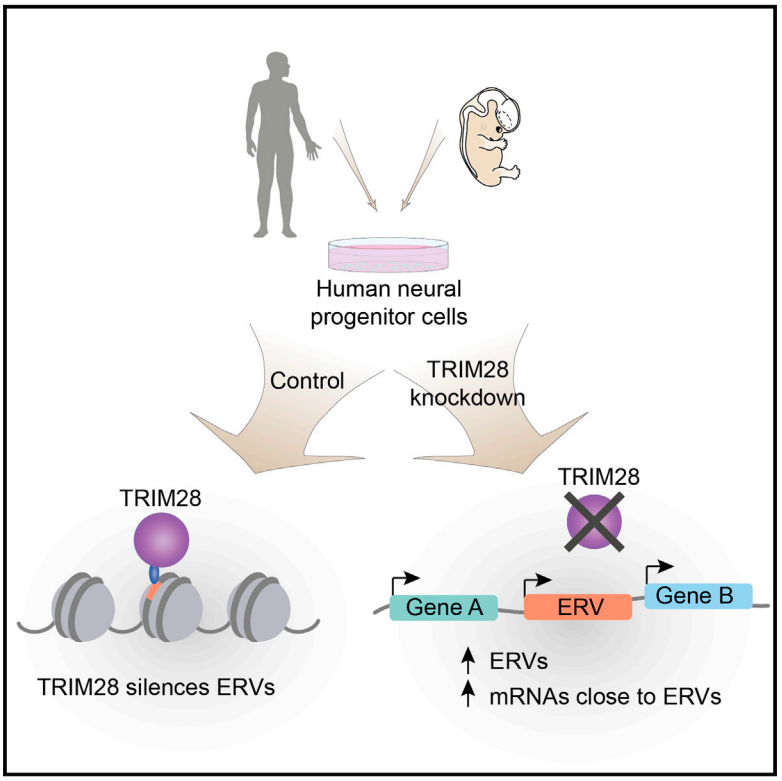
Cell Reports

Report

TRIM28 Controls a Gene Regulatory Network Based on Endogenous Retroviruses in Human Neural Progenitor Cells

Graphical Abstract



Authors

Per Ludvik Brattås, Marie E. Jönsson, Liana Fasching, ..., Patric Jern, Malin Parmar, Johan Jakobsson

Correspondence

johan.jakobsson@med.lu.se

In Brief

Brattås et al. report that endogenous retroviruses are bound by TRIM28 in human neural progenitor cells. This results in the establishment of local heterochromatin that affects nearby gene expression, suggesting a role for endogenous retroviruses in the control of transcriptional networks in the developing human brain.

Highlights

- Stage- and region-specific expression of ERVs during human brain development
- TRIM28 binds to ERVs and induces hetereochromatin in human neural progenitor cells
- Knockdown of TRIM28 in hNPCs results in the upregulation of ERV expression
- Protein-coding genes located near upregulated ERVs are upregulated

Accession Numbers

GSE84259

Brattås et al., 2017, Cell Reports *18*, 1–11
January 3, 2017
http://dx.doi.org/10.1016/j.celrep.2016.12.010

CellPress

Cell Reports
Report

TRIM28 Controls a Gene Regulatory Network Based on Endogenous Retroviruses in Human Neural Progenitor Cells

Per Ludvik Brattås,[1] Marie E. Jönsson,[1] Liana Fasching,[1] Jenny Nelander Wahlestedt,[4] Mansoureh Shahsavani,[2] Ronny Falk,[2] Anna Falk,[2] Patric Jern,[3] Malin Parmar,[4] and Johan Jakobsson[1,5,*]
[1]Laboratory of Molecular Neurogenetics, Department of Experimental Medical Science, Wallenberg Neuroscience Center and Lund Stem Cell Center, BMC A11, Lund University, 221 84 Lund, Sweden
[2]Department of Neuroscience, Karolinska Institute, Retziusväg 8, 171 77 Stockholm, Sweden
[3]Science for Life Laboratory, Department of Medical Biochemistry and Microbiology, Uppsala University, 751 23 Uppsala, Sweden
[4]Developmental and Regenerative Neurobiology, Department of Experimental Medical Science, Wallenberg Neuroscience Center and Lund Stem Cell Center, BMC A11, Lund University, 221 84 Lund, Sweden
[5]Lead Contact
*Correspondence: johan.jakobsson@med.lu.se
http://dx.doi.org/10.1016/j.celrep.2016.12.010

SUMMARY

Endogenous retroviruses (ERVs), which make up 8% of the human genome, have been proposed to participate in the control of gene regulatory networks. In this study, we find a region- and developmental stage-specific expression pattern of ERVs in the developing human brain, which is linked to a transcriptional network based on ERVs. We demonstrate that almost 10,000, primarily primate-specific, ERVs act as docking platforms for the co-repressor protein TRIM28 in human neural progenitor cells, which results in the establishment of local heterochromatin. Thereby, TRIM28 represses ERVs and consequently regulates the expression of neighboring genes. These results uncover a gene regulatory network based on ERVs that participates in control of gene expression of protein-coding transcripts important for brain development.

INTRODUCTION

The complexity of human brain development differs markedly from other mammals and is thought be important for the emergence of higher cognitive functions. However, the precise genetic changes, as well as the existence of human-specific gene regulatory networks underlying the evolution of the human brain, remain poorly explored (Vallender et al., 2008). Most of our knowledge about human brain development is restricted to evolutionarily conserved developmental pathways, whereas much less is known about primate- and human-specific developmental mechanisms. Identification of further mechanisms that regulate human brain development is important for our understanding of the human brain and may also provide additional links to the biology of human brain disorders.

Several recent studies suggest that transposable elements (TEs), which make up more than one-half of the human DNA, have the capacity to establish primate-specific gene regulatory networks. TEs are mobile genetic sequences with potential to alter the genetic landscape and to influence gene expression from their integration sites in the host genome by providing various *cis*-regulatory elements including promoters, repressors, enhancers, and insulators. It is becoming increasingly clear that TEs are adapted to influence gene expression through their regulatory sequences and that they play important roles in controlling and fine-tuning host gene networks (Cowley and Oakey, 2013).

About 8% of the human genome is composed of endogenous retroviruses (ERVs). These sequences are derived from retroviruses that have invaded vertebrate hosts for millions of years leaving traces as inherited ERVs through germline infection and subsequent transposition (Stoye, 2012). Several studies have found that ERV transcription is tightly controlled at multiple levels in early human development and human pluripotent stem cells (hPSCs), including positive regulation by a unique combination of transcription factors (Bourque et al., 2008; Santoni et al., 2012; Wang et al., 2014). There are also several repressive transcriptional regulators that silence ERV transcription through the formation of heterochromatin (Rowe and Trono, 2011). In addition, certain ERVs act as long non-coding RNAs (lncRNAs) in hPSCs, are expressed in a cell-type-specific manner in the two-, four-, eight-cell, and blastocyst stage, and expression of certain ERVs can be used to identify naive-like hPSCs populations (Göke et al., 2015; Lu et al., 2014; Wang et al., 2016). These studies demonstrate the existence of a dynamic and cell-type-specific transcriptional control of ERVs during early development and indicate that these elements play a critical role in this process.

Because the ERV repertoire is highly divergent across host species (Hayward et al., 2013, 2015), it is possible that ERVs contribute to species-specific regulatory networks. ERVs have colonized mammalian genomes during at least 100 million years with an apparent peak of insertions around 35–45 million years ago, at the time of the split between Old and New World

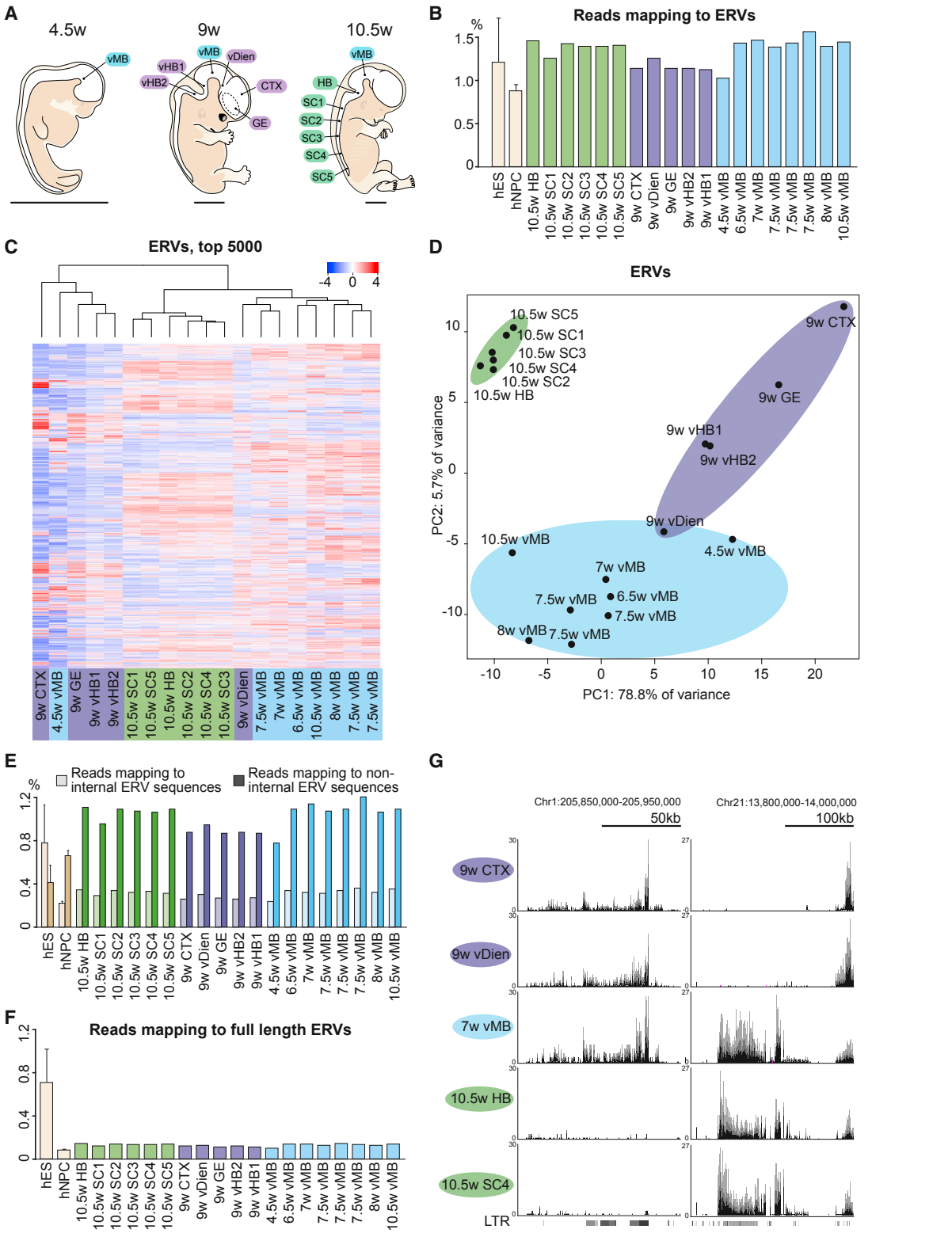


Figure 1. Stage- and Region-Specific Expression of ERVs in the Developing Human Brain

(A) Schematic drawings of dissected human embryos.

