

GENETICS

Primate-restricted KRAB zinc finger proteins and target retrotransposons control gene expression in human neurons

Priscilla Turelli¹, Christopher Playfoot¹, Dephine Grun^{1*}, Charlène Raclot^{1*}, Julien Pontis¹, Alexandre Coudray¹, Christian Thorball¹, Julien Duc¹, Eugenia V. Pankevich^{1,2}, Bart Deplancke¹, Volker Buskamp^{3,4}, Didier Trono^{1†}

In the first days of embryogenesis, transposable element–embedded regulatory sequences (TEeRS) are silenced by Kruppel-associated box (KRAB) zinc finger proteins (KZFPs). Many TEeRS are subsequently co-opted in transcription networks, but how KZFPs influence this process is largely unknown. We identify ZNF417 and ZNF587 as primate-specific KZFPs repressing HERVK (human endogenous retrovirus K) and SVA (SINE-VNTR-Alu) integrants in human embryonic stem cells (ESCs). Expressed in specific regions of the human developing and adult brain, ZNF417/587 keep controlling TEeRS in ESC-derived neurons and brain organoids, secondarily influencing the differentiation and neurotransmission profile of neurons and preventing the induction of neurotoxic retroviral proteins and an interferon-like response. Thus, evolutionarily recent KZFPs and their TE targets partner up to influence human neuronal differentiation and physiology.

INTRODUCTION

Some 4.5 million transposable element (TE)–derived sequences are disseminated across the human genome, many of which integrated within the last few tens of million years (1). TEs are typically enriched in transcription factor (TF) binding sites and correspondingly influence gene expression in a broad range of biological events (2–5). However, TEeRS are silenced during the earliest phase of embryogenesis by KZFPs (Kruppel-associated box zinc finger proteins), which dock KAP1 (KRAB-associated protein 1, also known as TRIM28) and associated heterochromatin inducers at TE loci (6–8). The rapid evolutionary selection of KZFP genes was initially interpreted as solely reflecting the host component of an arms race, but recent data suggest that KZFPs team up with TEs to build species-restricted layers of epigenetic regulation (8, 9). The present work provides direct support for this model.

RESULTS

We previously determined that a 35–base pair (bp)–long TE sequence encompassing the HERVK14C (human endogenous retrovirus K14C) primer binding site (PBS)–encoding region (coined HERVK-R) confers KAP1-induced repression to a nearby PGK promoter in human embryonic stem cell (hESC) (10). As part of a large-scale screen, we identified ZNF417 and ZNF587 as selectively enriched at loci containing this HERVK sequence (9). Depleting these two KZFPs from hESC restored expression of an HERVK-R-containing PGK-GFP (green fluorescent protein) lentivector (LV) (Fig. 1A), while producing them in murine ESC silenced this vector, demonstrating the sequence-specific repressor potential of ZNF417

and ZNF587 and the likely absence of a mouse ortholog (fig. S1A). Phylogenetic analyses confirmed that *ZNF417* emerged in the human ancestral genome (11) ahead of the New World monkey split ~43 million years ago and that *ZNF587* arose by duplication some 24 million years later (fig. S1, B and C). *ZNF417* and *ZNF587* display 98% amino acid homology with some differences in their zinc fingerprints (ZFPs), the series of amino acid trios predicted to dictate the sequence specificity of their DNA binding (Fig. 1B and fig. S1C). Only rare individuals harbor homozygous loss-of-function (LoF) mutations in *ZNF417* or *ZNF587* in the Genome Aggregation Database (gnomAD) repertoire (<https://gnomad.broadinstitute.org/>) (Fig. 1B) and the two genes exhibit fairly comparable patterns of expression across tissues according to Genotype-Tissue Expression (GTEx) (<https://www.gtexportal.org/home/>) and the BrainSpan Atlas of the Developing Human Brain (human.brain-map.org), with higher levels of transcripts in adult pituitary gland, thyroid, ovary, uterus, and prenatal compared to postnatal brain structures (Fig. 1C and fig. S1D). Of note, the *ZNF814* gene is located between *ZNF417* and *ZNF587*, but its product shares only 46% amino acid homology with the two paralogs, including a markedly divergent ZFP, and displays a different pattern of expression (fig. S1, D and E).

Chromatin immunoprecipitation sequencing (ChIP-seq) of H1 hESC overexpressing hemagglutinin (HA)–tagged versions of ZNF417 and ZNF587 identified 321 and 451 peaks, respectively, including 171 in common. About 85% mapped to primate-restricted long terminal repeat (LTR)/ERVK, SINE-VNTR-Alu (SVA), and LTR/ERV1 (Fig. 1D and fig. S2, A and B), among which 12 human-specific LTR/ERVK and 4 of 8 HML-2 HERVK previously noted to be polymorphic in the population (table S1) (12). KAP1, which binds both KZFPs (fig. S2C), and H3K9me3, the repressive mark instated by the KAP1-associated histone methyltransferase SETDB1 (13), were enriched at the PBS-coding sequence of ZNF417/ZNF587-bound LTR/ERVKs in hESC (fig. S2D). Most bound ERV sequences correspond to the PBS_{Lys1.2}-coding region, and a highly homologous motif is found in SVA-D integrants (Fig. 1E). Furthermore, SMILE-seq (selective microfluidics-based ligand enrichment followed by

Copyright © 2020
The Authors, some
rights reserved;
exclusive licensee
American Association
for the Advancement
of Science. No claim to
original U.S. Government
Works. Distributed
under a Creative
Commons Attribution
NonCommercial
License 4.0 (CC BY-NC).

¹School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland. ²CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, Austria. ³Center for Regenerative Therapies, Technische Universität Dresden, Dresden, Germany. ⁴Faculty of Medicine, Department of Ophthalmology, University of Bonn, Bonn, Germany.

