence of epigenetic marks associated with an active chromatin state. We focused on the enhancer/promoter associated marks H3K4me3 and H3K27ac and assessed their genome-wide distribution by ChIP-seq in *ZNF91* KO and WT hESCs. H3K4me3 is mainly found at transcription start sites (TSSs) and is associated with active transcription (Bernstein et al. 2005), whereas H3K27ac histone modifications are associated with active enhancers and promoters (Wang et al. 2008; Creyghton et al. 2010). In *ZNF91* KO hESCs, 29.2% of SVAs became H3K4me3 positive compared to 0.8% in WT hESCs (Fig. 3C). H3K27ac was detected at 6% of SVAs in *ZNF91* KO hESCs, and nearly all of these were also H3K4me3 positive (Fig. 3C). We noticed that the level of epigenetic activation of SVAs roughly correlates with their evolutionary age: predominantly the youngest classes of SVAs (SVA-D/E/F) gained an H3K4me3 mark upon *ZNF91* deletion, with 92% of SVA-E and 65% of SVA-F elements displaying signs of epigenetic activation (Fig. 3D). The H3K4me3 peak colocalizes with the 5′ peak of ZNF91 binding, suggesting that ZNF91 binds to the domain that is also important for the activation of the SVA (Fig. 3E). Furthermore, H3K4me3-
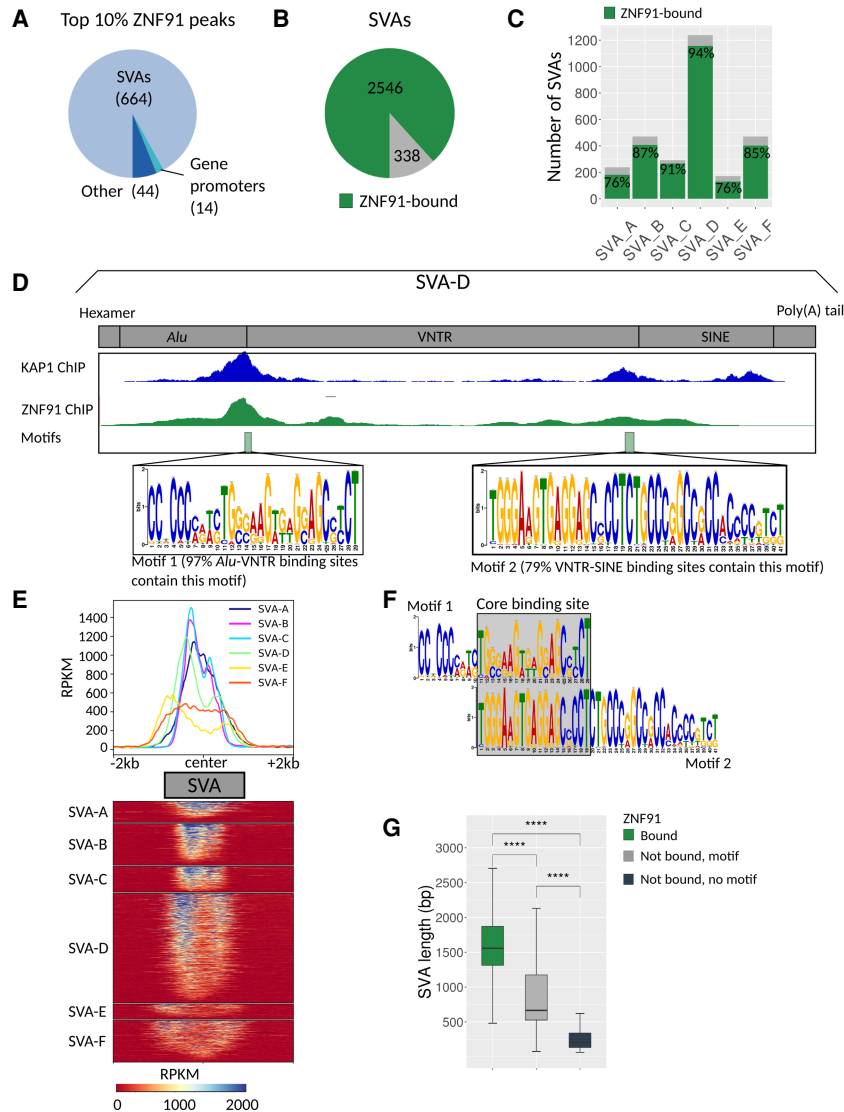
Figure 1. ZNF91 binds at the *Alu*-VNTR and VNTR-SINE border of all SVA subclasses. (*A*) Pie chart showing the distribution of the top 10% ZNF91 peaks based on MACS score. (*B*) Pie chart showing the number of SVA elements that are bound (green, 2545) and not bound (gray, 338) by ZNF91. (*C*) Frequency plot showing fraction of SVA subclasses bound by ZNF91 (green). (*D*) Schematic of SVA-D with different domains indicated. A pile-up of KAP1 summits on the SVA-D consensus is displayed in blue (Jacobs et al. 2014), and a representative example of ZNF91 coverage on an SVA-D is in green. Two ZNF91 binding motifs were identified localizing to the *Alu*-VNTR border and VNTR-SINE border. The fraction of SVA elements that contain motif 1 in the peaks localizing to the *Alu*-VNTR border and motif 2 in the peaks localizing to the VNTR-SINE border are expressed in percentages. (*E*) Heatmap and profile plots of ZNF91 ChIP showing coverage in RPKM of the mean of replicates 1 and 2 with SVA center as reference point. Average signal plotted per SVA subclass. (*F*) Comparison of motifs 1 and 2 showing overlap at the *center*. (*G*) Length of SVA elements bound (2546, green) and not bound (gray) by ZNF91. Light gray indicates that at least one of the two SVA binding motifs (199) was present, and dark gray indicates that no motif was detected (139). For statistical testing, an unpaired Wilcoxon rank-sum test was used followed by an FDR correction: (****) $P < 0.0001$.

for both H3K4me3 and H3K27ac are located significantly closer to expressed genes than their H3K4me3-positive and H3K4me3/H3K27ac-negative counterparts (Fig. 3G). Recently, a ChIP-exo screen, including more than 200 KZNF proteins, revealed that ZNF611 also binds SVAs, although to a lesser extent than ZNF91. Fifty-eight percent of SVAs is bound by ZNF611, and SVA-A and SVA-F are bound most frequently (Supplemental Fig. S3A,B; Imbeault et al. 2017). Because both *ZNF91* and *ZNF611* are expressed in hESCs, we generated *ZNF611* KO lines to assess the role of ZNF611 in repressing SVAs (Supplemental Fig. S3C). To account for possible redundancy of both KZNF proteins, also *ZNF91/ZNF611* double KO (dKO) cell lines were created and analyzed (Supplemental Fig. S3C). RNA-seq revealed no transcriptional activation of SVAs upon *ZNF611* deletion (Supplemental Fig. S3D,E). No additional effect on SVA transcriptional activation was observed in *ZNF91/ZNF611* dKO cells, indicating there is no functional redundancy or compensation between ZNF91 and ZNF611 (Supplemental Fig. S3D,E). In conclusion, our data reveal that ZNF91, but not ZNF611, is essential for SVA repression in hESCs. In the absence of ZNF91-mediated repression, SVAs lose epigenetic repression, become aberrantly transcribed, and bear histone marks associated with active chromatin.

SVAs function as gene regulators in the absence of repression

SVAs are shown to harbor strong gene-regulatory properties (Savage et al. 2013, 2014; Jacobs et al. 2014; Trizzino et al. 2017, 2018; Pontis et al. 2019). To provide further support for this, we analyzed whether human genes with an SVA nearby showed differential gene expression compared to the orthologous genes in the rhesus macaque monkey, which is a species that lacks SVAs completely. This was done in human and rhesus ESC lines by comparing gene expression profiles throughout human and rhesus neuronal differentiation (Field et al. 2019). This comparative analysis revealed that on average, genes that had a nearby SVA insertion in the human lineage, were more likely to display a higher expression level in humans compared to rhesus (Supplemental Fig. S4A,B), indicating an enhancing effect of the SVA insertions on gene regulation.

We next examined gene expression changes potentially caused by activated SVAs in *ZNF91* KO hESCs. Genes located in

negative SVAs were shorter than their H3K4me3-positive counterparts, indicating that shorter elements, likely owing to shorter VNTRs, have fewer domains required for activation (Fig. 3F).
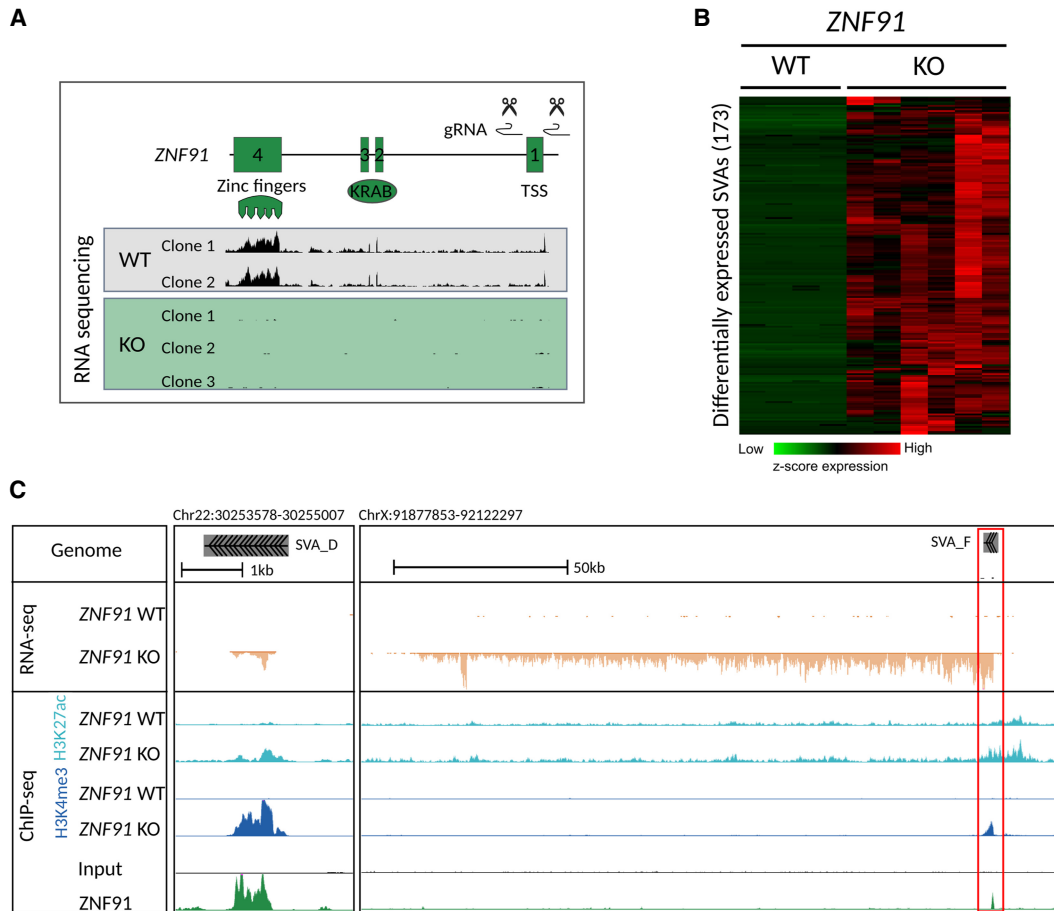
Next, we assessed whether SVA insertions located in transcriptionally active regions were more likely to become activated upon *ZNF91* deletion. To this end, we looked at the distance between activated SVAs and genes expressed in hESCs. SVAs positive