(B) Fraction of RNA-seq reads mapping to ERVs out of all reads mapping to the genome.

(C) The 5,000 ERVs with the largest variation are shown in a heat map, visualizing the dynamic expression of ERVs in the embryonic brain.

(legend continued on next page)

monkeys (Bannert and Kurth, 2006). This invasion results in a largely primate-specific makeup of ERVs in the human genome, and it is therefore tempting to speculate that ERVs have participated in the evolution of primates and humans, including the emergence of the complex human brain development. In addition, several studies have found aberrant transcriptional activation of ERVs in a number of neurological disorders (e.g., amyotrophic lateral sclerosis [ALS], schizophrenia, and bipolar disorders) that are thought to be, in part, the result of neurodevelopmental alterations (Christensen, 2016). It is possible that ERVs serve important roles as regulatory elements in brain development controlling gene networks that become dysregulated in diseases. These roles are, however, poorly explored both in terms of normal brain development and disease states.

In this study, we find a region- and developmental stage-specific expression pattern of ERVs in the developing human embryonic CNS, which is linked to a gene regulatory network involving ERVs. We demonstrate that several thousand ERVs, many that are primate specific, act as docking platforms for the epigenetic co-repressor protein TRIM28, which results in the establishment of local heterochromatin around these ERVs. This repressive transcriptional network modulates expression of protein-coding transcripts important for brain development, thereby providing an additional layer of transcriptional regulation.

RESULTS

Stage-Specific ERV Expression in Human Brain Development

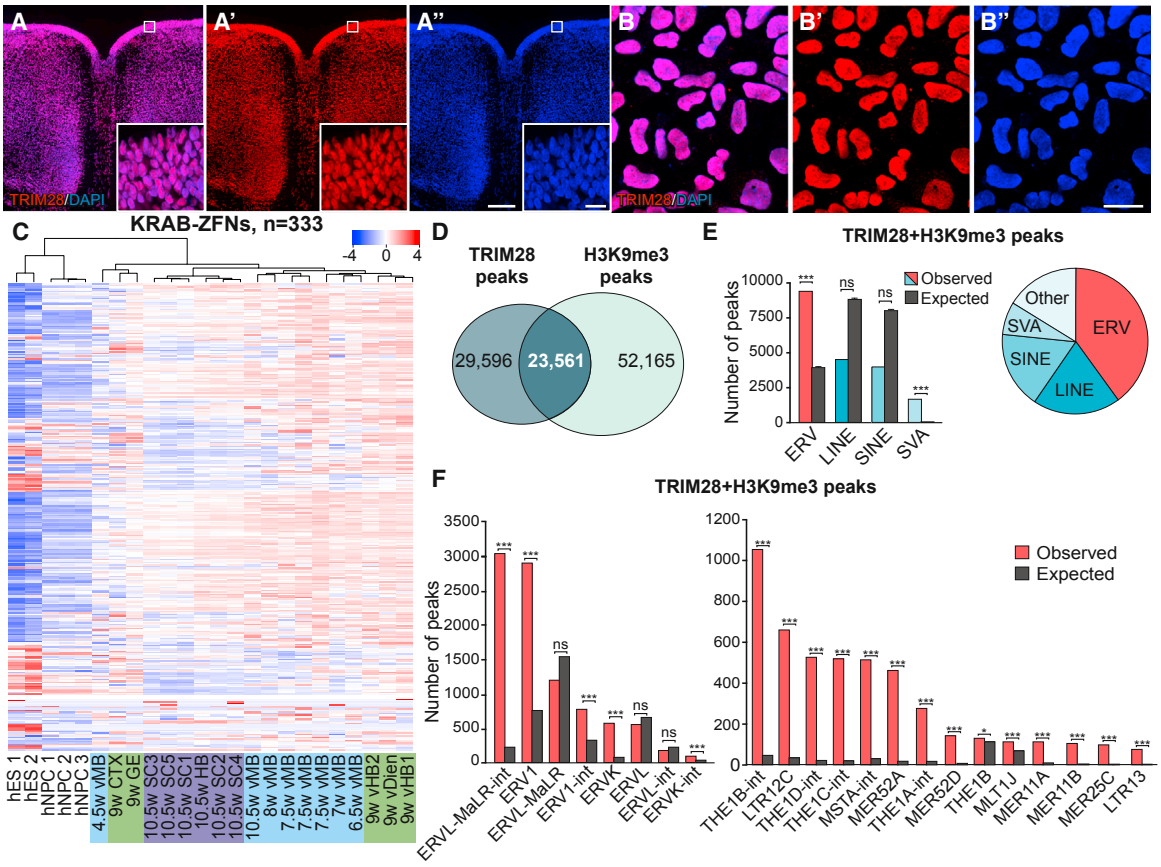
To investigate ERV expression patterns in human brain development, we sub-dissected and sampled sections from different rostrocaudal levels of the developing CNS from human embryos at developmental stages spanning from onset of neurogenesis to peak production of neurons (Figure 1A and Table S1). We included 19 samples from ten embryos in a poly(A)-selected RNA sequencing (RNA-seq) transcriptome analysis. The resulting 1,164,889,003 sequence reads were mapped to the human genome and reads for all of the approximately 750,000 ERV elements identified by RepeatMasker (class "LTR") were quantified.

We calculated the relative abundance of ERV elements identified by RepeatMasker (class "LTR") and found that ERV transcripts comprised 1.0%–1.5% of total mapped reads in all samples (Figure 1B). ERVs overlapping annotated protein-coding exons were excluded from the analysis. This fraction is slightly higher than in human embryonic stem cells (hESCs) and cultured human neural progenitor cells (hNPCs) and in the same range as previously described in early human development (Göke et al., 2015). Similarly, we found that other classes of TEs, such as LINE-1, Alu, and SVA elements, were also highly expressed in the embryonic brain samples (Figure S1A).

We could verify that many different ERV loci were expressed (25,733 showing more than five reads/element in at least one sample) and demonstrated a dynamic expression across samples (Figure 1C). To further investigate whether ERV expression is specific to developmental stage or region, we performed a principal-component analysis (PCA). Remarkably, we found that the two first principal components clearly separated both developmental stage and region (Figure 1D). By including ventral midbrain samples derived from embryos of different gestational stages, we performed a detailed time course analysis confirming a stage-dependent expression. We also analyzed several closely located structures in the same embryo, such as hindbrain and spinal cord (gestational age: 10.5 weeks) as well as cortex, ganglionic eminence, and diencephalon (gestational age: 9 weeks). The PCA demonstrated that samples from closely located structures derived from the same individual display a region-specific ERV expression (Figure 1D). It is worth noting that the PC1 component, which contains 78.8% variance, provides a good distinction between samples and corresponds to the level of ongoing neurogenesis: on one end, samples with a large proportion of post-mitotic neurons (samples from 10.5-week-old embryos) and, on the other end, samples with a large proportion of expanding neural progenitors (cortex from a 9-week-old embryo). This demonstrates a dynamic ERV expression during neurogenesis when progenitors differentiate into neurons. A similar pattern was also observed when performing a cluster analysis including SINE, LINE, and SVA, revealing a highly region- and stage-specific expression of all main groups of TEs in the embryonic brain (Figures S1B and S1C).

When we analyzed ERV expression in detail, we found distinct differences between pluripotent hESCs, which correspond to a developmental stage before germline commitment, and samples obtained from human embryonic brain. We observed that, although the majority of reads in neural cells originate from non-internal ERV fragments, hESCs transcribe a large number of internal ERV fragments ("-int" or "-I" annotation in Repbase) (Figure 1E), indicating that more complete ERV loci are primarily expressed in hESC, whereas long terminal repeat (LTR) fragments dominate in embryonic brain samples. We therefore used the RetroTector software (Sperber et al., 2007) to identify ERVs with relatively complete structure, and found 3,841 full-length ERVs (FL-ERVs) in the reference human genome (GRCh38; Table S2). Interestingly, the expression of these FL-ERV loci was dramatically higher in hESC than in samples from the developing human brain (Figure 1F). Because the majority of ERVs expressed during human brain development were incomplete fragments, these loci may primarily be passively expressed due to their position in a transcriptionally active genomic region. Indeed, we found several truncated ERV fragments, with

(D) PCA using the 500 ERVs with the largest variance in the embryonic samples shows region- and stage-specific expression of ERVs.
(E) Fraction of reads mapping to internal versus non-internal ERV fragments out of all reads mapping to the genome.
(F) Fraction of reads mapping to FL-ERVs out of all reads mapping to the genome.
(G) Read coverage at clusters of ERVs with highly region- and stage-specific expression patterns. LTR track: RepeatMasker class LTR. vMB, ventral midbrain; vHB, ventral hindbrain; vDien, ventral diencephalon; CTX, cortex; GE, ganglionic eminences; SC, spinal cord. See also Figure S1 and Tables S1 and S2.
In (B), (E), and (F), the data are shown as the mean fractions of three and two replicates for hNPC and hES, respectively. The error bars show the 95% confidence limits.



(legend on next page)

high and region-specific expression that were parts of longer endogenous transcripts (Figure 1G).

Together, these data demonstrate a high-level and dynamic expression of ERVs during human brain development, which is controlled independently from protein-coding genes at different developmental stages. The data also demonstrate that the transcriptional control of ERVs in the developing brain is markedly different when compared to pluripotent stem cells resulting in the expression of primarily ERV fragments in the human embryonic CNS. This suggests the existence of a dynamic transcriptional repressor system with the capacity to induce local heterochromatin on and around ERVs at different stages during human brain development.

TRIM28 and KRAB-ZFPs Are Highly Expressed during Human Brain Development

KRAB-zinc finger proteins (KRAB-ZFPs), a family of transcription factors that has undergone a rapid expansion in mammalian genomes in parallel with the expansion of ERVs (Thomas and Schneider, 2011), and their associated epigenetic co-repressor protein TRIM28 (also known as KAP1 or TIF1B) represent a candidate repressor system, which previously has been implicated in the silencing of ERV expression in hPSCs and during early development (Rowe et al., 2010; Turelli et al., 2014). KRAB-ZFPs recognize DNA sequences through zinc finger domains and recruit TRIM28 through the KRAB domain (Wolf and Goff, 2009). The role of this transcriptional repressor system in human brain development has not been investigated previously.

Immunohistochemistry on sectioned human embryonic tissue confirmed high expression of TRIM28 throughout the developing human brain and in cultured hNPCs (Figures 2A and 2B). RNA-seq reads from human embryonic tissue showed that most KRAB-ZFPs were expressed during human brain development, and many of these showed a region- and stage-dependent expression profile (Figure 2C). Thus, differences in KRAB-ZFP expression pattern and repression patterns of different ERVs in various cell types during brain development could explain the dynamic expression of ERVs during human brain development.