*These authors contributed equally to this work.

†Corresponding author. Email: didier.trono@epfl.ch

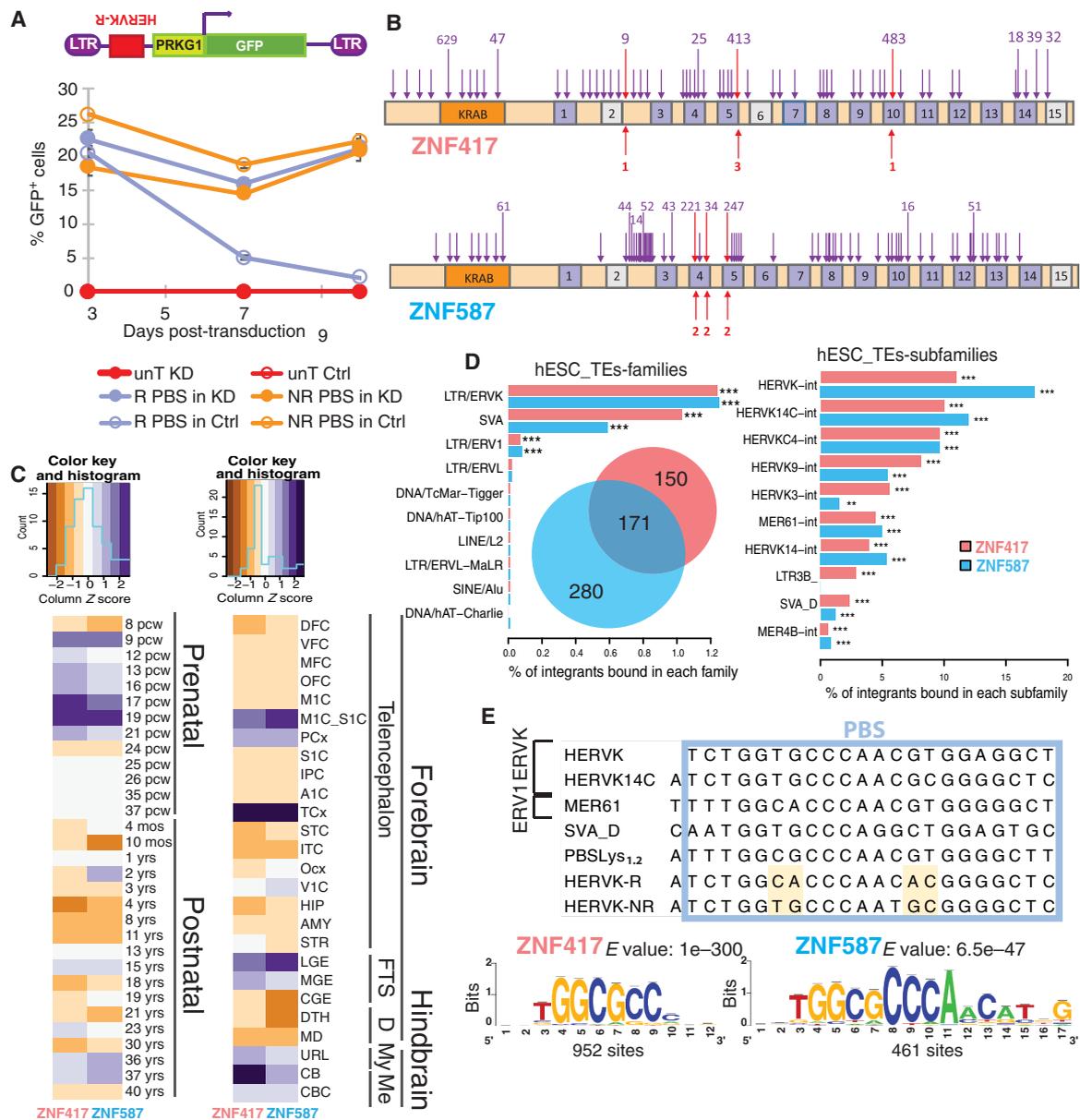


Fig. 1. Genomic characterization of ZNF417/587. (A) Expression from a PGK-GFP cassette cloned downstream of KAP1-restricted (R) or nonrestricted (NR) HERVK PBS sequences in control (Ctrl) or ZNF417/587 KD hESC (average and SD values of duplicates). unT, untransduced. (B) LoF mutations for ZNF417/587 with numbers of most frequent alleles among >140,000 individuals (in red homozygous LoF mutations). Dark and light purple boxes indicate intact and degenerated ZFs. (C) ZNF417/587 expression in brain development and substructures according to BrainSpan Atlas of the Developing Human Brain. pcw, post-conception weeks; FTS, forebrain fetal transient structures; D, diencephalon; My, myelencephalon; Me, metencephalon; M1C_S1C, primary motor sensory cortex; PCx, parietal cortex; TCx, temporal neocortex; LGE, lateral ganglionic eminence; MGE, medial ganglionic eminence; URL, upper rhombic lip; CB, cerebellum; CBC, cerebellar cortex. (D) Percentage of integrants from indicated TE families and subfamilies overlapping with ZNF417 and ZNF587 peaks as obtained by ChIP-seq triplicates in hESC. Only families with at least five bound integrants are shown. (***) $P < 0.001$ and ***) $P < 0.01$, hypergeometric test). (E) Top: PBS consensus sequences of bound LTR/ERVs and SVAs compared with PBS_{Lys1,2} and R- and NR-HERVK14C sequences. Bottom: Predicted binding motifs according to Rsat.

sequencing) (14) revealed that ZNF417 and ZNF587 had a higher affinity for methylated than unmethylated versions of this sequence (fig. S2E). Correspondingly, these KZFPs inefficiently repressed the HERVK-R-PGK-GFP LV in hESC depleted for the de novo DNA methyltransferases DNMT3A and DNMT3B, although this might also reflect indirect effects (fig. S2F).