Figure 2. SVA elements become transcriptionally active in the absence of *ZNF91*. (*A*) CRISPR-Cas9 design to delete 172 bp around the transcription start site (TSS) of *ZNF91* and RNA-sequencing coverage tracks showing absence of *ZNF91* expression in *ZNF91* KO clonal hESC lines. (*B*) Heatmap of hierarchical clustering (Euclidean distance) using *Z*-score of scaled counts of SVA elements that are up-regulated in *ZNF91* KO hESCs (defined as $\log_2$ fold change ≥ 3). (*C*) Coverage of RNA-seq and H3K4me3 and H3K27ac ChIP-seq data for *ZNF91* WT and KO hESCs and ZNF91-GFP ChIP on HEK293 cells: (*left*) transcripts running from 5′ to 3′ end of SVA; (*right*) red box indicates SVA element as driver of lncRNA. RNA-seq coverage tracks were scaled using DESeq2 scaling factors. ChIP-seq coverage tracks were scaled based on control regions.

proximity of an H3K4me3-positive SVA were overrepresented (Fisher's exact test, $P < 0.00001$) among 54 differentially expressed genes. The vast majority (13/15) of differentially expressed genes within 25 kb of an SVA was up-regulated (Fig. 4A; Supplemental Table S1). Furthermore, significantly up-regulated genes are located closer to H3K4me3-positive SVAs compared to down-regulated genes (Fig. 4B). In addition, genes located within a 25-kb window from an H3K4me3-positive SVA showed a mild, but significant, up-regulation compared to a group of randomly selected genes (Fig. 4C). Together this further confirms that derepressed SVAs have a positive regulatory effect on the expression of nearby genes. Besides potentially functioning as classical enhancers, we observed that some of the up-regulated genes had an activated SVA near their TSSs, of which the H3K4me3 signal in the *ZNF91* KO cells extended the H3K27ac and H3K4me3 signal that is present in the gene promoter under normal conditions (Fig. 4D; Supplemental Table S1). Some activated SVAs become alternative TSSs in *ZNF91* KO cells, generating chimeric SVA-gene transcripts with the genes that are located downstream (Supplemental Table S1). Of the 15 differentially expressed genes located within 25 kb of an H3K4me3 marked SVA 7 functioned as alternative TSS. An example of an alternative ectopically expressed transcript identified

through de novo transcript assembly is the SVA-*HORMAD1* fusion transcript. *HORMAD1* is normally exclusively expressed in the testis, but aberrantly induced in hESCs upon *ZNF91* deletion (Fig. 4D). In the expressed sequence tags (EST) database we found ESTs that covered this fusion transcript, indicating that SVA-*HORMAD1* fusion transcripts also exist in a (patho-) physiological context (Supplemental Fig. S4C).

In line with what is observed for other KZNF proteins (Imbeault et al. 2017; Farmiloe et al. 2020), ZNF91 also binds to genomic regions outside SVAs. We found that ZNF91 binds to 1062 gene promoters. However, differential gene expression analysis using RNA-seq data of *ZNF91* KO hESCs revealed no collective expression changes of ZNF91-bound genes in hESCs (Supplemental Fig. S4D). Even if we only consider ZNF91 peaks in promoters that overlap with KAP1 binding in hESCs (Supplemental Fig. S4E), we find no evidence for a direct regulatory role of ZNF91 on the promoters where it binds independently of SVAs (Supplemental Fig. S4F). Therefore, our data suggest that the main role of ZNF91 is to repress SVA transcriptional activity and the SVA-mediated gene-regulatory potential as a result of the propensity of SVAs to gain activating epigenetic marks in the absence of repression.
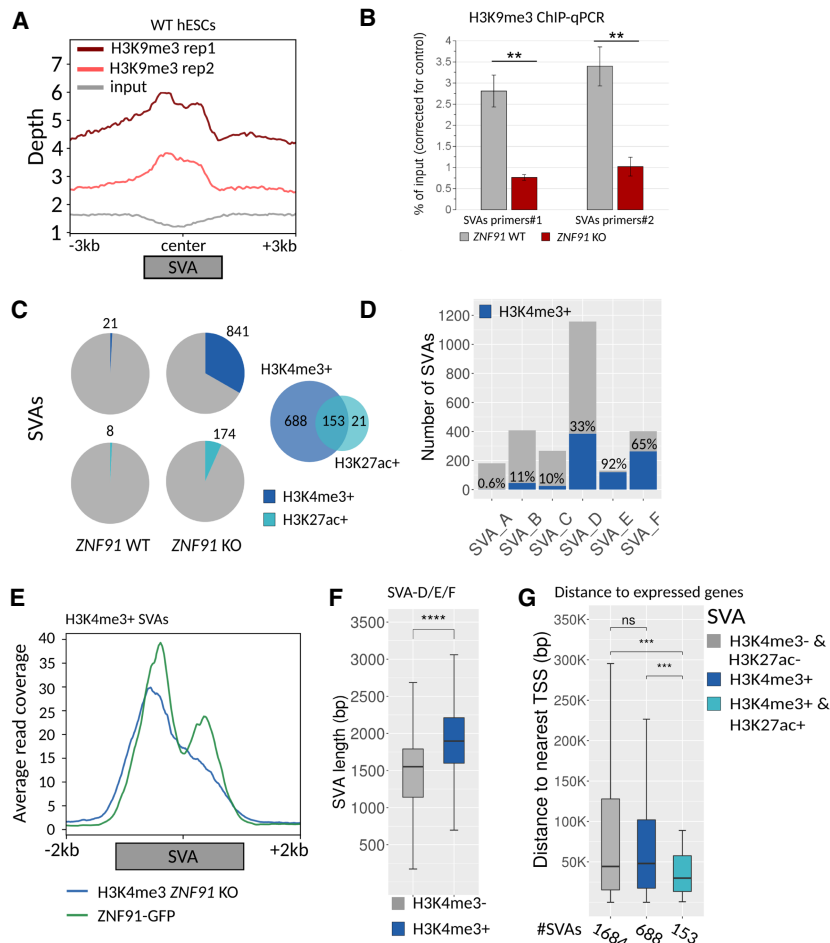
Figure 3. Evolutionarily young SVA elements become epigenetically active in the absence of ZNF91. (*A*) Profile plots showing coverage of H3K9me3 centered on SVA elements in WT hESCs, and 3 kb upstream of and downstream from SVAs is displayed. (*B*) H3K9me3 ChIP-qPCR on *ZNF91* WT and KO hESCs using two SVA primer pairs. Percentage of input was normalized against a control (LTR12). A one-sided, unpaired *t*-test was used for statistical analysis. (*C*) Pie charts showing number of H3K4me3 (dark blue) and H3K27ac (light blue) positive SVA elements in *ZNF91* WT and KO hESCs; Venn diagram showing overlap of H3K4me3 and H3K27ac peaks. (*D*) Absolute number (gray) and percentage of H3K4me3-positive (blue) SVA elements in each subclass. (*E*) Profile plot showing ZNF91 ChIP signal in HEK293 cells and H3K4me3 signal in *ZNF91* KO hESCs. Average read coverage on H3K4me3-positive SVAs was displayed 2 kb upstream of and downstream from the element. (*F*) Box plot showing length of H3K4me3-positive and -negative SVA subtypes D, E, and F. (*G*) Box plots showing distance from H3K4me3/H3K27ac-negative, H3K4me3-positive, and H3K4me3/H3K27ac-positive SVA elements to nearest TSS of a gene expressed in hESCs (baseMean > 10) in base pairs. Outliers are not displayed. For statistical testing in *F* and *G*, an unpaired Wilcoxon rank-sum test was used, for *G* followed by an FDR correction: (****) $P < 0.0001$; (***) $P < 0.001$; (**) $P < 0.01$; (ns) not significant. For H3K4me3 and H3K27ac ChIP analyses, only SVA elements bound by ZNF91 were included.