TRIM28 Binds Directly to ERVs in hNPCs

The recruitment of TRIM28 to ERVs by KRAB-ZFPs results in the formation of local heterochromatin characterized by the presence of the repressive histone mark H3K9me3 that can spread into adjacent DNA (Groner et al., 2010). To investigate whether TRIM28 directly controls ERVs in hNPCs, we therefore performed chromatin immunoprecipitation sequencing (ChIP-seq) for TRIM28 and H3K9me3. The hNPCs used for this experiment were homogeneous cultures of induced pluripotent stem cell (iPS)-derived neuroepithelial-like stem cells, which can be extensively expanded as homogeneous progenitor cell cultures and also differentiated to homogeneous neuronal cultures, with a high amount (90%) of neurons (Falk et al., 2012). These cells are positive for NPC markers, and negative for pluripotency markers (Figures S2A–S2H). We found high expression of ERV fragments in hNPCs (Figures 1E and 1F), at similar levels as the embryonic brain. hNPCs also displayed dynamic expression of both ERVs and KRAB-ZFPs during neuronal differentiation, making them a suitable culture model to study the role of TRIM28 in human brain development (Figures S2I and S2J).

The ChIP-seq analysis identified more than 23,000 genomic regions that were co-occupied by TRIM28 and H3K9me3 in hNPCs (Figure 2D). Remarkably, over 85% of the TRIM28-H3K9me3 peaks were located in TEs, with a highly significant enrichment in ERV and SVA binding (Figure 2E). Using Repbase annotations, we also identified multiple ERV superfamilies and subfamilies that were significantly enriched for TRIM28-H3K9me3 binding (Figures 2F and 2G).

We analyzed TRIM28 binding of the FL-ERVs (mapped with RetroTector; Table S2) and found that 482 out of 3,841 were co-bound by TRIM28 and H3K9me3. We performed a phylogenetic analysis (as described in Hayward et al., 2013, 2015) based on the 778 FL-ERVs with sufficient sequence information for the analysis, of which 98 were co-bound by TRIM28 and H3K9me3, to investigate whether sequence relationship of FL-ERVs influence TRIM28 binding, and found that TRIM28 binds and induces heterochromatin at FL-ERVs of several distinct ERV-lineages (Figure 2H). Taken together, these data show that TRIM28 binds many different types of ERV elements in hNPCs, including full-length elements as well as short fragments, and indicates that many KRAB-ZFPs participate in this network.

To investigate whether TRIM28 binding is related to the evolutionary age of ERVs, we estimated the age of the FL-ERVs, by computing sequence divergence between 5′- and 3′-LTRs, because LTRs are identical at the time of integration into the host DNA and accumulate neutral mutations at a rate of about 0.2% per million year (Hardison et al., 2003). Interestingly, we found that the majority of FL-ERVs co-bound by TRIM28 and H3K9me3 are around 35–55 million years old. This corresponds

Figure 2. TRIM28 Directly Binds ERVs in hNPCs

(A) Immunohistochemistry shows TRIM28 expression in the hindbrain from an 8-week-old human embryo.

(B) Immunocytochemistry reveals TRIM28 expression in expanding hNPCs.

(C) A heat map of the expression of KRAB-ZFPs shows a dynamic expression in pluripotent stem cells, hNPCs, and the embryonic brain.

(D) A total of 23,561 out of 53,122 TRIM28 ChIP-seq peaks overlapped with H3K9me3 peaks.

(E) TRIM28-H3K9me3 co-peaks at different classes of TEs, versus shuffled coordinates (10,000 bootstraps). The data are shown as mean number of expected binding events in the 10,000 bootstraps. Error bars show SD.

(F) TRIM28-H3K9me3 binding to various superfamilies (left) and subfamilies (right) of ERVs, according to Repbase nomenclature.

(G) Read coverage of TRIM28 and H3K9me3 ChIP-seq reads at two FL-ERVs (FL-ERV-track, detected by RetroTector) and a THE1C element. LTR track, RepeatMasker; ChIP, intervals of TRIM28 and H3K9me3-peaks.

(H) Phylogenetic relationship of 778 FL-ERVs. FL-ERVs bound by TRIM28-H3K9me3 in hNPCs are marked by red.

(I) The age distribution of FL-ERVs bound by TRIM28-H3K9me3 versus all FL-ERVs, where FL-ERVs were identified using RetroTector.

(J) The age distribution of ERVs bound by TRIM28-H3K9me3 versus all ERVs, where ERVs were identified using RepeatMasker. Scale bars represent 200 μm in (A) and 20 μm in both (B) and insets in (A). See also Figures S2 and S3 and Table S2.

to the time of the split between New World and Old World monkeys (33–57 mya) and the divergence of hominoids from the Old World monkeys (30–45 mya) (Figure 2I) (Bannert and Kurth, 2006). To complement this analysis, we also estimated ages by calculating divergence from consensus sequence in Repbase and found a similar age distribution of co-bound ERV elements (Figure 2J). We also found a similar enrichment of TRIM28-H3K9me3 binding to of primate-specific LINE-1 and SVA subfamilies (Figures S3A and S3B). Together, these results demonstrate that TRIM28 binds and induces heterochromatin at a diverse set of primarily primate-specific ERVs and other TEs in hNPCs.

Loss of TRIM28 in hNPCs Results in ERV Transcriptional Activation

To assess the functional role of TRIM28, we performed knockdown (KD) experiments in hNPCs using a small hairpin RNA (shRNA) lentiviral vector (Figures 3A and S4). *TRIM28*-KD resulted in efficient loss of TRIM28 at both the RNA and protein level (Figures 3B and 3C). RNA-seq of *TRIM28*-KD and control NPCs identified significantly increased expression of 411 ERVs following *TRIM28*-KD, whereas only 72 ERVs showed decreased expression (Benjamini-Hochberg [BH]-corrected p value < 0.05), demonstrating that TRIM28 represses the expression of multiple ERVs in hNPCs (Figure 3D).

In many cases, TRIM28-H3K9me3-co-bound ERVs caused upregulation of adjacent ERVs (Figure 3E, left panel). In addition, we found that activation of some ERVs resulted in the formation of lncRNAs, extending up to 100 kb into the adjacent genomic sequence (Figure 3E, right panel). Interestingly, a relatively small fraction (~11%) of all upregulated ERVs were directly co-bound by TRIM28-H3K9me3 (Figure 3F), and less than 1% (47/9,540) of all TRIM28-H3K9me3-bound ERVs were upregulated. However, almost all upregulated ERVs were either covered by or close to H3K9me3 and more than 81% (335/411) were located within 50 kb of a H3K9me3-TRIM28 peak, of which most were bound to other ERVs or TEs (Figures 3G and 3H). In summary, these data show that loss of TRIM28 binding results in transcriptional activation of a few directly bound ERVs as well many other ERVs located in close vicinity to TRIM28-bound ERVs.

In addition to ERVs, we found that several other classes of TEs, including many LINE-1 elements, were upregulated following *TRIM28*-KD (Figure S3C). Also in this case, the majority of upregulated LINE-1 elements were not always directly bound by TRIM28 but were often in close vicinity to an H3K9me3-TRIM28-bound TE (Figure S3D).

We validated these results by performing a complete replication experiment, including RNA-seq analysis, using an additional hNPCs cell line, which was derived from human embryonic brain tissue. We found similar results when performing *TRIM28*-KD in this hNPC line including upregulation of mostly the same ERV loci (Figures S3E and S3F). Together, these data suggest that TRIM28-bound TEs form a localized heterochromatin region characterized by silencing of nearby transcription.

Transcriptional Activation of ERVs Is Coupled with Increased Expression of Protein-Coding Genes

We next investigated whether TRIM28 binding to ERVs also influences the expression of nearby protein-coding genes. Remarkably, we found that out of 163 protein-coding genes located within 50 kb of an upregulated ERV, 93 were also upregulated whereas only 6 genes were downregulated (BH-corrected p value < 0.05) (Figures 4A and 4B). In addition, we found that 22 of 31 genes within 50 kb from upregulated ERVs that were also co-bound by TRIM28-H3K9me3 were significantly upregulated, whereas only one gene was downregulated. These changes were significantly different from genes located near non-affected ERVs after *TRIM28*-KD ($p < 0.0001$, Wilcoxon rank sum test; Figure 4C). It is worth noting that the transcriptional impact of ERV elements on nearby genes was rather modest, with on average a 2-fold change in expression level (Figure 4C), suggesting that this network fine-tunes gene expression rather than strictly switching genes on or off. There was no clear sense versus anti-sense or upstream versus downstream bias of the ERVs relative to the genes (Figure 4D), suggesting that the transcriptional activation is primarily caused by the removal of repressive epigenetic marks and not by the ERV acting as an alternative promoter.

Together, these data demonstrate a striking and highly significant correlation between the expression of TRIM28-controlled ERV elements and almost 100 nearby protein-coding genes (Table S3). Some of these genes are likely to have an important role during brain development such as *BMP3* (Figure 4E), which is a conserved morphogen that may play a role in brain development (Yamashita et al., 2016). *BMP3* has three TRIM28-bound ERVL-MaLR elements (THE1C, THE1, and THEB) located just upstream of the open reading frame (Figure 4E). We also found that transcriptional regulators, such as *TRIM24*, and genes related to apoptosis, such as *STK17B* (Figure 4E), are under the control of the TRIM28-ERV network. We confirmed that the same set of genes was upregulated in an additional hNPC line derived from embryonic brain tissue and also confirmed the upregulation of *BMP3* and *STK17B* with qRT-PCR (Figure 4F). Together, these data support that an ERV-based network influences the expression level of genes with important regulatory functions during human brain development.

Knockdown of *TRIM28* is likely to cause effects on the transcriptome that is unrelated to ERVs, and a key question is therefore the functional relevance of the TRIM28-ERV network. We confirmed the broad influence of TRIM28-controlled ERVs on adjacent gene expression, as our global assessment of upregulated mRNAs after *TRIM28*-KD ($n = 3466$, BH-corrected p value < 0.05) showed that about one-half of these loci were close to a TRIM28-H3K9me3 peak, frequently associated with TE binding (Figure 4G). It is noteworthy that gene ontology analysis showed that downregulated genes are primarily associated with neuronal development and differentiation, whereas upregulated genes are related to other developmental fates, thereby linking TRIM28 to transcriptional control of neurodevelopmental gene programs (Figure 4H). Together, these data suggest that ERV-based repression mediate significant changes in gene expression that influence transcriptional programs related to brain development.