The knockdown (KD) of ZNF417/ZNF587 in hESC resulted in up-regulating [fold change > 2, false discovery rate (FDR) < 0.05]

857 TEs, most notably members of the LTR/ERV1, LTR/ERVL-MaLR, SVA ($P < 0.001$, hypergeometric test), and LTR/ERVK ($P = 0.055$) subgroups, many of which harbored a ZNF417/ZNF587 binding motif (Fig. 2A, left) and were normally bound by KAP1 (Fig. 2A, middle) ($P < 0.001$, one-sided Fisher's exact test). Correspondingly, TEs normally bound by these KZFPs lost H3K9me3 and gained H3K4me1 and H3K27ac and were more expressed than their unbound counterparts in KD ESC (Fig. 2B, top). Expression of 854 genes

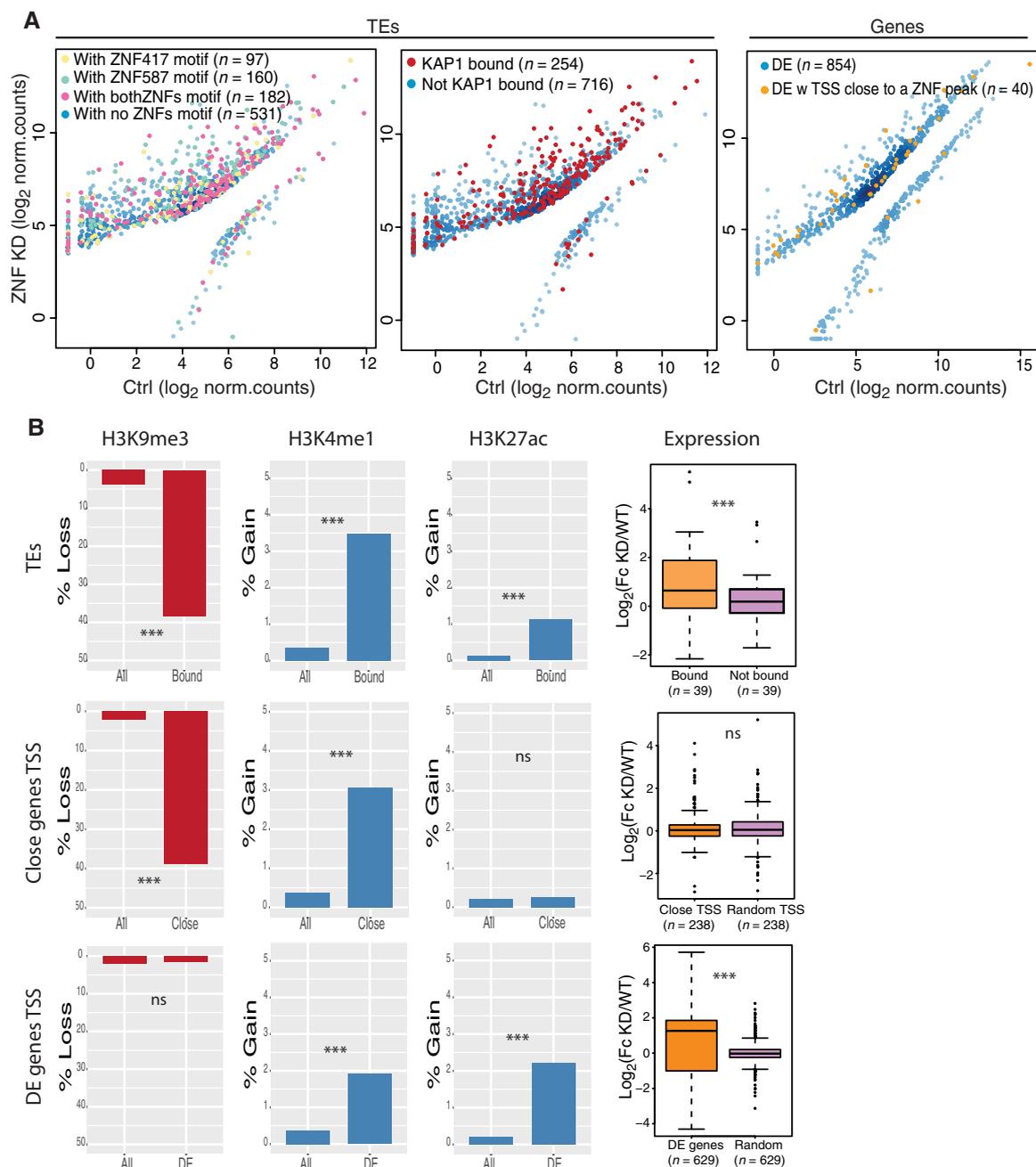


Fig. 2. Impact of ZNF417/587 depletion in pluripotent stem cells. (A) Dot plots of RNA-seq from ZNF417/587 KD quadruplicates versus control hESC triplicates, outlining DE TEs and genes (fold change > 2 , FDR < 0.05). TEs with predicted ZNF417 (yellow), ZNF587 (green), or both (pink) binding motifs (left) or bound by KAP1 in hESC (middle) or genes with a TSS closer than 100 kb from a ZNF binding site (right) are highlighted. (B) Bar plots depicting loss of H3K9me3 or gain in H3K4me1 or H3K27ac at indicated loci as obtained by independent ChIP-seq duplicates in hESC. Top: ZNF417/587-bound versus ZNF417/587-unbound TEs; middle: TSS of coding genes close to < 20 kb versus distant from a KZFP peak; bottom: TSS of coding DE genes versus all genes (P values, hypergeometric test). Right: Fold change (Fc) in expression in KD versus wild-type (WT) hESC of loci illustrated on the left (P values, Wilcoxon test). ns, not significant.