## Up-regulation of SVA-dense KZNF clusters upon *ZNF91* deletion

Although only 54 differentially expressed genes met the stringent statistical threshold of the DESeq2 analysis, we noticed a modest collective up-regulation of KZNFs in *ZNF91* KO cells compared to WT. The differential expression was limited to KZNFs located in clusters of KZNFs, but solitary KZNFs were unaffected (Fig. 5A). KZNF clusters are found on a number of chromosomes and vary in the amount of KZNFs per cluster. The KZNF density on Chromosome 19 is very high, harboring many KZNF clusters and a high number of KZNFs (Fig. 5B). A subset of these clusters shows a collective up-regulation in the absence of ZNF91 (Fig. 5C). Remarkable is that Chromosome 19 also shows an exceptionally high

density of SVAs (Fig. 5B), which is highest in the areas where the up-regulated KZNF clusters reside. A proportion of SVAs in these clusters become H3K4me3 positive upon *ZNF91* deletion (Fig. 5C). This suggests that the collective up-regulation of KZNFs may be a consequence of activated SVA insertions in close vicinity of KZNFs (Fig. 5D). This is further supported by the fact that only KZNFs within 100 kb of an H3K4me3-positive SVA showed an up-regulation that was significant compared to a set of randomly selected genes (Fig. 5E). This regulatory effect was restricted to KZNFs because other genes located inside KZNF clusters on Chromosome 19 do not show a collective up-regulation (Supplemental Fig. S5A). An alternative possibility is that KZNFs are regulated by ZNF91 directly through ZNF91 binding to their promoters. However, for the KZNFs that showed evidence for ZNF91 binding in their promoter, even after reducing the threshold of ZNF91-peak detection, ZNF91 binding did not correlate with differential expression upon *ZNF91* KO (Supplemental Fig. S5B,C), in line with what was observed for other ZNF91-bound genes (Supplemental Fig. S4D,F). This suggests that, although we cannot fully rule this out, up-regulation of KZNFs does not seem to be mediated by loss of ZNF91 binding to their promoters. This provides further circumstantial support for the hypothesis that KZNF up-regulation is mediated by derepressed SVAs in the KZNF clusters, although we were not able to provide conclusive evidence for this. A few studies have linked KZNF up-regulation to the induction of an innate immune response by retroviral transcripts (Li et al. 2018; Tie et al. 2018). However, we found no significant up-regulation of genes associated with a response to viral infections in *ZNF91* KO hESCs (Supplemental Fig. S5D). Although the mechanism remains elusive, our results are in line with previous studies reporting a sim-

ilar up-regulation of KZNFs in response to loss of KAP1-mediated repression of TEs (Kauzlaric et al. 2017; Tie et al. 2018). Whether or not derepressed SVAs are mediating those observed effects remains to be established, but our findings show a possible trace of a feedback mechanism through which our genome senses the presence of uncontrolled TEs, where up-regulation of KZNFs may be the pioneering response to counteract their invasion.

Discussion

Here, we describe the genome-wide effects of genetic deletion of *ZNF91*, which binds to SVAs, one of the few currently active TEs in our genome. We show that SVAs become derepressed upon
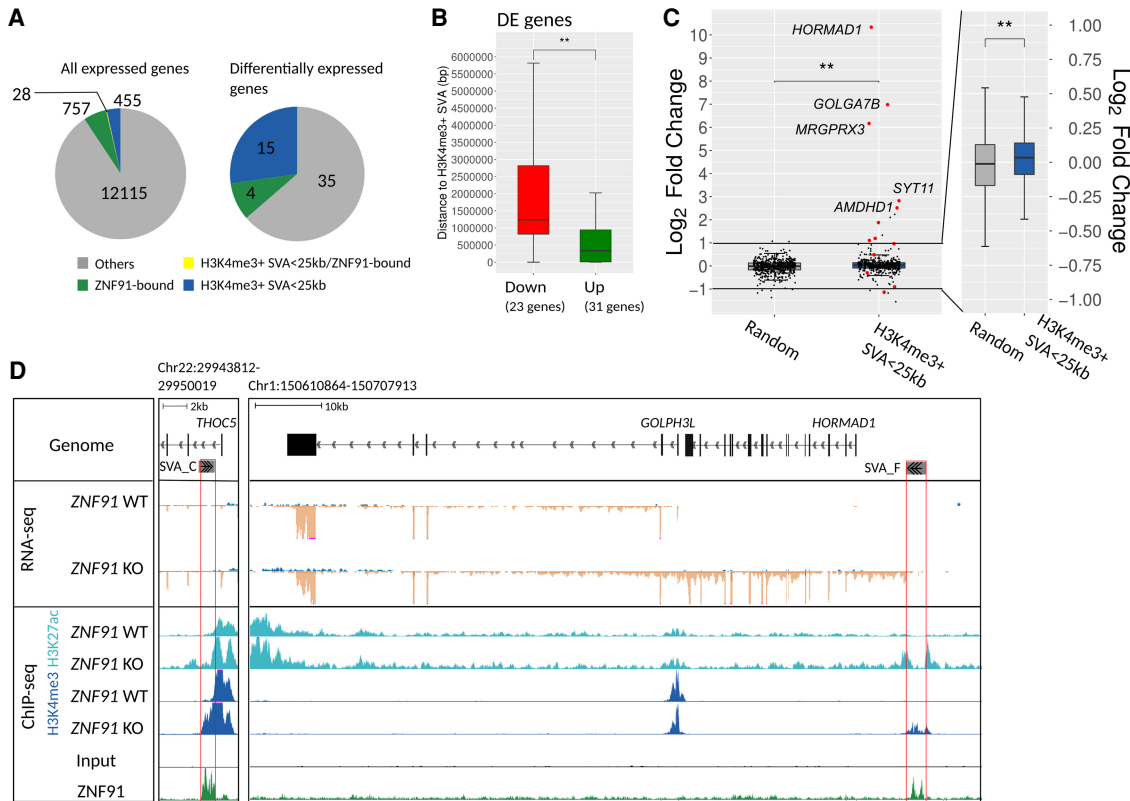
Figure 4. Differentially expressed genes in *ZNF91* KO cells are bound by ZNF91 and driven by activated SVA elements. (*A*) Pie charts of all expressed (baseMean > 10) and differentially expressed (DE) genes (DESeq2 analysis, $P_{adj} < 0.05$) indicating how many genes are bound by ZNF91 at their promoter (green), located within 25 kb of an H3K4me3-positive SVA (blue), both (yellow), and neither (gray). (*B*) Box plots of distance to nearest H3K4me3-positive SVA from TSS of DE genes (23 genes down-regulated and 31 up-regulated). (*C*) Box plots showing log$_2$ fold change of genes that are located within 25 kb of an H3K4me3-positive SVA (420 genes) compared to a set of randomly selected genes (556 genes). Red data points indicate genes that were differentially expressed (FDR < 0.05). Only expressed genes (baseMean > 10) were included in this graph. (*D*) Coverage tracks of RNA-seq (blue is transcription from the positive strand, and orange is from the negative strand) and ChIP-seq data: (*left*) activation of an intronic SVA-C results in extension of H3K4me3 and H3K27ac signals at the promoter, leading to increased expression of *THOC5*; (*right*) activation of an SVA-F upstream of *HORMAD1* leads to generation of an SVA-*HORMAD1*-*GOLPH3L* fusion transcript. Red boxes indicate activated SVA elements. RNA-seq coverage tracks were scaled using DESeq2 scaling factors. ChIP-seq coverage tracks were scaled based on control regions. For all statistical testing, an unpaired Wilcoxon rank-sum test was used, if necessary, followed by an FDR correction: (**) $P < 0.01$.

genetic deletion of *ZNF91*, but not by deleting *ZNF611*. In addition, our data confirm previous observations that suggested a hidden gene-regulatory potential of SVAs in our genome (Savage et al. 2013, 2014; Jacobs et al. 2014; Trizzino et al. 2017; Pontis et al. 2019). Loss of *ZNF91* leads to derepression of the majority of SVAs in our genome, but among individual SVA insertions we observed a large variation in the levels of epigenetic and transcriptomic signatures associated with the ectopic activation of SVAs. Derepressed SVAs occasionally generate SVA-gene fusion transcripts. This phenomenon called exonization is also observed for some other TE classes (Faulkner et al. 2009; Hancks et al. 2009; Pontis et al. 2019). By gaining a novel TE-derived promoter, such TE-gene fusion transcripts may display highly different characteristics from the native transcripts regarding developmental timing and cell type specificity. That this phenomenon can occur under normal conditions is evidenced by many reports of de novo SVA insertions leading to disease through mRNA dysregulation and alternative splicing (Kobayashi et al. 1998; Wilund et al. 2002; Deng et al. 2008; Van der Klift et al. 2012; Nakamura et al. 2015). In addition, our data show that epigenetic activation of SVAs can lead to gene dysregulation. This is in line with previous studies showing a global derepression of young TEs under pathological cellular con-

ditions that affect the epigenetic status of our genome such as cancer and aging-associated diseases (Bollati et al. 2009; Szpakowski et al. 2009; Jintaridth and Mutirangura 2010; Barchitta et al. 2014). As an example, we showed that in the absence of ZNF91-mediated repression, SVAs drive the expression of a chimeric *HORMAD1* transcript in hESCs, a gene exclusively expressed in the testis. *HORMAD1* is found to be aberrantly expressed in various types of cancer (Shahzad et al. 2013; Watkins et al. 2015; Nichols et al. 2018). Further analyses of RNA-seq and histone ChIP data of these cancer cells could point out if pathological derepression of SVAs is the cause of ectopic *HORMAD1* expression in cancers.