DISCUSSION

In this study, we demonstrate that TRIM28 binds to ERVs in hNPCs and mediates local heterochromatin that spreads in

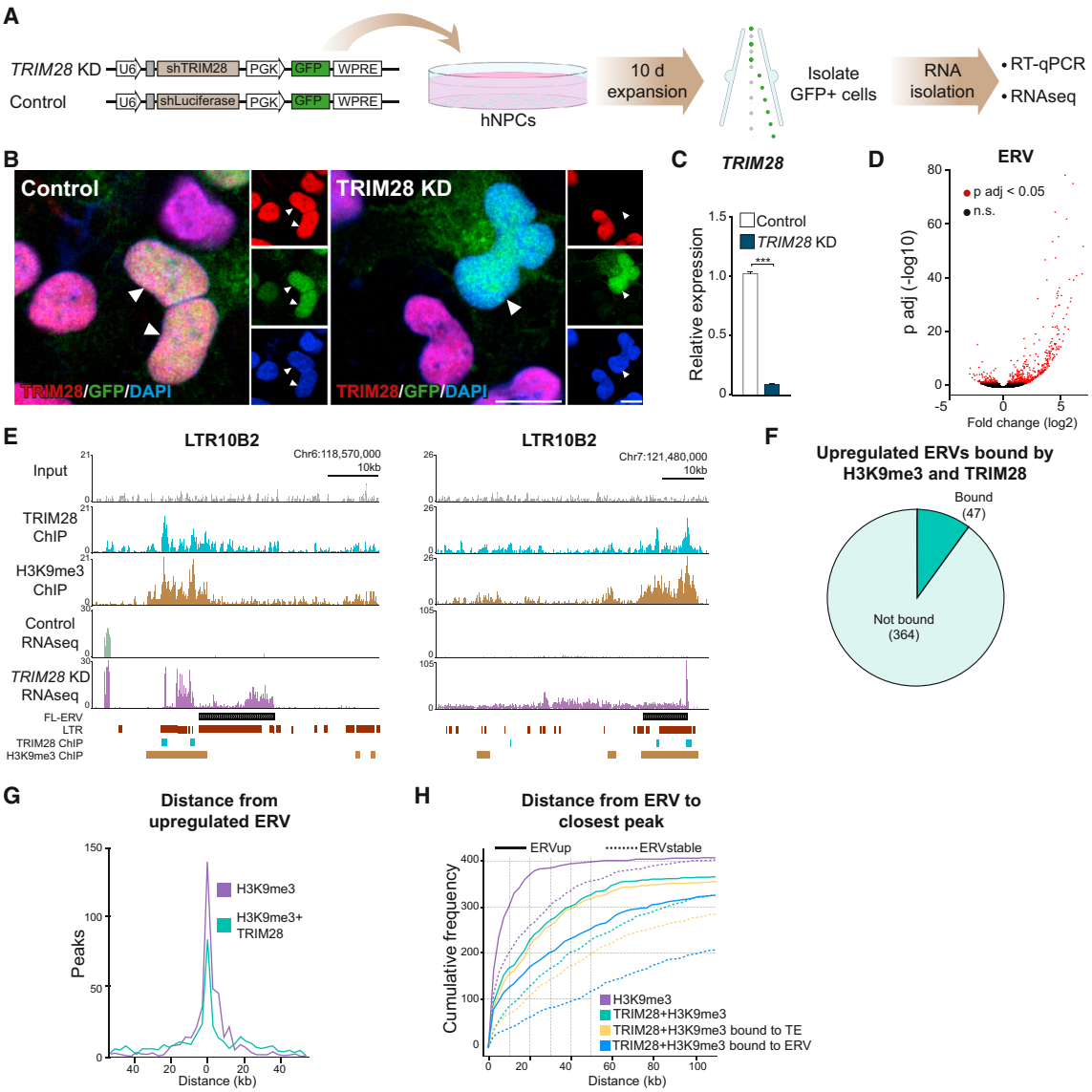


Figure 3. Knockdown of TRIM28 in hNPCs Results in Upregulation of ERV Expression

(A) Schematics of *TRIM28*-KD strategy.

(B) Immunocytochemistry demonstrates the loss of TRIM28 upon *TRIM28*-KD.

(C) qRT-PCR analysis demonstrates the loss of *TRIM28* upon *TRIM28*-KD, displayed as mean with error bars representing SEM; unpaired Student's t test, $p < 0.0001$.

(D) A total of 411 ERV elements was significantly upregulated upon *TRIM28*-KD. Red points represent ERVs with BH-corrected p value < 0.05. Fold change: log2 of KD/CTR.

(E) Read coverage of upregulated ERVs co-bound by TRIM28-H3K9me3, as well as nearby ERVs. FL-ERV track, RetroTector; LTR, RepeatMasker; ChIP, intervals of TRIM28 and H3K9me3 peaks.

(F) Eleven percent of all upregulated ERVs were co-bound by TRIM28-H3K9me3.

(G) Upregulated ERVs were close to a H3K9me3-peak (green) or a TRIM28-H3K9me3 (purple) co-peak. A 3-kb window was used to calculate the frequencies.

(H) Cumulative frequencies of distance from upregulated ERVs to the closest peaks in the different categories. Dotted lines show distance from a random selection of stable ERVs (p value > 0.9, n = 411). Scale bars represent 10 μm in (A), (B), and insets. See also Figures S3 and S4.

the vicinity of the surrounding genome. De-repression of this network results in transcriptional activation of nearby ERVs and other TEs as well as protein-coding genes and lncRNAs. This ERV-based gene regulatory network provides an additional layer of transcriptional control in hNPCs.

TRIM28 and KRAB-ZFPs have previously been described as a defense system against the invasion of mobile genetic elements, suggesting that TRIM28/KRAB-ZFPs establish silencing of ERVs during early development, which then is replaced by stable silencing of ERVs in somatic tissues characterized by DNA

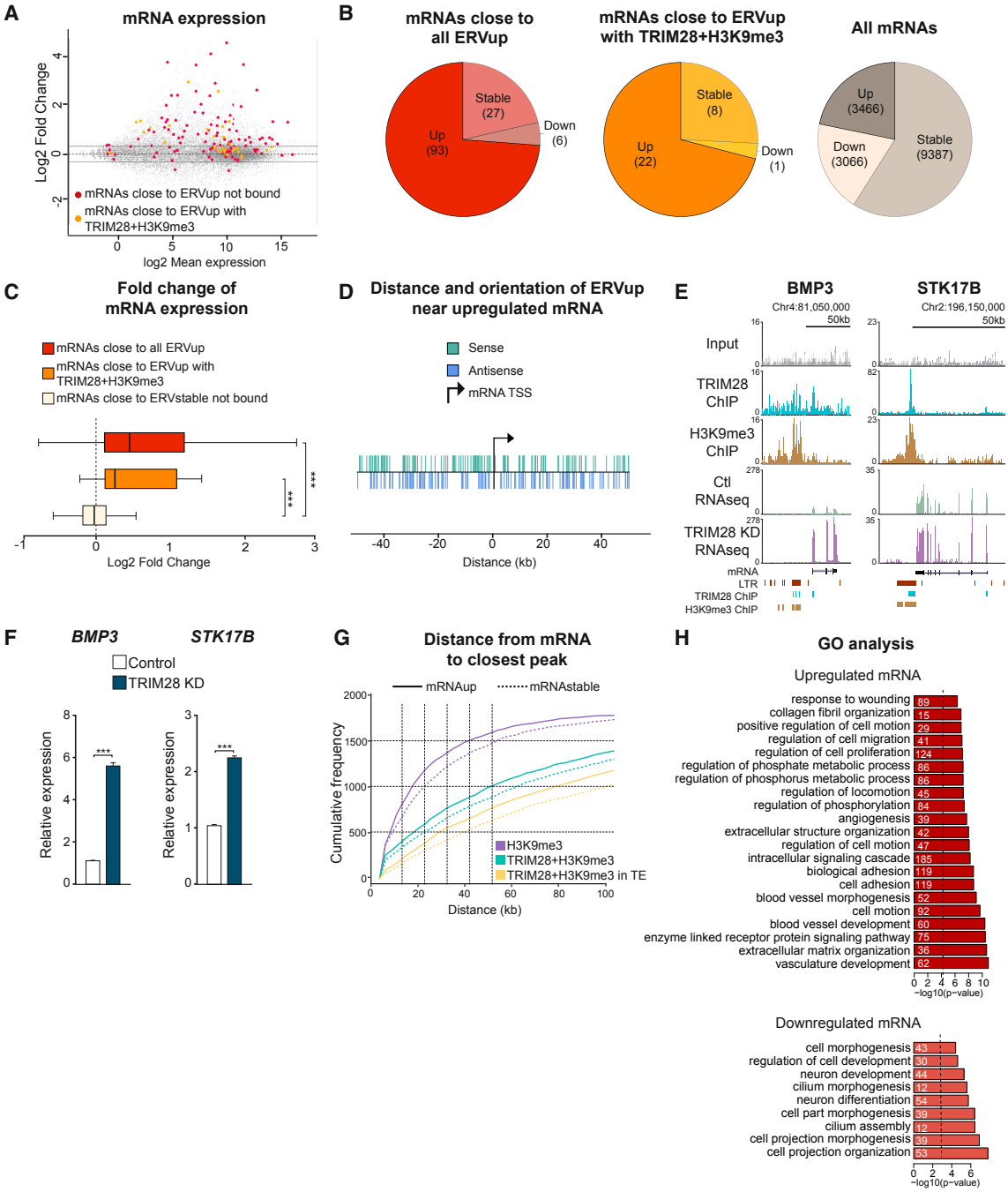


Figure 4. TRIM28-Bound ERVs Control the Expression Level of Nearby Protein-Coding Genes

(A) The majority of protein-coding genes with transcription start site (TSS) within 50 kb from an upregulated ERV (red) and an upregulated TRIM28-H3K9me3-bound ERV (orange) was significantly upregulated (BH-corrected p value < 0.05). The grey dots represent all mRNA. The x axis shows mean expression in control and *TRIM28*-KD samples combined.

(B) The pie charts visualize how many protein-coding genes with TSS within 50 kb from an upregulated ERV (red) and an upregulated TRIM28-H3K9me3-bound ERV (orange), that are stable, up- or downregulated upon *TRIM28*-KD. The grey pie chart to the right represents expression of all mRNAs upon *TRIM28*-KD.

(C) The fold changes of protein-coding genes close (<50 kb) to upregulated ERVs were significantly higher than fold changes of genes close to stable ERVs ($p < 0.0001$, Wilcoxon rank sum test, using $\log_2$-transformed fold changes). Boxes represent first to third quartiles. Whiskers represent 1.5*interquartile range.

(D) Plotting upstream/downstream or sense/antisense orientation of upregulated ERVs relative to TSS of the nearby upregulated genes reveals no preferences.

(E) Read coverage plots of *BMP3* and *STK17B* and associated ERVs; mRNA track, RefSeq; LTR, RepeatMasker; ChIP, intervals of TRIM28 and H3K9me3 peaks.