[differentially expressed (DE) genes] was also altered (Fig. 2A, right, and table S2), a majority up-regulated (fold change > 2 , FDR < 0.05), some ($P = 0.039$, one-sided Fisher's exact test) with a ZNF417/587 peak within 100 kb of the transcription start site (TSS). Those within 20 kb of a ZNF417/587-bound TE lost H3K9me3 and gained H4K4me1 at the TSS upon KZFP KD, but rarely displayed higher levels of H3K27Ac and transcription, consistent with a

poised promoter state (Fig. 2B, middle). In contrast, the TSS of genes induced in this setting displayed, on average, increased levels of H3K4me1 and H3K27ac but no change in H3K9me3 (Fig. 2B, bottom), probably because many of these genes, including 130 interferon (IFN)-sensitive genes (ISGs) and the putative targets of 31 up-regulated TFs, were indirectly rather than directly controlled by the KZFPs.

Using publicly available data, we found that TE integrants bound by ZNF417/ZNF587 in hESC were induced during embryonic genome activation (EGA), repressed upon naïve-to-primed hESC conversion (Fig. 3A), and that genes controlled by these two KZFPs were relatively more expressed during human than macaque EGA (fig. S3A), consistent with our recent proposal that KZFPs tame the transcriptional activity of EGA-promoting TE enhancers (8). ZNF417/587-targeted TEs were also more expressed in brain and testis than in other tissues (fig. S3B), and we found 40% of these loci to overlap with regions classified as brain and spinal cord enhancers in EnhancerAtlas 2.0 (15). Accordingly, several genes normally

expressed in the brain stood out among transcriptional units up-regulated in hESC depleted for ZNF417/587. For instance, *AADAT*, the product of which facilitates the synthesis of the neuroprotective kynurenic acid (16), and *PRODH*, a gene highly expressed in the brain where it influences GABAergic (γ -aminobutyric acid-mediated) neurotransmission and previously linked to schizophrenia (17, 18), both harbor ZNF417/587-recruiting HERVKs upstream of their promoters and were markedly induced by depletion of these KZFPs (Fig. 3B). Correspondingly, levels of *ZNF417/587* transcripts anti-correlate during development and in many regions of the adult brain with those of HERVKs and *PRODH* (Fig. 3C and fig. S3, C and D).