Our data suggest that despite the ability of multiple KZNF proteins to bind to SVAs, including ZNF611, genetic deletion of *ZNF91* is sufficient to elicit derepression of the majority of SVAs in our genome. We further show that although ZNF611 binds to all SVA subclasses, genetic deletion of *ZNF611* in hESCs does not lead to SVA derepression. We cannot rule out, however, that ZNF611 or any of the other SVA-interacting KZNFs, contribute to H3K9me3 deposition and SVA repression in other cell types or during different developmental stages. A recent report shows that ZNF611 is capable of repressing H3K27ac-positive SVAs upon overexpression in naive hESCs (Pontis et al. 2019). This raises the
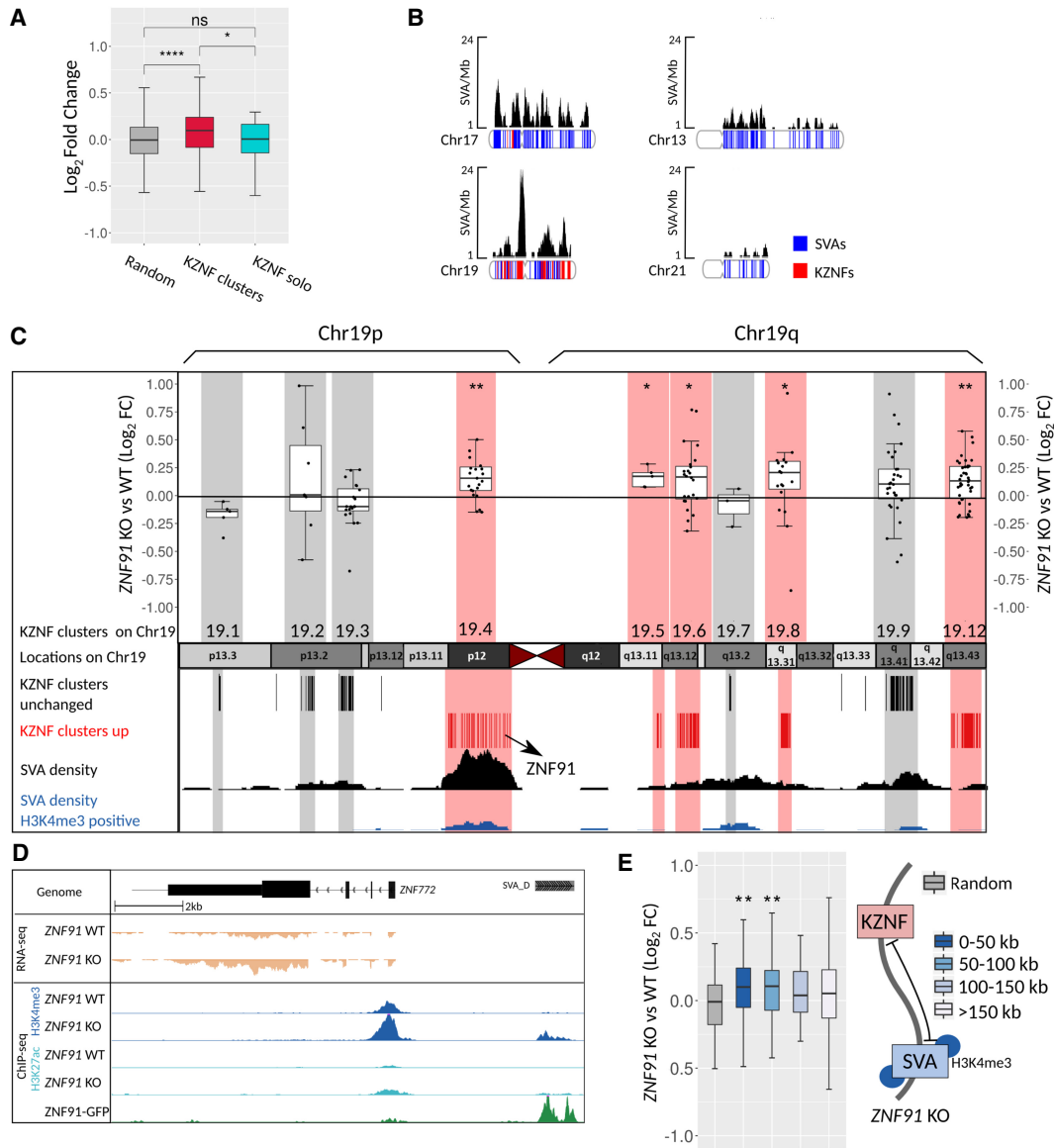
Figure 5. Up-regulation of clustered KZNF genes in proximity of activated SVA elements. (*A*) Box plots showing log$_2$ fold change of a random gene set (257), KZNF located in clusters (265), and solitary KZNF genes (38). Clusters and solitary KZNF were defined by Thomas and Schneider (2011). Only genes with a baseMean > 10 were included. (*B*) SVA coverage (SVA/Mb) is shown for Chromosomes 17 and 19 with higher SVA density than expected and Chromosomes 13 and 21 with lower SVA density than expected (based on Wang et al. 2005). Chromosome density plots show distribution of SVA elements (blue) and KZNF genes (red). (*C*) Schematic representation of Chromosome 19 showing the location of KZNF clusters. Box plots of log$_2$ fold changes on *top* show if clusters are up-regulated (red) or unchanged (gray). Only genes with baseMean > 10 were included. Number of expressed genes per cluster: 19.1 = 9 genes, 19.12 = 42 genes, 19.2 = 7 genes, 19.3 = 20 genes, 19.4 = 23 genes, 19.5 = 5 genes, 19.6 = 24 genes, 19.7 = 3 genes, 19.8 = 19 genes, 19.9 = 26 genes. Coverage tracks show SVA/Mb for all SVA elements (black) and H3K4me3-positive SVA elements (blue). (*D*) Coverage of RNA-seq and ChIP-seq (H3K4me3 and H3K27ac) on *ZNF91* WT and KO hESCs and ZNF91 in HEK293 cells showing ZNF772 and upstream SVA-D. RNA-seq coverage tracks were scaled using DESeq2 scaling factors. ChIP-seq coverage tracks were scaled based on control regions. (*E*) Box plots showing log$_2$ fold change of KZNF grouped per distance to nearest SVA that gained H3K4me3 upon *ZNF91* deletion (random = 88 genes, 0–50 kb = 112 genes, 50–100 kb = 60 genes, 100–150 = 46 genes, >150 kb = 85 genes). For all analyses only expressed (baseMean > 0) genes were included. For all statistical testing an unpaired Wilcoxon rank-sum test was used, followed by an FDR correction: (ns) $P > 0.05$; (*) $P < 0.05$; (**) $P < 0.01$; (****) $P < 0.0001$.

possibility that in primed hESCs that were used in our study, endogenous *ZNF611* expression levels are too low to repress SVAs effectively.

The *ZNF91* gene is fixed in the human population, and based on 250,000 genomes included in the gnomAD database of genetic variants, no homozygous null mutations exist in the human population. Sporadically, low-confidence loss-of-function (LOF)

mutations are found in genomes but according to gnomAD "the variant annotation and/or quality of these calls is dubious." (See Karczewski et al. 2020.) Similar sporadic low-confidence LOF mutations are found in any essential gene. The absence of null alleles indicates that *ZNF91* is an essential gene in humans, despite its recent emergence in primate genomes. Furthermore, there is no evidence of aberrant *ZNF91* expression levels being associated with

human diseases. Possibly, ZNF91 has an even greater and more essential role in early embryonic stages, which prevents deleterious mutations to be carried over to viable offspring.

Our data showed that activation of SVAs results in a collective up-regulation of KZNFs, specifically affecting KZNFs located in SVA-rich clusters. This supports a hypothetical model in which TE derepression elicits a defensive response by the host to counteract TE outbursts. The mechanism that drives this response remains elusive. It was recently reported that the presence of TE-derived transcripts could elicit an innate immune response resulting in up-regulation of KZNFs (Li et al. 2018; Tie et al. 2018). However, in our study we did not find support for this. Possibly, higher levels of KZNF protein lead to more aspecific binding capacities of KZNFs, which could be a short-term solution of our genome to establish repression of TEs. In the long term, the KZNFs that were most able to repress the TE may be subject to evolutionary pressure and subsequently undergo structural evolutionary changes to optimize their repressive capacities. In our study, the collective up-regulation of KZNFs is likely mediated by the high density of activated SVAs in the KZNF clusters. These findings are supported by an earlier study that attributed a collective up-regulation of KZNF clusters in mouse KAP1 KO cells to aberrant activation of nearby TEs (Kauzlaric et al. 2017). Our data provide further support for this hypothesis and show a possible mechanism in which the epigenetic activity of the invading TEs is coopted by the host to establish a genome-wide defense against their outburst. Our genome has endured ongoing retrotransposition activity by SVAs for many millions of years, and ZNF91 has a crucial role in reducing their invasive behavior. By unleashing this active class of TEs in the human genome by removing its main repressor protein, our data shed light on the complex interactions between TE invasions and the KZNFs that evolve to repress them to protect our genome's integrity.

Methods

HEK293 cell culture and transfection

HEK293 cells were cultured in DMEM (Invitrogen) supplemented with 10% fetal bovine serum (FBS, Gibco) and 1% penicillin/streptomycin (Invitrogen). To passage cells, 0.25% trypsin/0.01% EDTA (Invitrogen) in PBS was used 1:10 every 3–4 d. One day before transfection, 2 million cells were seeded per 60 mm dish to reach 70% confluency on the day of transfection, and 4.1 µg pCAG.ZNF91-GFP or pCAG.GFP plasmid was used to transfect each 60 mm dish using 5.5 µL Lipofectamine 3000 (Invitrogen) and 8.3 µL P3000 (Invitrogen). Forty-eight hours after transfection, cells were isolated for chromatin immunoprecipitation (ChIP).

Human embryonic stem cell culture

H9 human embryonic stem cells were grown on matrigel-coated dishes (Corning). They were cultured in hESC medium that was incubated with mouse embryonic fibroblasts (MEFs) for 24 h. hESC medium consisted of DMEM/F12 supplemented with 2 mM L-glutamine (Invitrogen), 20% knockout serum replacement (Gibco), penicillin/streptomycin (Invitrogen), nonessential amino acids (Invitrogen), and 0.1 mM β-mercaptoethanol (Invitrogen). Medium was supplemented with basic fibroblast growth factor (Sigma-Aldrich, 8 ng/µL) and changed daily to secure pluripotency of hESCs. For maintenance of the culture, cells were grown in colonies and passaged manually by cutting the colonies with a needle. For transfection and clonal expansion hESCs were grown as single cells and passaged using accutase (Sigma-Aldrich) at ratios 1:4–1:10. For single cell culturing, culture medium was supplemented with ROCK inhibitor Thiazovivin (2 µM, Sigma-Aldrich) 1 h before dissociation and during plating. For H3K9me3 ChIP-seq experiments, hESCs were cultured on MEFs as described in Jacobs et al. (2014).

hESC transfection and clonal expansion

Two gRNAs targeting a few hundred base pairs around the TSS of *ZNF91* or *ZNF611* were cloned into pX330 (Supplemental Table S2). For transfection 400,000 hESCs per well were seeded on a six-well plate 1 d before transfection. Per well, a total of 3 µg pX330(Cas9-gRNA), 1.5 µg of each of the two gRNAs, and 20 ng of pCAG.GFP was mixed with 5 µL P3000 reagent in a total volume of 50 µL Opti-MEM (Invitrogen). This DNA mix was combined with a second mix containing 3.75 µL Lipofectamine 3000 (Invitrogen) and 46.25 µL Opti-MEM (Invitrogen) and incubated for 15 min at room temperature. Medium was refreshed before adding 100 µL transfection mix dropwise. Forty-eight hours after transfection GFP positive cells were sorted using fluorescence activated cell sorting (FACS) on a FACS Aria III with a 100-µm nozzle. Cells were plated at low density (5000–9000 cells) on MEF-coated 60-mm plates for clonal expansion. Until colonies appeared, cells were cultured in recovery medium, consisting of hESC-conditioned medium and MEF-conditioned medium (1:1). For *ZNF91* and *ZNF611*, WT and KO clonal lines were expanded in parallel. For *ZNF91/ZNF611* dKO cells, one clonal *ZNF91* KO line was used to generate *ZNF611* WT and KO lines in parallel. To ensure clonality, colonies were passaged manually to a 96-well plate before they started to merge. All lines were frozen down and genotyped using PCR and Sanger sequencing (for genotyping primers, see Supplemental Table S1; for Sanger sequences, see Supplemental File 1). Lines were considered WT if only one WT band was found with PCR and sequencing confirmed no insertions/deletions were present. Lines were considered KO if only the band with the deletion was found and sequencing confirmed that the start codon was deleted. A selection of WT and KO lines was expanded for mRNA level detection with qPCR (for primers, see Supplemental Table S2) to select lines used for RNA-seq.