(legend continued on next page)

methylation (Rowe et al., 2010). As such, KRAB-ZFPs would have evolved to prevent aberrant expression and perhaps transposition of mobile genetic elements, thereby serving as guardians of the genome (Jacobs et al., 2014). Our results indicate that this model may be too simple and suggest that KRAB-ZFPs and ERVs have also co-evolved to form gene regulatory networks. Although TRIM28 represses the expression of ERVs and other TEs in hNPCs, it is not clear whether these elements provide functionality. In fact, ERVs that are expressed during human brain development are primarily small LTR fragments, which could represent transcriptional noise. The majority of FL-ERVs with established regulatory roles and protein-coding potential appear to be stably silenced in hNPCs by TRIM28-independent mechanisms. Our data are accordingly more in line with an important role for TRIM28 in meditating an indirect transcriptional repression of protein-coding genes located in the vicinity of ERVs.

Interestingly, ERVs bound by TRIM28 are mostly enriched for elements that integrated around 35–55 million years ago, thus not binding to the most recent ERVs as well as ancient elements. It has been speculated that a protein-based recognition system such as TRIM28/KRAB-ZFPs would co-evolve too slowly in order to silence novel ERVs entering a host. Rather, a small RNA-based system such as Piwi-interacting RNAs (piRNAs), which are adaptive in nature, could provide a more rapid response over evolutionary timescales, thereby silencing the most recently integrated ERV (Castro-Diaz et al., 2014; Jacobs et al., 2014). On a somewhat slower evolutionary timescale, the TRIM28/KRAB-ZFP network would then co-evolve with ERVs and subsequently take over transcriptional silencing, thereby explaining the bias for older elements. With time, ERVs will be domesticated, lose their mobile activity, and eventually become fixed in the population. Interestingly, TRIM28 binding to many of these elements is still kept in hNPCs, suggesting that ERVs and humans co-evolve to establish gene regulatory networks, similar to what has previously been demonstrated in *Drosophila* (Batut et al., 2013).

Because ERVs integrate into random sites of the genome, there is a considerable polymorphism across different hosts, which provides a powerful evolutionary mechanism to establish species-specific gene regulatory networks. In line with this, many of the ERVs that are bound by TRIM28 in hNPCs are primate-specific elements. In addition to this, KRAB-ZFPs have also undergone a rapid expansion in primate genomes, causing a species-specific KRAB-ZFP battery (Thomas and Schneider, 2011). Interestingly, many KRAB-ZFPs are highly expressed in the brain, and the transcriptional activity of several KRAB-ZFP orthologous differs greatly in the human and chimpanzee brain (Nowick et al., 2009). Together, these data points to the existence of a human-specific KRAB-ZFP/ERV network and raise the exciting hypothesis that ERVs have played a key role in the evolution of the primate brain.

Aberrant ERV expression has been implicated in several neurological disorders, including ALS, schizophrenia, and bipolar disorders (Christensen, 2016). It has been speculated that RNA or peptides derived from such ERV expression could have detrimental effects. However, our data suggest that the expression of ERVs detected in these syndromes could be linked to aberrancies in the TRIM28/KRAB network. In line with this, genetic variants in KRAB-ZFPs have recently been linked to bipolar disorders and depression (Subaran et al., 2016). Considering the influence of TRIM28-bound ERVs on nearby gene expression in hNPCs, it will be valuable to further investigate KRAB-ZFPs, the variation in their binding to ERVs in the human population, and their role in cognitive disorders.

A fascinating aspect of the TRIM28/KRAB system is that the general mechanism of TE binding and repression is conserved between human and mouse despite major differences in the setup of ERVs and KRAB-ZFPs. We have previously found that TRIM28 controls ERVs in mouse NPCs, and there are also emerging data documenting a role for KRAB-ZFPs and TRIM28 in other somatic tissues, including the regulation of nearby genes (Ecco et al., 2016; Fasching et al., 2015). It is worth noting that we previously found that deletion of TRIM28 in the mouse hippocampus results in behavioral impairments similar to what is found in individuals with psychiatric disorders (Jakobsson et al., 2008). Parallel studies to investigate how the TRIM28 network has evolved differently to control brain-specific networks in closely related species such as human and other primates as well as in more distant related species such as the mouse will be very interesting.

In summary, we report the existence of a TRIM28-controlled gene regulatory network based on ERVs that is active in hNPCs. Our findings open up for several exciting future studies on the role of ERVs as potential drivers of human brain evolution, their contribution to individual variation, and the implication in human brain disorders.

EXPERIMENTAL PROCEDURES

Detailed experimental procedures can be found in Supplemental Experimental Procedures.

Human Tissue and Cell Culture

Human tissue was obtained from legally aborted embryos from post-conception (P.C.) week 4.5–10.5, with approval of the Swedish National Board of Health and Welfare.

Two human neural epithelial-like stem cell lines, Sai2 (embryo-derived) (Tailor et al., 2013) and AF22 (iPS-derived) (Falk et al., 2012), were used.

ChIP-Seq Analysis

ChIP-seq was performed through Active Motif Epigenetic Services according to their procedures. The 50-bp single-end reads were mapped to the human genome assembly (GRCh38) using the Bowtie2 short sequence aligner (Langmead et al., 2009).

(F) Upregulation of *BMP3* and *STK17B* was confirmed by qRT-PCR, displayed as mean with SEM, $p < 0.0001$.

(G) Cumulative frequency of distances from upregulated (solid line) or stable (dotted line) protein-coding genes to the different peak categories. Stable genes were a random sample (n = 3,466) of mRNAs with p value > 0.9 upon *TRIM28*-KD.

(H) GO analysis of upregulated or downregulated genes upon *TRIM28*-KD (BH-corrected p value < 0.05). See also Figure S3 and Table S3.

RNA Analysis

Total RNA from both tissue and cells was isolated using the miRNeasy Micro Kit (QIAGEN) and used either for RNaseq or qPCR (for primers, see Table S4). The 50-bp single-reads were mapped to the human genome assembly (GRCh38) using STAR (2.4.0j) (Dobin et al., 2013). RNA-seq data from hESCs were downloaded from the NCBI GEO (GEO: GSE57866).

FL-ERV Identification and Analysis

We used the RetroTector software (Sperber et al., 2007) to identify FL-ERVs in the human genome (GRCh38).

Statistical Methods

For the ChIP-seq data, a one-sided binomial test was performed to test for enrichment in observed data relative to shuffled coordinates. For the RNA-seq data, clustering and differential expression analysis was done with DESeq2. Differences between groups in the qRT-PCR data were analyzed with unpaired t test (two groups) or one-way ANOVA with Bonferroni post hoc test (more than two groups).

ACCESSION NUMBERS

The accession number for the RNA-seq and ChIP-seq data reported in this paper is NCBI GEO: GSE84259.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, four figures, and four tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2016.12.010.

AUTHOR CONTRIBUTIONS

P.L.B, M.E.J, L.F., J.N.W, M.S, R.F., A.F., P.J., M.P., and J.J. designed and performed research and analyzed data. P.L.B and P.J. performed bioinformatics analysis. P.L.B, M.E.J., and J.J. wrote the paper. All authors reviewed the manuscript.

ACKNOWLEDGMENTS

We are grateful to all members of the J.J. lab. We also thank M. Persson Vejgården, U. Jarl, J. Johansson, A. Hammarberg, E. Ling, Christina Isaksson, B. Mattson, S. da Rocha Baez, M. Sparrenius, and I. Nilsson for technical assistance. The work was supported by grants from the Swedish Research Council (J.J.: K2014-62X-22527-01-3; P.J.: 2015-02429), the Swedish Foundation for Strategic Research (FFL12-0074), the Swedish Brain Foundation (FO2015-0040), the Swedish Cancer Foundation (15 0279), the Swedish Excellence Project Basal Ganglia Disorders Linnaeus Consortium (Bagadilico; 70862601), and the Swedish Government Initiative for Strategic Research Areas (MultiPark and StemTherapy).

Received: August 18, 2016
Revised: October 26, 2016
Accepted: December 2, 2016
Published: January 3, 2017

REFERENCES

Bannert, N., and Kurth, R. (2006). The evolutionary dynamics of human endogenous retroviral families. Annu. Rev. Genomics Hum. Genet. *7*, 149–173.

Batut, P., Dobin, A., Plessy, C., Carninci, P., and Gingeras, T.R. (2013). High-fidelity promoter profiling reveals widespread alternative promoter usage and transposon-driven developmental gene expression. Genome Res. *23*, 169–180.

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

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

Christensen, T. (2016). Human endogenous retroviruses in neurologic disease. APMIS *124*, 116–126.

Cowley, M., and Oakey, R.J. (2013). Transposable elements re-wire and fine-tune the transcriptome. PLoS Genet. *9*, e1003234.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Ecco, G., Cassano, M., Kauzlaric, A., Duc, J., Coluccio, A., Offner, S., Imbeault, M., Rowe, H.M., Turelli, P., and Trono, D. (2016). Transposable elements and their KRAB-ZFP controllers regulate gene expression in adult tissues. Dev. Cell *36*, 611–623.

Falk, A., Koch, P., Kesavan, J., Takashima, Y., Ladewig, J., Alexander, M., Wiskow, O., Tailor, J., Trotter, M., Pollard, S., et al. (2012). Capture of neuroepithelial-like stem cells from pluripotent stem cells provides a versatile system for in vitro production of human neurons. PLoS One *7*, e29597.

Fasching, L., Kapopoulou, A., Sachdeva, R., Petri, R., Jönsson, M.E., Männe, C., Turelli, P., Jern, P., Cammas, F., Trono, D., and Jakobsson, J. (2015). TRIM28 represses transcription of endogenous retroviruses in neural progenitor cells. Cell Rep. *10*, 20–28.

Göke, J., Lu, X., Chan, Y.S., Ng, H.H., Ly, L.H., Sachs, F., and Szczerbinska, I. (2015). Dynamic transcription of distinct classes of endogenous retroviral elements marks specific populations of early human embryonic cells. Cell Stem Cell *16*, 135–141.

Groner, A.C., Meylan, S., Ciuffi, A., Zangger, N., Ambrosini, G., Dénervaud, N., Bucher, P., and Trono, D. (2010). KRAB-zinc finger proteins and KAP1 can mediate long-range transcriptional repression through heterochromatin spreading. PLoS Genet. *6*, e1000869.

Hardison, R.C., Roskin, K.M., Yang, S., Diekhans, M., Kent, W.J., Weber, R., Elnitski, L., Li, J., O'Connor, M., Kolbe, D., et al. (2003). Covariation in frequencies of substitution, deletion, transposition, and recombination during eutherian evolution. Genome Res. *13*, 13–26.

Hayward, A., Grabherr, M., and Jern, P. (2013). Broad-scale phylogenomics provides insights into retrovirus-host evolution. Proc. Natl. Acad. Sci. USA *110*, 20146–20151.

Hayward, A., Cornwallis, C.K., and Jern, P. (2015). Pan-vertebrate comparative genomics unmasks retrovirus macroevolution. Proc. Natl. Acad. Sci. USA *112*, 464–469.