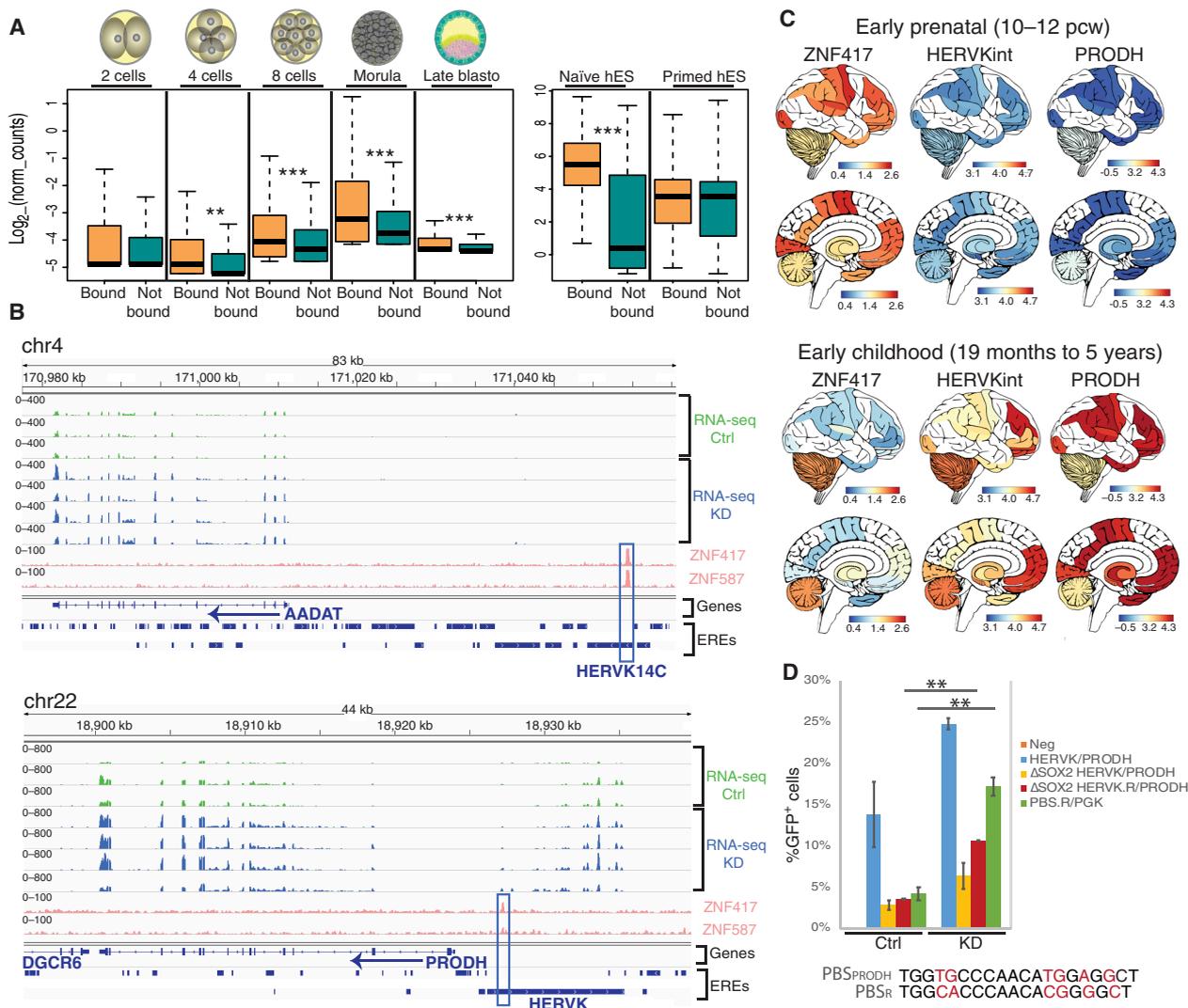


Fig. 3. ZNF417/587-mediated repression of TEeRs and neuron-specific genes. (A) Expression at indicated stages of human development using single-cell (left) or in naïve versus primed hESC (right) RNA-seq data (*P* values, Wilcoxon test) of TEs found or not found to be bound by ZNF417/587 in ESC. (B) IGV (Integrative Genomics Viewer) screenshots of independent RNA-seq replicates from control (Ctrl) and ZKFP KD hESC with boxed ZNF417/587 peaks at HERVK integrants upstream of *AADAT* (top) and *PRODH* (bottom). EREs, Endogenous Retroelement. (C) Spatial representation of *ZNF417*, *HERVK*, and *PRODH* expression in early prenatal and childhood brains, using RNA-seqs from the Brain Span Atlas of the Developing Human Brain. Prenatal brain is depicted as anatomically adult for consistency. (D) Repression assay in control or KD hESC using LVs carrying upstream of a GFP the genomic region encompassing the *PRODH* promoter and either the WT *HERVK* LTR5hs and PBS sequences (*HERVK/PRODH*), the sequence mutated in the two *SOX2* binding sites (Δ *SOX2 HERVK/PRODH*), or the PBS sequence mutated into an R PBS binding site (*HERVK.R/PRODH*). The previously described PGK-GFP LV with the strong KZFP recruiting 39-bp R.PBS sequence was used in parallel as a positive control of silencing (PBS.R/PGK). Data have been collected from independent duplicates of KD and control cells, at day 6 after transduction with equal amounts of LVs. Average and SD of the duplicates are shown. The WT PBS sequence as found on the WT *HERVK* upstream of the *PRODH* promoter (PBS_{PRODH}) and the mutated one (PBS_R) is shown.