Plasmids

For ZNF91-GFP ChIP experiments pCAG.ZNF91 from Jacobs et al. (2014) was used to generate pCAG.ZNF91-Glycinelinker-eGFP. For CRISPR-Cas9 deletion experiments gRNAs were cloned into pX330-U6-Chimeric_BB-CBh-hSpCas9 (Addgene 42230) and pCAG.GFP (Addgene 11150) was cotransfected for FAC-sorting of hESCs.

Western blotting

A fraction of the GFP ChIP on ZNF91-GFP and GFP-transfected HEK293 cells was used for western blot to confirm the presence of ZNF91 protein. DTT (final concentration 100 mM) and 2× Laemmli sample buffer (Bio-Rad) were added to half of the sample. Samples were subjected to SDS-PAGE on 4%–20% mini protean precast protein gels (Bio-Rad) and transferred to a blotting membrane according to the manufacturer's protocol. Blots were washed three times in water and then blocked for 1 h in 5% milk powder in TBS-T on a rocking platform. Blots were incubated with rabbit anti-β-Tubulin (Cell Signaling Technology 2144, 1:1000) or rabbit anti-GFP (Abcam ab290, 1:5000) in TBS-T overnight at 4°C on a rotator. Blots were washed in TBS-T and then exposed to goat-anti-rabbit-HRP (Thermo Fisher Scientific 656120, 1:25000) in TBS-T, both for 1 h on a rocking platform. Blots were washed twice for 30 min in TBS-T on a rocking platform. Signal was visualized with

SuperSignal West Dura Trial Kit (Thermo Fisher Scientific) on Odyssey FC Imaging System (LI-COR Biosciences).

RNA sequencing

Samples of *ZNF91* KO, *ZNF611* KO, and *ZNF91/ZNF611* dKO were used for RNA sequencing. Total RNA was isolated using TRIzol according to the manufacturer's protocol, followed by a DNase treatment and RNA purification using the RNA clean and concentrator kit (Zymo research). Ribosomal RNA was depleted from total RNA with the rRNA depletion kit (NEB E6310) and subsequently prepared for RNA-seq with NEB Next Ultra Directional RNA Library Prep Kit (NEB E7420) at GenomeScan. Samples were sequenced at 150-bp paired end at an Illumina HiSeq 4000 device.

Mapping and analysis of RNA-seq data

A Snakemake (Köster and Rahmann 2012) pipeline was written to quality check, trim, map, and quantify RNA-seq data. The complete pipeline can be found at https://zenodo.org/record/ 2581199 and as Supplemental Code. A short description of the separate steps composing the pipeline follows. Paired-end Illumina RNA-seq FASTQ files were used as input. Read quality was assessed with FastQC. Trimmomatic (Bolger et al. 2014) was used to clip adapter sequences and trim low-quality reads. Reads were aligned to the human genome (hg19 version) using STAR (Dobin et al. 2013) with default settings except for the following: outFilterMismatchNmax = 2 and outFilterMultimapNmax = 10, outWigType = bedGraph, outSAMtype = BAM SortedByCoordinate. (Based on our verification that almost all SVA sequences are shared between hg19 and hg38, we state that mapping on hg38 instead of hg19 would not significantly alter the outcome or conclusions of our study.) We assessed in a representative RNA-seq sample that with 150-bp reads and paired-end sequencing, the percentage of multimapping reads that end up at SVAs is very low: Of all reads at SVAs, only 5.8% was derived from multimapping reads, which was not higher than the average 8% of multimapping reads genome wide. Because STAR assigns a multimapping read to only one, highest scoring place, we feel that this is the best approach for our research question. Base-by-base coverage tracks were generated from each bedGraph file with bedGraphToBigWig. Coverage tracks were displayed on UCSC Genome Browser while scaled using a scaling factor determined with DESeq2, including all RNA-seq samples in one session. Raw read counts of genes were determined with featureCounts (Liao et al. 2014) using hg19 KnownGenes.GTF from UCSC for annotation (https://genome.ucsc.edu/index.html; downloaded September 12, 2016). Only properly paired reads were counted (-B) and the library was reversely stranded (-s 2). Genes were summarized at metalevel (default), whereas TEs were summarized at the feature level (-f). StringTie (Pertea et al. 2015) was used to build de novo transcript models from *ZNF91* WT and KO hESC samples mapped with STAR (BAM files sorted by coordinate were used as input). Predicted transcripts were included with a minimum length of 100 bp (-m) and minimum read coverage of 1.5 (-c). One annotation file was generated from all separate StringTie transcript models that could be used by featureCounts to estimate raw counts of chimeric transcripts. DESeq2 (Love et al. 2014) was used to normalize raw counts, perform principal component analysis and differential expression analysis of genes, TEs, and transcripts assembled with StringTie. For differential expression analysis of SVAs, Log_2 FC > 3 was considered differentially expressed, because SVA expression levels were too low to determine adjusted *P*-values adequately. R packages ggplot2 and ggpubr were used to generate plots and perform statistical testing. Heatmaps were generated with MeV (vs 4.8.1).

Previously published RNA-seq data of human and rhesus macaque (*Macaca mulatta*) hESCs and cortical organoids was used (Field et al. 2019). Genes with an average basemean (DESeq) expression level below 30 over all analyzed human and macaque tissues were excluded to ensure only orthologous genes, expressed in both species, were studied. Transcripts without expression in any of the time points were excluded. Expression of transcripts with an SVA >1000 bp within 100 kb of their TSS (GSE106245_hg19.fantom.lv3) were compared to those without. Significant differences between human/rhesus expression ratios for genes with an SVA nearby versus all genes was tested using the Wilcoxon rank-sum test, and 95% confidence intervals of the mean were calculated by 10,000 times bootstrap of genes without an SVA nearby for a similar sized sample as those with an SVA nearby. Only results for cortical organoids shown.

Chromatin immunoprecipitation (ChIP)

HEK293 and *ZNF91* WT and KO hESCS were harvested the following ways: HEK293 cells, transfected with ZNF91-GFP or GFP two replicates each, were harvested from a confluent 60-mm dish by mechanical detachment with a cell scraper in cold PBS. One replicate of *ZNF91* WT and KO hESC colonies were detached from two 60-mm dishes with a cell scraper and collected at the bottom of a 15-mL tube through mild centrifugation (3 min at 400 rpm). Medium was replaced with 10 mL cold PBS. For all ChIP experiments, cells were cross-linked immediately after harvest by adding 1 mL of 11× cross-linking buffer (50 mM Tris-HCL pH 7.5, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 11% Formaldehyde) to 10 mL of cold cell suspension for 10 min at room temperature on a rocking platform. To quench formaldehyde, glycine was added at a final concentration of 0.11 M and samples were incubated for 5 min at room temperature on a rocking platform. Cells were pelleted through centrifugation for 5 min at 2000 rcf at 4°C. To remove all traces of cross-linking buffer, the cells were washed two times with 10 mL cold PBS. hESC samples were split in two for H3K4me3 and H3K27ac ChIP. For cell lysis, 1 mL lysis buffer 1 (50 mM Hepes-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100) supplemented with protease inhibitor cocktail (Roche) was added to cell pellets and incubated on a rocking platform for 10 min, GFP and H3K4me3 at 4°C, and H3K27ac at room temperature (RT). Cells were pelleted again by centrifugation (5 min at 2000 rcf at 4°C). Next, 1 mL lysis buffer 2 (10 mM Tris-HCL pH 8, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA) supplemented with protease inhibitor cocktail (Roche) was added to cell pellets, followed by incubation on a rocking platform for 5 min, GFP and H3K4me3 at 4°C, and H3K27ac at RT. Cells were pelleted again by centrifugation (5 min at 2000 rcf at 4°C). The pellet was resuspended in 100 μL lysis buffer 3 (10 mM Tris-HCL pH 8, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% Na-Deoxycholate, 0.5% N-Lauroylsarcosine) supplemented with protease inhibitor cocktail (Roche). Chromatin was sheared to fragments of ~500 bp in 12 cycles of sonication (intensity high, 30 sec on and 1 min off) in a Bioruptor. Then, 450 μL lysis buffer 3 and 50 μL 10% Triton X-100 were added to the sonicated cell lysate. Samples were centrifuged at maximum speed for 10 min at 4°C. Fifty microliters supernatant was stored at −20°C to be used as input sample, and the rest was used for chromatin immunoprecipitation (ChIP). Fifty microliters of Dynabeads M-280 sheep anti-rabbit IgG (Invitrogen) were washed three times with 0.5% BSA in PBS before incubating with 5 μg rabbit anti-GFP (Abcam ab290), 3.75 μL rabbit anti-H3K4me3 (Millipore 07-473, lot 2930138), or 5 μg rabbit anti-H3K27ac (Abcam ab4729, lot GR3205521-1) on a rotator for 4 h at 4°C. Excess antibody was removed by three wash steps with lysis buffer 3. Beads were resuspended in 500 μL