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

Jakobsson, J., Cordero, M.I., Bisaz, R., Groner, A.C., Busskamp, V., Bensadoun, J.C., Cammas, F., Losson, R., Mansuy, I.M., Sandi, C., and Trono, D. (2008). KAP1-mediated epigenetic repression in the forebrain modulates behavioral vulnerability to stress. Neuron *60*, 818–831.

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

Lu, X., Sachs, F., Ramsay, L., Jacques, P.E., Göke, J., Bourque, G., and Ng, H.H. (2014). The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. Nat. Struct. Mol. Biol. *21*, 423–425.

Nowick, K., Gernat, T., Almaas, E., and Stubbs, L. (2009). Differences in human and chimpanzee gene expression patterns define an evolving network of transcription factors in brain. Proc. Natl. Acad. Sci. USA *106*, 22358–22363.

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

Rowe, H.M., Jakobsson, J., Mesnard, D., Rougemont, J., Reynard, S., Aktas, T., Maillard, P.V., Layard-Liesching, H., Verp, S., Marquis, J., et al. (2010).

KAP1 controls endogenous retroviruses in embryonic stem cells. Nature *463*, 237–240.

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

Sperber, G.O., Airola, T., Jern, P., and Blomberg, J. (2007). Automated recognition of retroviral sequences in genomic data—RetroTector. Nucleic Acids Res. *35*, 4964–4976.

Stoye, J.P. (2012). Studies of endogenous retroviruses reveal a continuing evolutionary saga. Nat. Rev. Microbiol. *10*, 395–406.

Subaran, R.L., Odgerel, Z., Swaminathan, R., Glatt, C.E., and Weissman, M.M. (2016). Novel variants in ZNF34 and other brain-expressed transcription factors are shared among early-onset MDD relatives. Am. J. Med. Genet. B Neuropsychiatr. Genet. *171*, 333–341.

Tailor, J., Kittappa, R., Leto, K., Gates, M., Borel, M., Paulsen, O., Spitzer, S., Karadottir, R.T., Rossi, F., Falk, A., and Smith, A. (2013). Stem cells expanded from the human embryonic hindbrain stably retain regional specification and high neurogenic potency. J. Neurosci. *33*, 12407–12422.

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

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

Vallender, E.J., Mekel-Bobrov, N., and Lahn, B.T. (2008). Genetic basis of human brain evolution. Trends Neurosci. *31*, 637–644.

Wang, J., Xie, G., Singh, M., Ghanbarian, A.T., Raskó, T., Szvetnik, A., Cai, H., Besser, D., Prigione, A., Fuchs, N.V., et al. (2014). Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. Nature *516*, 405–409.

Wang, J., Singh, M., Sun, C., Besser, D., Prigione, A., Ivics, Z., Hurst, L.D., and Izsvák, Z. (2016). Isolation and cultivation of naive-like human pluripotent stem cells based on HERVH expression. Nat. Protoc. *11*, 327–346.

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

Yamashita, K., Mikawa, S., and Sato, K. (2016). BMP3 expression in the adult rat CNS. Brain Res. *1643*, 35–50.

Cell Reports, Volume *18*

Supplemental Information

TRIM28 Controls a Gene Regulatory Network Based on Endogenous Retroviruses in Human Neural Progenitor Cells

Per Ludvik Brattås, Marie E. Jönsson, Liana Fasching, Jenny Nelander Wahlestedt, Mansoureh Shahsavani, Ronny Falk, Anna Falk, Patric Jern, Malin Parmar, and Johan Jakobsson

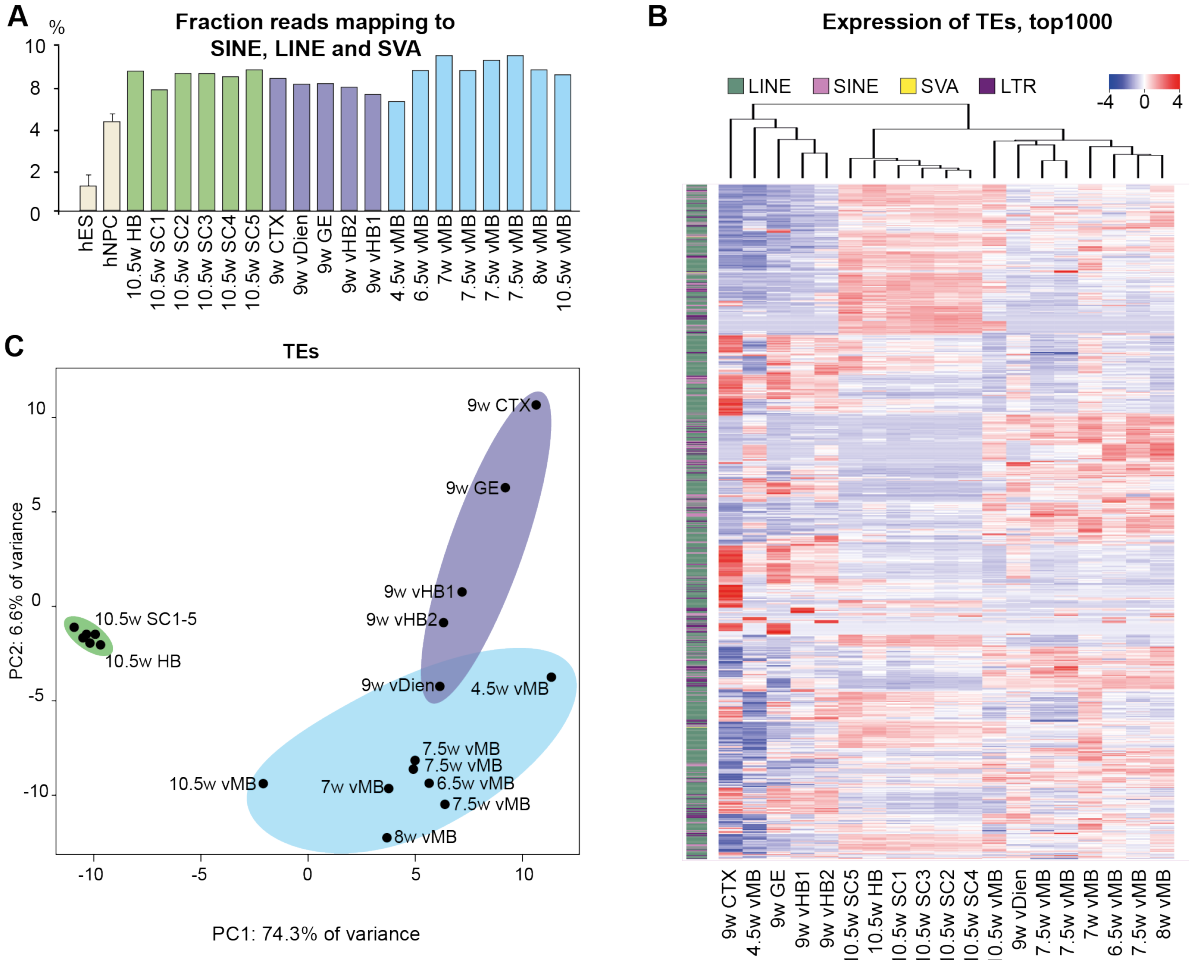


Figure S1. Stage- and region-specific expression of different TE classes in the developing human brain. Related to Figure 1.

A) Fraction of RNA-seq reads mapping to SINE, LINE and SVA elements out of total number of reads mapping to the genome. B) Dynamic expression of TEs in the fetal brain. The heatmap shows the 2000 individual TEs with the largest variance over the samples. Color bar to the left indicates the class of the TE. C) PCA of the 500 TEs with the largest variance in expression (including SINE, LINE, SVA and LTR retrotransposons).

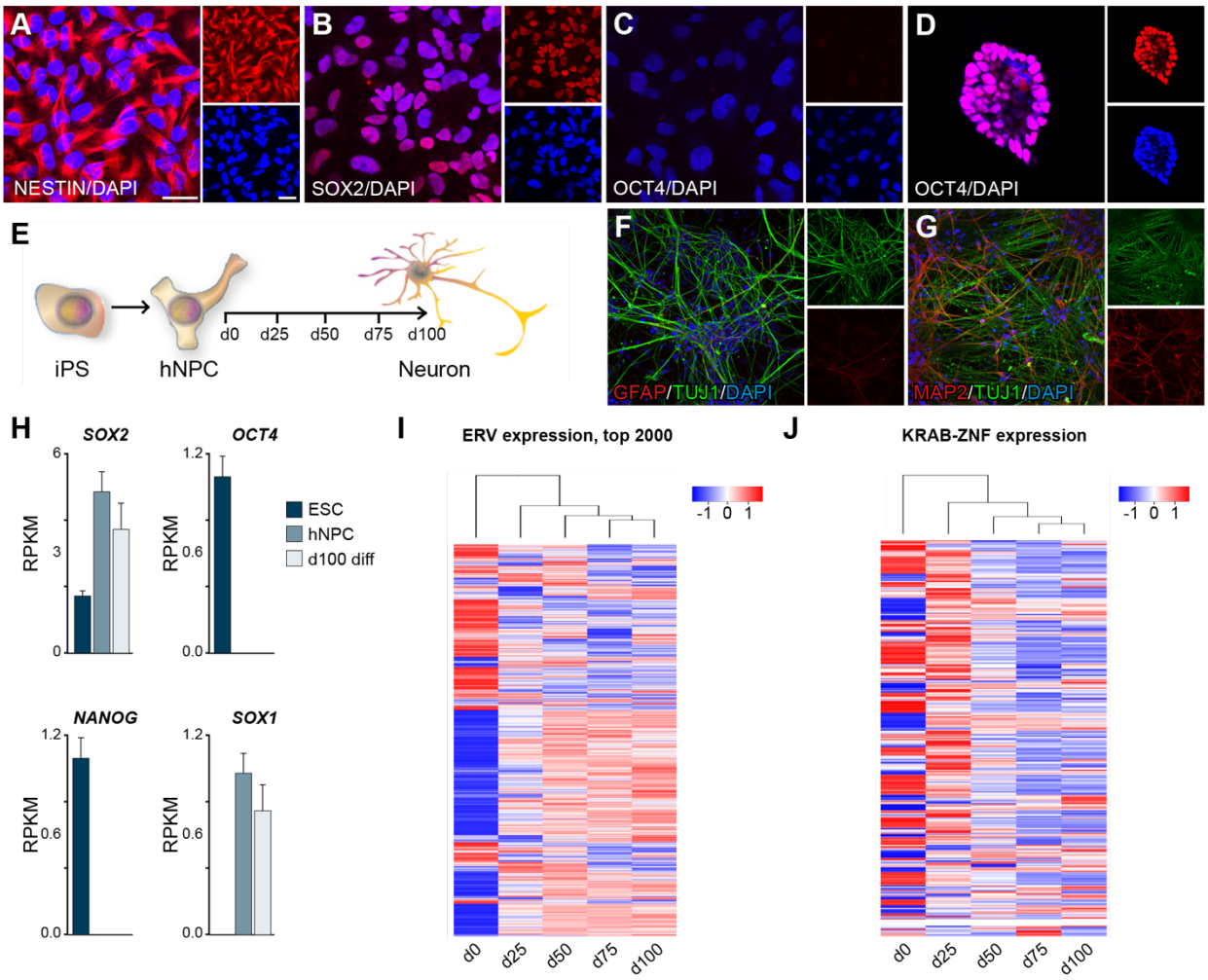


Figure S2. Characterization of the iPS-derived hNPC line, AF22. Related to Figure 2.
A-D) The hNPCs express the NPC markers NESTIN, SOX2, but not the pluripotent marker OCT4, which is strongly expressed in hESCs (H9). E) Schematic of differentiation assay. Differentiated cells are GFAP negative (F) and MAP2 positive (G). H) RNA-seq; hNPCs are positive for SOX1 and SOX2, while negative for NANOG and OCT4. Reads were normalized on reads mapping to the genome and length of transcript. I-J) Dynamic expression profiles of the 2000 ERV elements with the highest variance through the samples (I) and all genes containing KRAB-ZPF domains (J) upon differentiation from NPCs to neurons. Scalebar A-D 20µm.