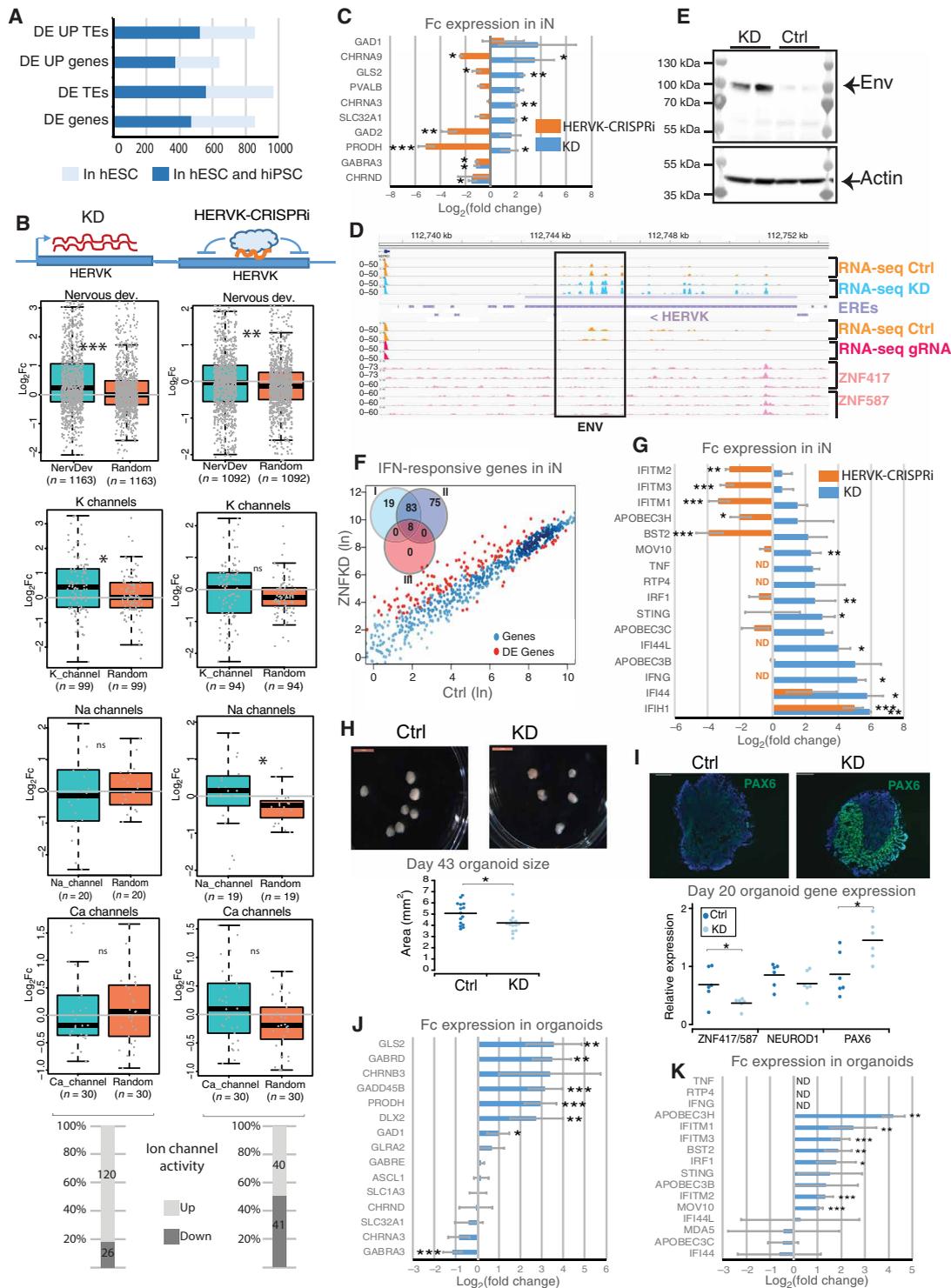


Fig. 4. Impact of ZNF417/587 on neuronal differentiation, function, and homeostasis. (A) Number of DE elements in indicated cells. (B) Expression in KD (left) or HERVK-silenced (right) versus control iN of genes classified with Allen Brain Atlas Gene Ontologies (GOs) compared to random genes (*P* values, Wilcoxon test). (C) Examples of DE genes related to GABAergic pathway. (D) HERVK with (top) RNA-seqs of control versus KD and (middle) control versus HERVK-CRISPRi iN, and (bottom) KZFPs ChIP-seqs in hESC. (E) Western blot of NCCIT cell lysates probed with anti-ENV or anti-actin antibodies. (F) Expression of ISG (with fold change > 5 in IFN-treated normal tissues or cells according to the Interferome database) in KD versus control iN. In red, DE ISG (fold change > 2, FDR < 0.05). Venn diagram indicates number of DE genes stimulated by each IFN type. (G) Fold change expression of antiviral ISG in indicated conditions versus control iN (ND, not detected). (H) Forty-three-day organoids, with size quantification underneath (*P* values, Mann Whitney *U* test). (I) PAX6 immunostaining and RT-qPCR quantification (*P* values, two-tailed *t* test), and (J) fold change expression of neuronal function-related genes and (K) ISG in 20-day organoids. Independent duplicates (A to G) or triplicates (D, J, and K) were used. Error bars represent the SD (***FDR < 0.001, **FDR < 0.01, and *FDR < 0.05).

Transcription from the *PRODH* promoter was found to be enhanced by SOX2 recruitment to the sequences in the upstream HERVK LTR (19). Accordingly, the LV-mediated transduction of hESC with a GFP cassette placed downstream of a 3215-bp sequence encompassing these two *PRODH* regulatory elements resulted in SOX2 binding site-dependent expression of the reporter, which was boosted by knocking down ZNF417 and ZNF587 (Fig. 3D). Furthermore, when the PBS in this vector was replaced by the PBS-R element originally identified as a strong ZNF417/587 recruiter, GFP expression was weak in control hESC, indicating that, in this case, the repressors prevailed over the activator, the action of which was revealed in the KD cells. This experiment illustrates both the TF-regulated enhancer potential of a TEeRS and the subtlety of its control by cognate KZFPs.

To test functionally whether ZNF417/ZNF587-targeted TEeRS act as neuronal enhancers, we first used an in vitro neuronal differentiation system where the doxycycline-inducible expression of neurogenin-1 and neurogenin-2 in human induced pluripotent stem cells (hiPSCs) triggers their high-efficiency differentiation into bipolar neurons (20) with TE expression tightly regulated during the differentiation process (fig. S4A). We perturbed this system by either decreasing (via RNA interference) or increasing (via overexpression) the levels of the two KZFPs or by repressing some of their HERVK and SVA targets with a CRISPR-based system (CRISPRi) (21). ZNF417/587-depleted iPSCs displayed a dysregulation of genes and TEs very reminiscent of that observed in hESC (Fig. 4A). Neurons derived therefrom were characterized by the aberrant expression of nonneuronal genes (e.g., endothelium) (22) and the up-regulation of transcripts related to potassium channel activity or to GABAergic neurotransmission (e.g., *PRODH*), whereas by contrast HERVK/SVA-CRISPRi-modified neurons displayed an induction of sodium channel-associated RNAs and a drop in GABAergic-related transcripts (Fig. 4, B and C, and fig. S4, B to D). Furthermore, among 160 HERVKs predicted to encode for at least fragments of a retroviral envelope protein (ENV) recently demonstrated to be neurotoxic in the mouse brain and up-regulated in cortical and anterior horn neurons of patients with sporadic amyotrophic lateral sclerosis (23), we found 15 to be targeted by ZNF417/587 and 6 of these to be up-regulated upon ZNF417/587 KD in iPSC-derived neurons (Fig. 4D). While ENV protein could not be easily detected in these cells, its induction was verified in NCCIT cells depleted for ZNF417/587 (Fig. 4E). We also observed an up-regulation of alternative transcripts of the ENV mRNA coding for NP9 and Rec proteins, known for their oncogenic potential and ability to stimulate expression of the IFN-induced transmembrane protein 1 (IFITM1) antiviral factor (24–28). In addition, KZFP-depleted iN (induced neurons) were characterized by the up-regulation of IFN γ and ISGs such as *TNF*, *IFITMs*, *APOBEC3B*, *IRF1*, *IFIH1* (also known as *MDA5*), *IFI44L*, *MOV10*, *RTP4*, and *Bst2* (Fig. 4, F and G, and table S3) (29). Aberrant cytoplasmic accumulation of DNA is known to activate the double-stranded DNA (dsDNA) sensor STING (also known as TMEM173) and to promote an innate immune reaction (30). However, the ISG induction observed in KZFP KD cells was only partly abrogated by inhibiting STING (Stimulator of IFN genes) (fig. S4E), suggesting that it was due not only to a cytoplasmic increase of retroviral reverse transcript dsDNAs but also likely to additional TE-derived products as observed in astrocytes and neuronal progenitor cells, respectively, deficient for the Aicardi-Goutières syndrome-associated DNA exonuclease TREX1 and the dsRNA-editing enzyme ADAR1, and upon Rec overexpression or treatment with inhibitors of DNA

methyltransferases promoting ERV RNA expression in cancer cells (26, 31–35). Reciprocally, levels of several ISG transcripts were decreased in HERVK-CRISPRi iN (Fig. 4G) and in ZNF417/587-overexpressing iPSCs (fig. S4F).