1% Triton X-100 in lysis buffer 3 and added to the lysate. IP was performed on a rotator overnight at 4°C. Immuno complexes were washed four times with 1 mL RIPA buffer (50 mM Hepes-KOH pH 7.5, 500 mM LiCl, 1 mM EDTA, 1% NP-40, 0.7% Na-Deoxycholate) supplemented with protease inhibitor cocktail (Roche) and once with 1 mL cold TBS (50 mM Tris-HCL pH7.6, 150 mM NaCl, GFP and H3K4me3) or TE (10 mM Tris-HCLpH8, 1 mM EDTA, H3K27ac) using a magnetic stand. After removal of all traces of TBS, the beads were resuspended in 200 µL elution buffer (50 mM Tris-HCLpH 8, 10 mM EDTA, 1% SDS). Then, 150 µL elution buffer was added to thawed input sample (50 µL). Chromatin was eluted and cross-linking reversed by incubating at 65°C on a rocking platform. Next, 200 µL TE (10 mM Tris-HCL pH 8, 1 mM EDTA) was added to each sample to dilute SDS in elution buffer. ChIP DNA was treated with Ambion RNase Cocktail (Invitrogen) for 30 min at 37°C and Proteinase K (Invitrogen) for 2 h at 55°C. DNA was extracted with one step phenol/chloroform and two steps chloroform. Ethanol (100%) was used to precipitate DNA for 45 min at −80°C. DNA was pelleted by centrifugation at 14,000 rpm for 45 min at 4°C. The pellet was washed with 70% ethanol and centrifuged at 14,000 rpm for 10 min at 4°C. DNA was dissolved in nuclease-free water. Finally, extracted DNA was purified using DNA Clean & Concentrator-5′ columns (Zymo). ChIP-seq for H3K9me3 on hESCs was performed according to the protocol described by Jacobs et al. (2014). For the ChIP, 5 µg of rabbit anti-H3K9me3 (Abcam ab8898) was used.

H3K9me3 ChIP-qPCR

Three replicates of *ZNF91* WT and KO hESC colonies were harvested by mechanical detachment from 60-mm dishes with a cell lifter and collected at the bottom of a 15-mL tube. Medium was replaced with 1 mL cold PBS. For all ChIP experiments, cells were cross-linked immediately after harvest by adding 100 µL of 11× cross-linking buffer (50 mM Tris-HCL pH 7.6, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 11% Formaldehyde) to 1 mL of cold cell suspension for 10 min at room temperature on a rocking platform. To quench formaldehyde, glycine was added at a final concentration of 0.11 M and samples were incubated for 5 min at room temperature on a rocking platform. Cells were pelleted through centrifugation at 500 rcf for 5 min at 4°C. To remove all traces of cross-linking buffer, the cells were washed two times with 1 mL cold PBS. For cell lysis, 1 mL lysis buffer 1 (50 mM Tris-HCL pH 7.6, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100) supplemented with protease inhibitor cocktail (Roche) was added to cell pellets. The cell suspension was transferred to a 1.7-mL low binding tube (Corning) and incubated on a rocking platform for 10 min at 4°C. Cells were pelleted again by centrifugation (5 min at 500 rcf at 4°C). Then, 1 mL lysis buffer 2 (10 mM Tris-HCL pH 8, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA) supplemented with protease inhibitor cocktail (Roche) was added to cell pellets, followed by incubation on a rocking platform for 10 min at 4°C. Cells were pelleted again by centrifugation (5 min at 500 rcf at 4°C). The pellet was resuspended in 100 µL lysis buffer 3 (10 mM Tris-HCL pH 8, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% Na-Deoxycholate, 0.5% N-Lauroylsarcosine) supplemented with protease inhibitor cocktail (Roche). Chromatin was sheared to fragments of ~500 bp in 12 cycles of sonication (intensity high, 30 sec on and 1 min off) in a bioruptor. Next, 550 µL lysis buffer 3 and 50 µL 10% Triton X-100 was added to the sonicated cell lysate. Samples were centrifuged at maximum speed for 10 min at 4°C. Fifty microliters supernatant was stored at −20°C to be used as input sample and the rest was used for chromatin immunoprecipitation (ChIP). Fifty microliters of Dynabeads M-280 sheep anti-rabbit IgG (Invitrogen, lot 00892017) were washed two times

with 0.5% BSA in PBS before incubating with 5 µg rabbit anti-H3K9me3 (Abcam ab8898, lot GR3365652-1) on a rotator for 6 h at 4°C. Excess antibody was removed by two wash steps with 500 µL lysis buffer 3. Beads were resuspended in 500 µL 1% Triton X-100 in lysis buffer 3 and added to the lysate. IP was performed on a rotator overnight at 4°C. Immuno complexes were washed four times with 500 µL RIPA buffer (50 mM Tris-HCL pH7.6, 500 mM LiCl, 1 mM EDTA, 1% NP-40, 0.7% Na-Deoxycholate) and once with 500 µL cold TE (10 mM Tris-HCL pH 8, 1 mM EDTA, 50 mM NaCl) using a magnetic stand and resuspended in 500 µL TE (without NaCl). The cell suspension in TE was transferred to a new 1.7-mL tube, the TE was removed, and the beads were resuspended in 210 µL elution buffer (50 mM Tris-HCL pH 8, 10 mM EDTA, 1% SDS). Then, 150 µL elution buffer was added to thawed input sample (50 µL). Chromatin was eluted and cross-linking reversed by incubating at 65°C on a rocking platform (Eppendorf, ThermoMixer C). Next, 200 µL TE (10 mM Tris-HCLpH8, 1 mM EDTA) was added to each sample to dilute SDS in the elution buffer. ChIP DNA was treated with Ambion RNase Cocktail (Invitrogen AM2286, lot 0906010) for 2 h at 37°C and Proteinase K (Invitrogen) for 2 h at 55°C at 550 rpm on a rocking platform. DNA was extracted with two steps phenol/chloroform and two steps chloroform. Ethanol (100%) was used to precipitate DNA overnight at −80°C. DNA was pelleted by centrifugation at 14,000 rpm for 45 min at 4°C. The pellet was washed with 70% ethanol and centrifuged at 14,000 rpm for 10 min at 4°C. DNA was dissolved in nuclease-free water. Finally, extracted DNA was purified using DNA Clean & Concentrator-5′ columns (Zymo). qPCR using the QuantiTect SYBR Green PCR kit (Roche) and 2× diluted DNA and DNA input control was performed on a Roche Lightcycler 480 II to check ChIP enrichment. The qPCR was performed by 5 min at 95°C, 40 cycles of 10 sec at 95°C, 30 sec at 60°C, and 30 sec at 72°C. Primers used for the qPCR are listed in Supplemental Table S3. The qPCR Ct values of the duplicates were analyzed by calculating the fold difference between ChIP DNA samples and DNA input control with the following formula: ΔCt [normalized ChIP] = Ct [ChIP] − (Ct [Input] − $\log_2$ (Input Dilution Factor)). Furthermore, the fold enrichment was calculated to normalize the background signal by dividing the ChIP DNA signal by the DNA input control signal (negative control, in this case, the LTR12C).

ChIP-seq library prep and sequencing

For TruSeq ChIP sample prep (Illumina, Lot 20194738), 25–30 ng of ChIP DNA was used as starting material. The provided protocol was followed with the following adjustments: DNA Clean & Concentrator-5 kit (Zymo) instead of beads was used for DNA purification. DNA fragment size selection with 2% agarose size selections gels (Invitrogen) using E-Gel SizeSelect II (Invitrogen) was done before and after PCR amplification of indexed ChIP DNA to obtain the 400–500 bp fraction. H3K4me3 and ZNF91-GFP ChIP samples were pooled and subjected to 75-bp paired-end sequencing at a depth of 130 million reads, which was performed by MAD:Dutch Genomics Service and Support Provider of the University of Amsterdam on an Illumina NextSeq 550 device. H3K27ac ChIP samples were pooled and subjected to 150-bp paired-end sequencing by GenomeScan at an Illumina HiSeq 4000 device. For H3K9me3 ChIP-seq on hESCs, library prep was described in Jacobs et al. (2014).

Mapping and analysis of ChIP-seq data

Next to our own ChIP-seq data we reanalyzed the ZNF611 ChIP-exo data published by Imbeault et al. (2017), accession number

GSE78099. A Snakemake (Köster and Rahmann 2012) pipeline was written to trim and map ChIP-seq data. The complete pipeline can be found at https://zenodo.org/record/2581325 and as Supplemental Code. The following is a short description of the separate steps of the pipeline, using paired-end Illumina sequencing FASTQ files as input. Read quality was assessed with FastQC. Trimmomatic (version 0.38) (Bolger et al. 2014) was used to clip adapter sequences and trim low-quality reads. Bowtie 2 (version 1.0.1) (Langmead and Salzberg 2012) in –end-to-end –very-sensitive mode was used to map reads to the human genome (hg19 version). Fragment length for valid paired-end fragments was set to a minimum of 80 (-I) and a maximum of 500 (-X). The resulting SAM files were converted to BAM files sorted by coordinate using SAMtools (version 1.7) (Li et al. 2009). The rmdup function of SAMtools was used to remove potential PCR duplicates from ChIP-seq samples. For mapping ChIP-seq data, we included multi-mapping reads, because this is now standard practice for analyzing ChIP-seq data with Bowtie 2 and widely done in our field (Imbeault et al. 2017; Fuentes et al. 2018). We used the default setting (search for multiple alignments, report the best one) with the --very-sensitive and --end-to-end parameters, to limit the inclusion of false positive/false negative read assignment. BEDTools genomecov and bedGraphToBigWig were used to generate unnormalized base-by-base bigWig files coverage tracks. For visualization on the UCSC Genome Browser, data was scaled according to *GAPDH* control regions. BamCoverage from the deepTools package (version 2.5.7) (Ramírez et al. 2016) was used to generate RPKM normalized base-by-base (--binSize 1) coverage bigWig file. BigwigCompare from the deepTools package was used to generate one coverage bigWig file displaying the mean of two ZNF91 ChIP replicates. Profile plots were generated with deepTools computeMatrix and plotHeatmap 2 kb around the center of SVAs using Galaxy (Galaxy deepTools version 3.1.2.0.1). Peak calling was done with MACS2 (version 2.1.1.20160309) (Zhang et al. 2008). All tags at a given location were included (--keep-dup all) and subpeaks were determined (--call-summits). We used the following cutoffs for ZNF91 ChIP: ZNF91 rep1 MACS score ≥ 50 and ZNF91 rep2 MACS score ≥ 100. Only peaks occurring in both replicates were included. This resulted in 7220 genuine ZNF91 peaks. For H3K4me3 and H3K27ac ChIPs, all peaks were included. Mapping of H3K9me3 ChIP-seq data was done according to steps described in Jacobs et al. (2014).