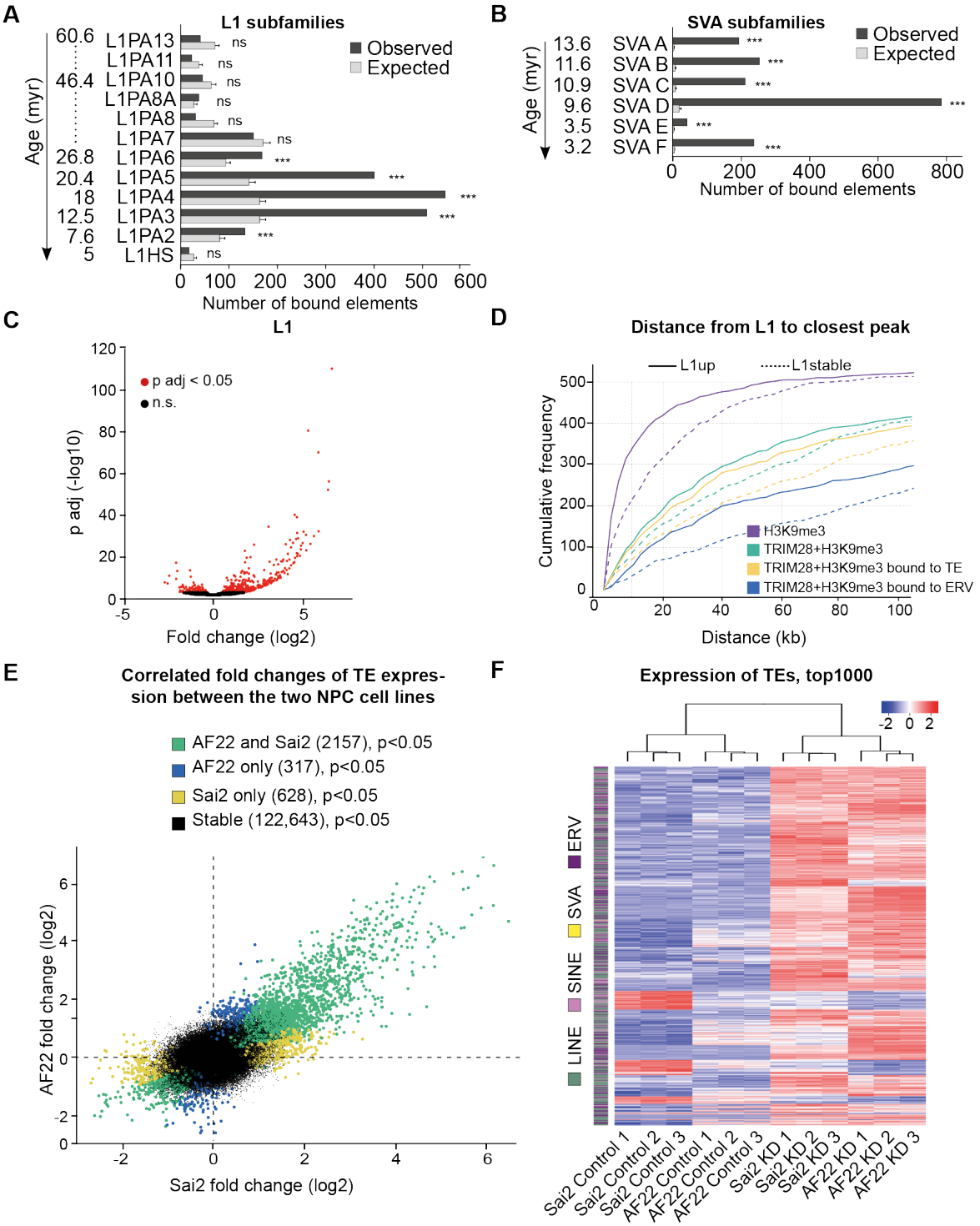


Figure S3. Related to Figure 2-4.

A-B) Number of TRIM28-H3K9me3 co-peaks overlapping L1 (A) or SVA (B) subfamilies. Shuffled values are the mean of 10,000 bootstraps of the shuffled coordinates. Error bars represent standard error of the mean. Estimated ages of L1 and SVA subfamilies were obtained from previously published analyses (Khan et al., 2006; Wang et al., 2005). C) 493 LINE-1 elements were upregulated upon TRIM28 KD. FC: KD/CTR. P-adj: BH-corrected p-value. D) Cumulative frequency of distance from each upregulated LINE-1 element to closest ChIP-seq peak of the different categories. L1stable: Random selection of LINE-1 elements (n=493) that are not significantly affected by TRIM28-KD (p-value > 0.9). E) Correlation of fold change of all TEs upon TRIM28-KD between iPS-derived (AF22) and fetal-derived (Sai2) NPCs. Fold change: TRIM28-KD/CTR. Green points: significant in both cell lines. Blue: Only significant in AF22. Yellow: Only significant in Sai2. Black: Stable in both cell lines. F) Correlated expression of TEs between the two cell lines. Heatmap displays the 1000 TEs with the lowest p-values upon TRIM28-KD in Sai2. RNA-seq reads were scaled to total number of reads mapping to the genome, log-transformed and row-normalized for visualization.

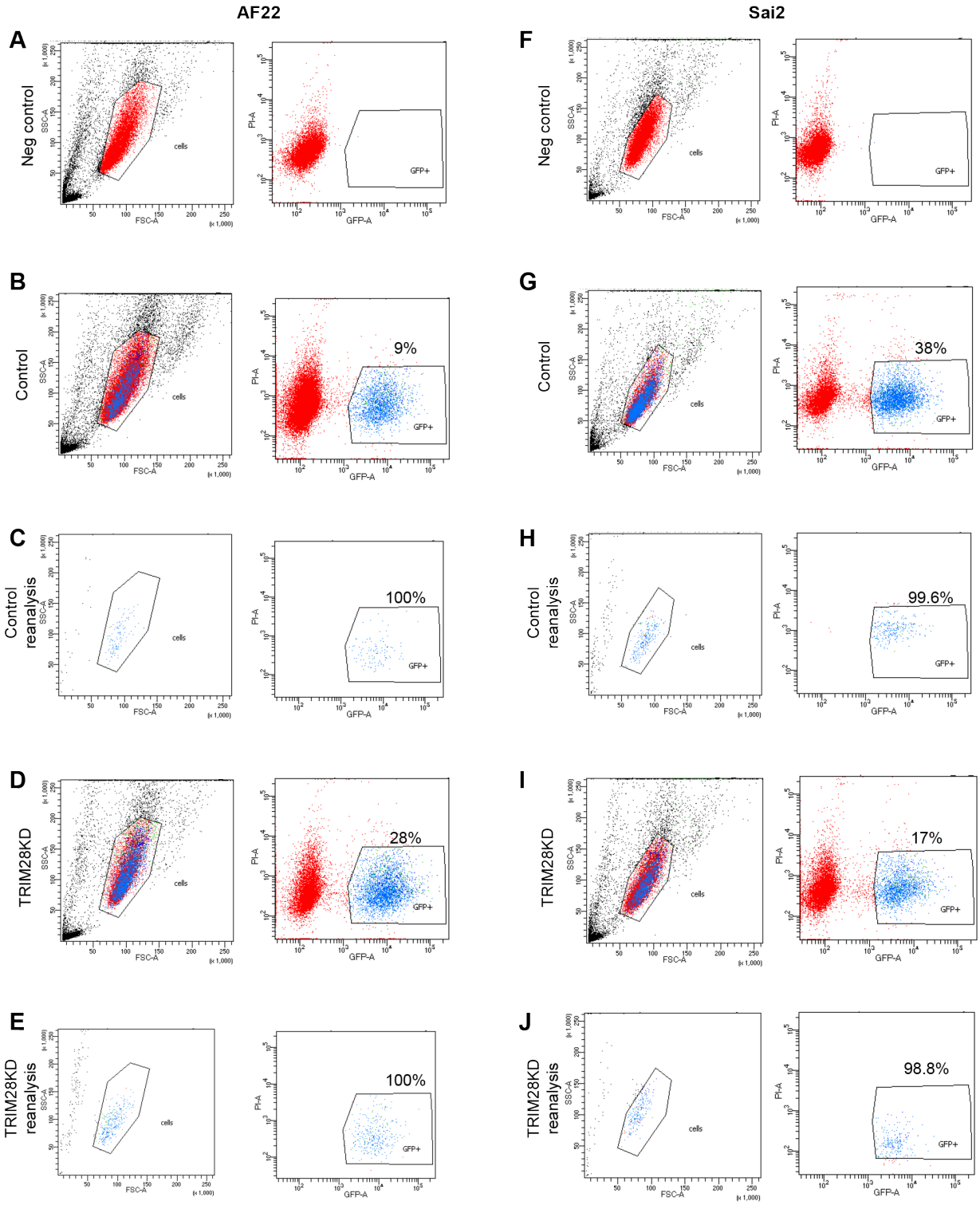


Figure S4. Cell sorting of hNPCs by FACS 10 days of expansion post lentiviral transduction. Related to figure 3. Control and KD groups were performed in triplicates for both cell lines. Non-transduced cells were used as a negative control for the GFP expression. Control cells were transduced with a vector expressing shLuciferase.GFP, while the KD cells were transduced with a vector expressing shTRIM28.GFP. Doublets were excluded (not shown). The re-analysis showed very high purity of sorted cells.
A-E) The FACS plots from the KD experiment in the AF22 cell line (iPS-derived). F-J) The FACS plots KD experiment in the Sai2 cell line (embryo-derived).

Table S1. Human embryonic tissue. Related to Figure 1.