Last, brain organoids derived from ZNF417/587-KD H1 hESC all exhibited neuroepithelial expansion upon embedding and were similar in size compared to controls at early time points of differentiation (fig. S5, A and B). However, a progressive decrease in KD-derived organoid size was observed at days 33 and 43, with less rosette-like structures of apical-basal cell polarity and more PAX6-expressing early progenitors (Fig. 4, H and I, and fig. S5, A to C). Staining with antibodies against the TUJ1 and SOX2 neuronal markers further emphasized the differential organization of control and KD organoids at day 20 of differentiation (fig. S5D). At this stage, we also found a greater abundance of Ki67-positive proliferating cells upon KZFP depletion, with no obvious difference in the frequency of cleaved caspase 3-positive apoptotic cells apart from the central region in control organoids where these cells were very abundant (fig. S5, E and F). The ultimately smaller size of the KD organoids, despite their initially high content in proliferating cells, thus likely reflected the global impact on their regional disorganization. Using publicly available single-cell RNA sequencing (RNA-seq) on the entire course of cerebral organoid differentiation (36), we determined that many genes up-regulated in KD organoids were normally expressed either in early neural stem cells or in regions of the brain rich in endothelial cells like the choroid plexus, whereas a substantial subset of down-regulated genes were normally expressed in cells further differentiated in hindbrain and midbrain, cortical, and ventral progenitors (fig. S5G). Last, increased levels of LTR/ERVK, SVA, and LTR/ERV1 RNA; alterations of neurotransmitter expression profiles; and an inflammatory response reminiscent of that observed in KZFP-depleted pluripotent stem cells and neuron derivatives could be detected in KD-derived organoids as well (Fig. 4, J and K, and table S4).

DISCUSSION

In summary, the present work demonstrates that rather than just silencing TE-embedded regulatory sequences during early embryogenesis, human KZFPs keep controlling their transcriptional impact later in development and in adult tissues. Our results further indicate that the evolutionary selection of some KZFP genes was key to the domestication of evolutionarily recent TEeRS toward the genesis of transcription networks active in human brain. They finally imply that interindividual differences in ZNF417, ZNF587, or their target TE-derived loci, many of which are species specific and display some polymorphism in the human population, might translate into variations in brain development, function, and disease susceptibility.

MATERIALS AND METHODS

Plasmid and LV production

LVs with a 39-bp sequence encompassing the HERVK14C PBS were described previously (10). The 3215-bp sequence encompassing the *PRODH* promoter and upstream antisense HERVK LTR5hs and PBS was amplified by polymerase chain reaction (PCR) from H1 cell genomic DNA using an antisense (a)LTR-*PRODH* F/R primers and cloned upstream of the GFP open reading frame (HERVK-*PRODH* LV). Mutations of the two Sox2 binding sites in the LTR were introduced by sewing PCR using fragments amplified with aLTR-*PRODH*

F/mutSox2.2R, mutSox2.1F/4R, and mutSox2.3F/aLTR-PRODH R primers and assembled in a final PCR with aLTR-PRODH F/R primers. Mutation in the HERVK PBS was inserted in HERVK-PRODH LV with the QuikChange II Site-Directed Mutagenesis Kit (Agilent) and primers PBS.R F/R to obtain the motif present in the so-called R PBS. The hairpin sequences against ZNF417/587 were designed using the Genetic Perturbation Portal of the Broad Institute and inserted with Age I/Eco RI into pLKO.1.puro LV (TRCN0000018002; Sigma-Aldrich). After assessing the efficiency of various hairpins for depleting the two KZFPs, the short hairpin RNA (shRNA) targeting the sequence GCAGCATATTGGAGAGAAATT was used alone in all experiments except for the RNA-seq in H1 hESC, where it was combined with one targeting AGTCGAAAGAGCAGCCTTATT in two of the four replicates. The pLKO.shDnmt3A/3B vectors were previously described (10). The pLKO.shLuc and pLKO empty vectors served as controls. KD efficiency was verified by quantitative reverse transcription PCR (qRT-PCR) after puromycin selection and at the time of sample collection for each experiment. LV containing Cas9KRAB and LTR5hs/SVAs single-guide RNA (ACCCCGTGCTCTCTGAAACAC) was designed and validated as previously described (8) using a Cas9KRAB-expressing empty backbone as control. The ZNF417-HA and ZNF587-HA coding sequences from the previously described pTRE vectors (9) were swapped with the GFP cassette of the pRRL.PGK.GFP LV (plasmid #12252; <http://www.addgene.org/>) to generate KZFP-overexpressing vectors used in stem cells. LV production and titration protocols are detailed at <http://tronolab.epfl.ch>, and lentiviral packaging plasmids are available at Addgene (plasmids #12260 and #12259). Viral titers were determined on HCT116, and an equivalent number of transducing units were used for all vectors in each experiment.