Motif analysis

MEME suite (version 5.0.4) (Bailey et al. 2009) was used for motif discovery (MEME), motif search (FIMO) (Grant et al. 2011) and motif comparison (TomTom). For motif discovery, sequences of peak summits extended with 50 bp to both sides were retrieved and used as input for MEME. For ZNF91 SVA motif discovery, SVAs were split in half and intersected with ZNF91 peaks. The top 200 peaks localizing to each half were used to generate two motifs. For promoter motif discovery, the top 200 of ZNF91 peaks overlapping with gene promoters (TSS plus/minus 1 kb) were used, respectively. The orientation of SVAs and genes were taken into consideration.

Data access

All raw and processed sequencing data generated in this study have been submitted to the NCBI Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) under accession number GSE162571.

Competing interest statement

Acknowledgments

References

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* **37:** W202–W208. doi:10.1093/nar/gkp335

Barchitta M, Quattrocchi A, Maugeri A, Vinciguerra M, Agodi A. 2014. LINE-1 hypomethylation in blood and tissue samples as an epigenetic marker for cancer risk: a systematic review and meta-analysis. *PLoS One* **9:** e109478. doi:10.1371/journal.pone.0109478

Batzer MA, Gudi VA, Mena JC, Foltz DW, Herrera RJ, Deininger PL. 1991. Amplification dynamics of human-specific (HS) alu family members. *Nucleic Acids Res* **19:** 3619–3623. doi:10.1093/nar/19.13.3619

Bernstein BE, Kamal M, Lindblad-Toh K, Bekiranov S, Bailey DK, Huebert DJ, McMahon S, Karlsson EK, Kulbokas EJ, Gingeras TR, et al. 2005. Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* **120:** 169–181. doi:10.1016/j.cell.2005.01.001

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30:** 2114–2120. doi:10.1093/bioinformatics/btu170

Bollati V, Schwartz J, Wright R, Litonjua A, Tarantini L, Suh H, Sparrow D, Vokonas P, Baccarelli A. 2009. Decline in genomic DNA methylation through aging in a cohort of elderly subjects. *Mech Ageing Dev* **130:** 234–239. doi:10.1016/j.mad.2008.12.003

Brouha B, Schustak J, Badge RM, Lutz-Prigge S, Farley AH, Moran JV, Kazazian HH Jr. 2003. Hot L1s account for the bulk of retrotransposition in the human population. *Proc Natl Acad Sci* **100:** 5280–5285. doi:10.1073/pnas.0831042100

Chen W, Schwalie PC, Pankevich EV, Gubelmann C, Raghav SK, Dainese R, Cassano M, Imbeault M, Jang SM, Russeil J, et al. 2019. ZFP30 promotes adipogenesis through the KAP1-mediated activation of a retrotransposon-derived *Pparg2* enhancer. *Nat Commun* **10:** 1809. doi:10.1038/s41467-019-09803-9

Chuong EB, Elde NC, Feschotte C. 2016. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science* **351:** 1083–1087. doi:10.1126/science.aad5497

Cordaux R, Batzer MA. 2009. The impact of retrotransposons on human genome evolution. *Nat Rev Genet* **10:** 691–703. doi:10.1038/nrg2640

Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, et al. 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci* **107:** 21931–21936. doi:10.1073/pnas.1016071107

Deng H, Le WD, Jankovic J. 2008. Genetic study of an American family with DYT3 dystonia (lubag). *Neurosci Lett* **448:** 180–183. doi:10.1016/j.neulet.2008.10.049

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29:** 15–21. doi:10.1093/bioinformatics/bts635

Farmiloe G, Lodewijk GA, Robben SF, van Bree EJ, Jacobs FMJ. 2020. Widespread correlation of KRAB zinc finger protein binding with

brain-developmental gene expression patterns. *Philos Trans R Soc B Biol Sci* **375:** 20190333. doi:10.1098/rstb.2019.0333

Faulkner GJ, Kimura Y, Daub CO, Wani S, Plessy C, Irvine KM, Schroder K, Cloonan N, Steptoe AL, Lassmann T, et al. 2009. The regulated retrotransposon transcriptome of mammalian cells. *Nat Genet* **41:** 563–571. doi:10.1038/ng.368

Field AR, Jacobs FMJ, Fiddes IT, Phillips APR, Reyes-Ortiz AM, LaMontagne E, Whitehead L, Meng V, Rosenkrantz JL, Olsen M, et al. 2019. Structurally conserved primate lncRNAs are transiently expressed during human cortical differentiation and influence cell-type-specific genes. *Stem Cell Reports* **12:** 245–257. doi:10.1016/j.stemcr.2018.12.006

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. doi:10.1371/journal.pone.0015082

Fuentes DR, Swigut T, Wysocka J. 2018. Systematic perturbation of retroviral LTRs reveals widespread long-range effects on human gene regulation. *eLife* **7:** e35989. doi:10.7554/eLife.35989

Grant CE, Bailey TL, Noble WS. 2011. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27:** 1017–1018. doi:10.1093/bioinformatics/btr064

Hancks DC, Ewing AD, Chen JE, Tokunaga K, Kazazian HH. 2009. Exon-trapping mediated by the human retrotransposon SVA. *Genome Res* **19:** 1983–1991. doi:10.1101/gr.093153.109

Huntley S, Baggott DM, Hamilton AT, Tran-Gyamfi M, Yang S, Kim J, Gordon L, Branscomb E, Stubbs L. 2006. A comprehensive catalog of human KRAB-associated zinc finger genes: insights into the evolutionary history of a large family of transcriptional repressors. *Genome Res* **16:** 669–677. doi:10.1101/gr.4842106

Imbeault M, Helleboid PY, Trono D. 2017. KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. *Nature* **543:** 550–554. doi:10.1038/nature21683

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

Jintaridth P, Mutirangura A. 2010. Distinctive patterns of age-dependent hypomethylation in interspersed repetitive sequences. *Physiol Genomics* **41:** 194–200. doi:10.1152/physiolgenomics.00146.2009

Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, et al. 2020. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581:** 434–443. doi:10.1038/s41586-020-2308-7

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. doi:10.1371/journal.pone.0173746

Kazazian HH, Wong C, Youssoufian H, Scott AF, Phillips DG, Antonarakis SE. 1988. Haemophilia A resulting from *de novo* insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature* **332:** 164–166. doi:10.1038/332164a0

Kobayashi K, Nakahori Y, Miyake M, Matsumura K, Kondo-Iida E, Nomura Y, Segawa M, Yoshioka M, Saito K, Osawa M, et al. 1998. An ancient retrotransposal insertion causes Fukuyama-type congenital muscular dystrophy. *Nature* **394:** 388–392. doi:10.1038/28653

Köster J, Rahmann S. 2012. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics* **28:** 2520–2522. doi:10.1093/bioinformatics/bts480

Kunarso G, Chia NY, Jeyakani J, Hwang C, Lu X, Chan YS, Ng HH, Bourque G. 2010. Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat Genet* **42:** 631–634. doi:10.1038/ng.600

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9:** 357–359. doi:10.1038/nmeth.1923

Li X, Ito M, Zhou F, Youngson N, Zuo X, Leder P, Ferguson-Smith AC. 2008. A maternal-zygotic effect gene, *Zfp57*, maintains both maternal and paternal imprints. *Dev Cell* **15:** 547–557. doi:10.1016/j.devcel.2008.08.014

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25:** 2078–2079. doi:10.1093/bioinformatics/btp352

Li X, Burton EM, Koganti S, Zhi J, Doyle F, Tenenbaum SA, Horn B, Bhaduri-McIntosh S. 2018. KRAB-ZFP repressors enforce quiescence of oncogenic human herpesviruses. *J Virol* **92:** e00298-18. doi:10.1128/JVI.00298-18

Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30:** 923–930. doi:10.1093/bioinformatics/btt656

Lippman Z, Gendrel AV, Black M, Vaughn MWMW, Dedhia N, Mccombie WRR, Lavine K, Mittal V, May B, Kasschau KDKD, et al. 2004. Role of transposable elements in heterochromatin and epigenetic control. *Nature* **430:** 471–476. doi:10.1038/nature02651

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15:** 550. doi:10.1186/s13059-014-0550-8

Najafabadi HS, Mnaimneh S, Schmitges FW, Garton M, Lam KN, Yang A, Albu M, Weirauch MT, Radovani E, Kim PM, et al. 2015. C2H2 zinc finger proteins greatly expand the human regulatory lexicon. *Nat Biotechnol* **33:** 555–562. doi:10.1038/nbt.3128

Nakamura Y, Murata M, Takagi Y, Kozuka T, Nakata Y, Hasebe R, Takagi A, Kitazawa Ji, Shima M, Kojima T. 2015. SVA retrotransposition in exon 6 of the coagulation factor IX gene causing severe hemophilia B. *Int J Hematol* **102:** 134–139. doi:10.1007/s12185-015-1765-5

Nichols BA, Oswald NW, McMillan EA, McGlynn K, Yan J, Kim MS, Saha J, Mallipeddi PL, LaDuke SA, Villalobos PA, et al. 2018. HORMAD1 is a negative prognostic indicator in lung adenocarcinoma and specifies resistance to oxidative and genotoxic stress. *Cancer Res* **78:** 6196–6208. doi:10.1158/0008-5472.CAN-18-1377