The RNA from the tissue samples were isolated using the miRNeasy micro kit (Qiagen) and the libraries were made using the KAPA Stranded mRNA-Seq kit (KAPA Biosciences). Samples were sequenced over two rounds of sequencing and at both timepoints the samples were pooled as part of 18 samples and run over 3 lanes in the HiSeq3000.

	Weeks post conception	Region disected	Date of dissection (initials)	Date of RNA isolation (initials)
Seq Round1	4.5	vVM	During 2013 (JN)	During 2013 (JN)
	6.5	vVM	130323 (JN)	151215 (MJ)
	7.5	vVM	150429 (MJ)	151215 (MJ)
	8	vVM	During 2013 (JN)	During 2013 (JN)
	10.5	vVM	During 2013 (JN)	During 2013 (JN)
	9	Cortex	141203 (MJ)	150217 (MJ)
		Ventral diencephalon	141203 (MJ)	150217 (MJ)
		Ganglionic eminences	141203 (MJ)	150217 (MJ)
		Hindbrain 1-2 (anterior -> posterior)	141203 (MJ)	150217 (MJ)
	10.5	Hindbrain	141203 (MJ)	150217 (MJ)
		Spinal cord 1-5 (anterior -> posterior)	141203 (MJ)	150217 (MJ)
Seq Round2	7	vVM	141105 (EM)	151215 (MJ)
	7.5	vVM	150225 (EM)	160307 (MJ)
	7.5	vVM	141105 (EM)	160307 (MJ)

Table S2. Genomic coordinates for all FL-ERVs with score > 300 detected by the RetroTector software, used in this study. Related to Figure 1-2.

See separate Excel file

Table S3. Significantly upregulated mRNAs. Related to Figure 4
List of significantly upregulated mRNAs, with additional information whether they are close to (<50kb) from an upregulated ERV, and if that nearby upregulated ERV is also TRIM28+H3K9me3 bound.

See separate Excel file.

Table S4. qRT-PRC primers. Related to Figure 3-4.

Gene	Forward (5'-3')	Reverse (5'-3')
B-ACTIN	CCTTGCACATGCCGGAG	GCACAGAGCCTCGCCTT
HPRT1	ACCCTTTCCAAATCCTCAGC	GTTATGGCGACCCGCAG
TRIM28	GTAATGATGCCCAGAAGGT	GTCACTCTCCAGAGCCCAAG
BMP3	AGCAGTGGATTGAACCTCGG	GCATCAAAGGACTTGGGGGA
STK17B	GGCCTACTAACTACAACTCCTCA	TGCAGCATATTCTTGGCCAGT

Supplemental Experimental Procedures

Cell culture

Two human neural epithelial-like stem cell lines, Sai2 (embryo-derived) (Tailor et al., 2013) and AF22 (iPS-derived) (Falk et al., 2012), were cultured in DMEM/F12 (Thermo Fisher Scientific) supplemented with Glutamine (2mM, Sigma), Penicillin/Streptomycin (1x, Gibco), N2 supplement (1x, Thermo Fisher Scientific), B27 (0.05x, Invitrogen) , EGF and FGF2 (both 10ng/ml, Thermo Fisher Scientific). Cells were grown on Nunc^{TM} T25 or T75 flasks pre-coated with Poly L-Ornithine (15μg/ml, Sigma) and Laminin (2μg/ml, Sigma). Cells were passaged every 2-3 days using TryplLE^{TM} Express enzyme (Life Technologies) and Defined Trypsin Inhibitor (Life Technologies) and plated at a density of $6x10^4$ per cm^2. Cells were differentiated by removal of EGF and FGF2 and changing media every second day.

Immunofluorescence

Prior to immunofluorescence the cells and tissue were fixed in 4% paraformaldehyde for 15min and 12 hours, respectively. The tissue was cryopreserved in 30% sucrose overnight and sectioned on a cryostat (12μm). The cells and tissue were and incubated overnight with the primary antibody (TRIM28 (MAB3662, Millipore), Nestin (611658, BD), SOX2 (MAB2018, R&D), GFP (6556, Abcam), OCT4 (5279, SantaCruz), GFAP (G3893, Sigma)) followed by a 2h incubation with a fluorophore-conjugated secondary antibody (Jackson Laboratories) and DAPI. Imaging of cells was performed with a fluorescence microscope (Leica), while tissue was imaged using a confocal microscope (Leica).

TRIM28 knockdown

pLKO.1.GFP shRNA vectors were used for *TRIM28* KD (Turelli et al., 2014), which are third-generation SIN vectors. Two different shRNA vectors targeting *TRIM28* were used in parallel and a shRNA vector targeting Luciferase was used as a control. Lentiviral vectors were produced as previously described (Zufferey et al., 1997) and were of titers 10^9 TU/ml which was determined using qRT-PCR (Georgievska et al., 2004).

hNPCs were transduced with an MOI 1 and allowed to expand 10days prior to FACS (FACSAria, BD sciences, Supplementary Figure 4). Cells were re-suspended in culture media, filtered (70μm, BD Bioscience) and propidium iodide was added allowing for dead cell exclusion. Isolated cells were spun down into pellets, the supernatant was removed before snap freezing the pellets on dry ice. Cells were kept at -80°C until RNA was isolated using the miRNeasy micro kit (Qiagen). All groups were performed in triplicates.

ChIP-seq

ChIP-seq was performed through Active Motif Epigenetic Services using antibodies for H3K9me3 (# 39161, Active Motif) and TRIM28 (ab10484, Abcam) according to their procedures. Sequencing was performed on Illumina HiSeq 2500. Quality control of the 50bp single-end reads was done using FastQC (Babraham Bioinformatics, http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/) and adapter sequences were removed. The trimmed reads were mapped to the human genome assembly (GRCh38) using the Bowtie2 short sequence aligner with *--sensitive-local* parameters (Langmead et al., 2009; Zhang et al., 2008). All alignments with MAPQ score lower than 10 were discarded, allowing for 3 mismatches. Peaks were called using the MACS2 software (Zhang et al., 2008),

normalizing to total input. Peaks with q<0.05 were kept for further analysis. H3K9me3 peaks were called with the -–*broad* parameter and TRIM28 peaks were called with default parameters. The samtools and bedtools suites (Li et al., 2009; Quinlan and Hall, 2010) were used to generate bigwig and bigbed files for visualization in the UCSC genome browser.

TRIM28 and H3K9me3 co-peaks were computed with the intersect module of the bedtools suite, defining a co-peak when at least 1bp of the peak regions overlapped. To test for enriched TRIM28-H3K9me3 binding at TE classes and ERV families, we used the TEanalysis software shuffle_bed.pl (https://github.com/4ureliek/TEanalysis, (Lynch et al., 2015)). GRCh38 assembly gap coordinates were downloaded from the UCSC genome table browser, and these gap coordinates were excluded from the peak shuffling.

For genomic coordinates and age estimation of ERVs annotated in RepeatMasker, we applied the output file from http://www.repeatmasker.org/species/hg.html, and the Kimura divergence from consensus with an estimated mutation rate of 0.2% per million years (Hardison et al., 2003). Estimated ages of SVA and LINE1 subfamilies (Supplemental Figure 3A-B) were obtained from previously published divergence studies (Khan et al., 2006; Wang et al., 2005).

Genomic coordinates for TEs were downloaded from the UCSC genome browser (RepeatMasker track), of the GRCh38 assembly. Downstream analyses were performed using in-house R and unix scripts.

RNA analysis

Total RNA from both tissue and cells was isolated using the miRNeasy Micro Kit (Qiagen) and used either for RNAseq or qPCR. cDNA was created by reverse transcription with random hexamer primers and Superscript III (Invitrogen) and analysed by quantitative PCR with SYBR Green I master (Roche) on a LightCycler 480 (Roche). Data are represented with the ΔΔCt method normalized to the housekeeping genes *GAPDH* and *HPRT1*. Error bars represent standard error of mean from three biological and three technical triplicates and differences between groups. Primers are listed in Supplemental Table 4.

Libraries from total RNA were made using the KAPA stranded mRNA-Seq kit. The 50-bp single-reads from the Illumina HiSeq 2500 (ChIPseq hNPCs and differentiating hNPCs) and 3000 (TRIM28-KD in hNPCs and human tissue) were mapped to the human genome assembly (GRCh38) using STAR (2.4.0j)(Dobin et al., 2013).
1,164,889,003 sequence reads were generated from the 19 human tissue samples, 469,778,298 from the 10 hNPC differentiation samples and 698,302,212 reads from the 12 samples from the TRIM28-KD experiments.
The RNA from the tissue samples were isolated using the miRNeasy micro kit (Qiagen) and the libraries were made using the KAPA Stranded mRNA-Seq kit (KAPA Biosciences). For details of tissue processing and sequencing see Supplementary Table 1.

Multiple mapping reads were discarded, and a maximum of 3 mismatches were allowed. mRNA and TE expression was quantified using the subread package FeatureCounts (Liao et al., 2014). Gene coordinates for mRNA and retroelements were obtained from NCBI and the UCSC genome browser RepeatMasker track (GRCh38). Read counts were normalized to the total number of reads mapping to the genome. Downstream analyses were performed using

DESeq2 (Love et al., 2014), in-house R and unix scripts. Gene ontology analysis was done with the Functional Annotation Tool of DAVID Bioinformatic Resources 6.7 (Huang da et al., 2009a, b).

RNA-seq fastq files from human embryonic stem cells were downloaded from the NCBI Gene expression omnibus (GEO accession number: GSE57866).

FL-ERV identification and analysis

We used the RetroTector software (Sperber et al., 2007) to identify 3841 ERVs (score>300) in the human genome (GRCh38). These are referred to as full-length ERVs in the text, however it should be noted that the algorithm allows smaller parts of the ERV to be missing. We then extracted 778 representative FL-ERVs with sufficient sequence information to construct a phylogenetic tree in FastTree2 (Price et al., 2010) within a framework of 85 reference sequences (Genbank and Repbase). To estimate FL-ERV integration ages, we compared sequence divergence between 5'- and 3-LTRs as these are identical at the time of integration into the host genome and assumed to accumulate neutral mutations at a rate of 0.2% per million year (Hardison et al., 2003).

Supplemental References

Georgievska, B., Jakobsson, J., Persson, E., Ericson, C., Kirik, D., and Lundberg, C. (2004). Regulated delivery of glial cell line-derived neurotrophic factor into rat striatum, using a tetracycline-dependent lentiviral vector. Human gene therapy *15*, 934-944.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res *37*, 1-13.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc *4*, 44-57.

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

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923-930.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550.

Lynch, V.J., Nnamani, M.C., Kapusta, A., Brayer, K., Plaza, S.L., Mazur, E.C., Emera, D., Sheikh, S.Z., Grutzner, F., Bauersachs, S.*, et al.* (2015). Ancient transposable elements

transformed the uterine regulatory landscape and transcriptome during the evolution of mammalian pregnancy. Cell Rep *10*, 551-561.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One *5*, e9490.

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

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