Nielsen AL, Ortiz JA, You J, Oulad-Abdelghani M, Khechumian R, Gansmuller A, Chambon P, Losson R. 1999. Interaction with members of the heterochromatin protein 1 (HP1) family and histone deacetylation are differentially involved in transcriptional silencing by members of the TIF1 family. *EMBO J* **18:** 6385–6395. doi:10.1093/emboj/18.22.6385

Notwell JH, Chung T, Heavner W, Bejerano G. 2015. A family of transposable elements co-opted into developmental enhancers in the mouse neocortex. *Nat Commun* **6:** 6644. doi:10.1038/ncomms7644

Oleksiewicz U, Gładych M, Raman AT, Heyn H, Mereu E, Chlebanowska P, Andrzejewska A, Sozańska B, Samant N, Fąk K, et al. 2017. TRIM28 and interacting KRAB-ZNFs control self-renewal of human pluripotent stem cells through epigenetic repression of pro-differentiation genes. *Stem Cell Reports* **9:** 2065–2080. doi:10.1016/j.stemcr.2017.10.031

Ostertag EM, Goodier JL, Zhang Y, Kazazian HH. 2003. SVA elements are nonautonomous retrotransposons that cause disease in humans. *Am J Hum Genet* **73:** 1444–1451. doi:10.1086/380207

Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* **33:** 290–295. doi:10.1038/nbt.3122

Pontis J, Planet E, Offner S, Turelli P, Duc J, Coudray A, Theunissen TW, Jaenisch R, Trono D. 2019. Hominoid-specific transposable elements and KZFPs facilitate human embryonic genome activation and control transcription in naive human ESCs. *Cell Stem Cell* **24:** 724–735.e5. doi:10.1016/j.stem.2019.03.012

Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. 2016. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44:** W160–W165. doi:10.1093/nar/gkw257

Rebollo R, Karimi MM, Bilenky M, Gagnier L, Miceli-Royer K, Zhang Y, Goyal P, Keane TM, Jones S, Hirst M, et al. 2011. Retrotransposon-induced heterochromatin spreading in the mouse revealed by insertional polymorphisms. *PLoS Genet* **7:** e1002301. doi:10.1371/journal.pgen.1002301

Riso V, Cammisa M, Kukreja H, Anvar Z, Verde G, Sparago A, Acurzio B, Lad S, Lonardo E, Sankar A, et al. 2016. ZFP57 maintains the parent-of-origin-specific expression of the imprinted genes and differentially affects non-imprinted targets in mouse embryonic stem cells. *Nucleic Acids Res* **44:** 8165–8178. doi:10.1093/nar/gkw505

Savage AL, Bubb VJ, Breen G, Quinn JP. 2013. Characterisation of the potential function of SVA retrotransposons to modulate gene expression patterns. *BMC Evol Biol* **13:** 101. doi:10.1186/1471-2148-13-101

Savage AL, Wilm TP, Khursheed K, Shatunov A, Morrison KE, Shaw PJ, Shaw CE, Smith B, Breen G, Al-Chalabi A, et al. 2014. An evaluation of a SVA retrotransposon in the *FUS* promoter as a transcriptional regulator and its association to ALS. *PLoS One* **9:** e90833. doi:10.1371/journal.pone.0090833

Schmitges FW, Radovani E, Najafabadi HS, Barazandeh M, Campitelli LF, Yin Y, Jolma A, Zhong G, Guo H, Kanagalingam T, et al. 2016. Multiparameter functional diversity of human C2H2 zinc finger proteins. *Genome Res* **26:** 1742–1752. doi:10.1101/gr.209643.116

Schultz DC, Friedman JR, Rauscher FJ. 2001. Targeting histone deacetylase complexes via KRAB-zinc finger proteins: the PHD and bromodomains of KAP-1 form a cooperative unit that recruits a novel isoform of the Mi-2α subunit of NuRD. *Genes Dev* **15:** 428–443. doi:10.1101/gad.869501

Schultz DC, Ayyanathan K, Negorev D, Maul GG, Rauscher FJ. 2002. SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes Dev* **16:** 919–932. doi:10.1101/gad.973302

Seah MKY, Wang Y, Goy PAV, Loh HM, Peh WJ, Loh DHP, Han BY, Wong E, Leong EL, Wolf G, et al. 2019. The KRAB-zinc finger protein ZFP708

mediates epigenetic repression at RMER19B retrotransposons. *Development* **146:** dev170266. doi:10.1242/dev.170266

Shahzad MMK, Shin YH, Matsuo K, Lu C, Nishimura M, Shen DY, Kang Y, Hu W, Mora EM, Rodriguez-Aguayo C, et al. 2013. Biological significance of HORMA domain containing protein 1 (HORMAD1) in epithelial ovarian carcinoma. *Cancer Lett* **330:** 123–129. doi:10.1016/j.canlet.2012.07.001

Sripathy SP, Stevens J, Schultz DC. 2006. The KAP1 corepressor functions to coordinate the assembly of de novo HP1-demarcated microenvironments of heterochromatin required for KRAB zinc finger protein-mediated transcriptional repression. *Mol Cell Biol* **26:** 8623–8638. doi:10.1128/MCB.00487-06

Sundaram V, Cheng Y, Ma Z, Li D, Xing X, Edge P, Snyder MP, Wang T. 2014. Widespread contribution of transposable elements to the innovation of gene regulatory networks. *Genome Res* **24:** 1963–1976. doi:10.1101/gr.168872.113

Szpakowski S, Sun X, Lage JM, Dyer A, Rubinstein J, Kowalski D, Sasaki C, Costa J, Lizardi PM. 2009. Loss of epigenetic silencing in tumors preferentially affects primate-specific retroelements. *Gene* **448:** 151–167. doi:10.1016/j.gene.2009.08.006

Takahashi N, Gray D, Strogantsev R, Noon A, Delahaye C, Skarnes WC, Tate PH, Ferguson-Smith AC. 2015. *ZFP57* and the targeted maintenance of postfertilization genomic imprints. *Cold Spring Harb Symp Quant Biol* **80:** 177–187. doi:10.1101/sqb.2015.80.027466

Takahashi N, Coluccio A, Thorball CW, Planet E, Shi H, Offner S, Turelli P, Imbeault M, Ferguson-Smith AC, Trono D. 2019. ZNF445 is a primary regulator of genomic imprinting. *Genes Dev* **33:** 49–54. doi:10.1101/gad.320069.118

Tao Y, Yen MR, Chitiashvili T, Nakano H, Kim R, Hosohama L, Tan YC, Nakano A, Chen PY, Clark AT. 2018. TRIM28-regulated transposon repression is required for human germline competency and not primed or naive human pluripotency. *Stem Cell Reports* **10:** 243–256. doi:10.1016/j.stemcr.2017.11.020

Thomas JH, Schneider S. 2011. Coevolution of retroelements and tandem zinc finger genes. *Genome Res* **21:** 1800–1812. doi:10.1101/gr.121749.111

Tie CH, Fernandes L, Conde L, Robbez-Masson L, Sumner RP, Peacock T, Rodriguez-Plata MT, Mickute G, Gifford R, Towers GJ, et al. 2018. KAP1 regulates endogenous retroviruses in adult human cells and contributes to innate immune control. *EMBO Rep* **19:** e45000. doi:10.15252/embr.201745000

Trizzino M, Park Y, Holsbach-Beltrame M, Aracena K, Mika K, Caliskan M, Perry GH, Lynch VJ, Brown CD. 2017. Transposable elements are the primary source of novelty in primate gene regulation. *Genome Res* **27:** 1623–1633. doi:10.1101/gr.218149.116

Trizzino M, Kapusta A, Brown CD. 2018. Transposable elements generate regulatory novelty in a tissue-specific fashion. *BMC Genomics* **19:** 468. doi:10.1186/s12864-018-4850-3

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

Van der Klift HM, Tops CM, Hes FJ, Devilee P, Wijnen JT. 2012. Insertion of an SVA element, a nonautonomous retrotransposon, in *PMS2* intron 7 as a novel cause of Lynch syndrome. *Hum Mutat* **33:** 1051–1055. doi:10.1002/humu.22092

Wang H, Xing J, Grover D, Hedges DJ, Han K, Walker JA, Batzer MA. 2005. SVA elements: a hominid-specific retroposon family. *J Mol Biol* **354:** 994–1007. doi:10.1016/j.jmb.2005.09.085

Wang Z, Zang C, Rosenfeld JA, Schones DE, Barski A, Cuddapah S, Cui K, Roh TY, Peng W, Zhang MQ, et al. 2008. Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet* **40:** 897–903. doi:10.1038/ng.154

Watkins J, Weekes D, Shah V, Gazinska P, Joshi S, Sidhu B, Gillett C, Pinder S, Vanoli F, Jasin M, et al. 2015. Genomic complexity profiling reveals that HORMAD1 overexpression contributes to homologous recombination deficiency in triple-negative breast cancers. *Cancer Discov* **5:** 488–505. doi:10.1158/2159-8290.CD-14-1092

Wilund KR, Yi M, Campagna F, Arca M, Zuliani G, Fellin R, Ho YK, Garcia JV, Hobbs HH, Cohen JC. 2002. Molecular mechanisms of autosomal recessive hypercholesterolemia. *Hum Mol Genet* **11:** 3019–3030. doi:10.1093/hmg/11.24.3019

Wolf D, Goff SP. 2009. Embryonic stem cells use ZFP809 to silence retroviral DNAs. *Nature* **458:** 1201–1204. doi:10.1038/nature07844

Wolf G, Greenberg D, Macfarlan TS. 2015. Spotting the enemy within: targeted silencing of foreign DNA in mammalian genomes by the Krüppel-associated box zinc finger protein family. *Mob DNA* **6:** 17. doi:10.1186/s13100-015-0050-8

Yan FJ, Fan J, Huang Z, Zhang JJ. 2017. ZNF300 tight self-regulation and functioning through DNA methylation and histone acetylation. *Cell Biosci* **7:** 33. doi:10.1186/s13578-016-0129-z

Yang P, Wang Y, Macfarlan TS. 2017. The role of KRAB-ZFPs in transposable element repression and mammalian evolution. *Trends Genet* **33:** 871–881. doi:10.1016/j.tig.2017.08.006

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. 2008. Model-based Analysis of ChIP-Seq (MACS). *Genome Biol* **9:** R137. doi:10.1186/gb-2008-9-9-r137