Cell culture

NCCIT cells were grown in RPMI 1640 (Thermo Fisher Scientific, Waltham, MA, USA), supplemented with 10% fetal bovine serum, 1× GlutaMAX, 1× non-essential amino acids, and 1× antibiotics. Murine J1 ESCs stably transduced with the various LV-HERVK vectors were grown on gelatin, split with Accutase, maintained in 2i + LIF (leukemia inhibitory factor) medium, and transduced with the doxycycline-inducible pTRE-ZNF LVs at a multiplicity of infection (MOI) of 60. hiPSC and H1 hESC (WA01, WiCell) were maintained on hESC-qualified Matrigel (BD Biosciences) as previously described (10) and as approved by the Swiss Federal Office of Public Health and the Canton of Vaud Ethics committee. Stem cells were transduced at an MOI of 4, 15, or 100 for KD, CRISPRi, or overexpression experiments, respectively, and selected with puromycin (0.5 μg/ml) for 3 days for KD and CRISPRi. Puromycin-selected hiPSCs were differentiated in induced neurons by addition of doxycycline (1 μg/ml) for 4 days to induce expression of the exogenous neurogenins as previously described (20). Lyophilized STINGi (37) was freshly reconstituted in dimethyl sulfoxide, added to hiPSCs at 0.5 μM 45 min before transduction with KD LVs, and replaced each day.

Brain organoids were differentiated from H1 ESCs transduced at the same MOI in different wells but, in parallel, with LVs expressing either the shZNF or the empty control vector. Three additional control wells were processed at the same time—one receiving an LV expressing a sh targeting a non-TE binding ZNF, one an irrelevant sh specific for luciferase, and one was not transduced—to control for normal macroscopic development of ZNF417/ZNF587 wild-type (WT)

organoids. Cells transduced with the various LVs were puromycin-selected for 3 days, and 9000 of the puromycin-selected cells from each condition were seeded into 96-well low-binding plate to begin organoid generation using the STEMdiff Cerebral Organoid Kit (#08570, STEMCELL Technologies) as per the manufacturer's protocol. Day 0 is denoting the transfer of H1 hESC to a 96-well plate for differentiation into embryoid bodies. Sample size was monitored at three different time points along differentiation (days 20, 33, and 43), and immunostaining and RNA-seq were performed with day 20 organoids, with KD and control organoids always treated in parallel and collected at the same time for downstream analyses.

Protein analyses

Proteins were extracted from pTRE plasmid-transfected 293T cells in 150 mM NaCl, 1% NP-40, and 0.5% sodium deoxycholate (NaDOC) and immunoprecipitated with either mouse immunoglobulin GM (IgGM) (#I5381, Sigma-Aldrich) or anti KAP1 (#MAB3662, Millipore) antibodies. Immunoprecipitated material was loaded onto 4 to 20% SDS-polyacrylamide gel electrophoresis and transferred to membranes for probing with anti-HA-Horseradish peroxidase conjugated (anti-HA)-HRPO (#12013819001, Roche) and rabbit anti-RBCC (#ab10483, Abcam). Proteins were extracted from NCCIT with radioimmunoprecipitation assay buffer, wet-transferred overnight on polyvinylidene difluoride membranes, and blotted with rabbit anti-ENV (#HERM-1811-5, AUSTRAL Biologicals, 1:3000) and anti-actin-HRPO (#ab20272, Abcam) antibodies.

GFP reporter assay

Murine cells treated or not with doxycycline and transduced with pTRE-ZNFs or hES KD and control cells selected for puromycin resistance were plated in 12-well plates and transduced in duplicates with the same numbers of transducing units for each LV-HERVK-PGK or LV-PRODH LVs. GFP signal was read by flow cytometry.

RNA extraction, quantification, and sequencing

At least duplicates from two independent experiments were collected per sample. RNA-seq was performed with triplicates of day 20 organoids derived in parallel from the shZNF417/587 or the non-relevant shLuc recipient H1 hESC. Total RNA was extracted with either the High Pure RNA Isolation Kit (Roche), NucleoSpin RNA XS (Macherey-Nagel), or RNeasy Kit (Qiagen) with an on-column deoxyribonuclease treatment. For RNA-seq, sample libraries were prepared using a TruSeq stranded mRNA sample preparation kit (Illumina). Libraries were sequenced on an Illumina Hi-Seq machine with stranded 75-base single or paired-end reads for the CRISPRi and ZNF KD or STINGi-related RNA-seqs, respectively. All RT-qPCRs were performed with RNAs from at least independent biological duplicates, using random hexamers and SuperScript II (Invitrogen) or Maxima H Minus (Thermo Fisher Scientific) to generate complementary DNAs (cDNAs). Each cDNA was quantified in triplicates with SYBR Green Mix (Applied Biosystems). The $-\Delta\Delta C_t$ method was used to calculate fold change. Negative controls without RT enzyme were processed in parallel. Primers used are described in table S5.

RNA-seq mapping and analysis

Raw data from time course RNA-seq triplicates used in fig. S4A were downloaded from GSE60548. For the BrainSpan Atlas and GTEx RNA-seq remapping, raw SRA (Sequence Read Archive) files were

downloaded from the dbGaP (Database of Genotypes and Phenotypes) data portal using the prefetch tool from National Center for Biotechnology Information (NCBI) SRA Toolkit v2.9.2 and then converted to FASTQ using the fastq-dump tool (with `-split-3` for paired-end data). Reads were mapped to the human genome (hg19) using HISAT2 (v2.1.0) (38) with default parameters. Gene counts were generated using uniquely mapped reads with HTSeq-count (for KD hESC RNA-seqs) (39) or featureCounts (iPS/iN and CRISPRi RNA-seqs) (40). TE counts were generated using the uniquely mapped reads using the multiBamCov tool from the BEDTools software (for KD hESC RNA-seqs) and featureCounts for all others. To avoid read assignment ambiguity between genes and TEs, a gtf file combining both features was provided to featureCounts. TE reads in exons were always dismissed. Genes and TEs with low counts (less reads than there are samples) were discarded. Normalization for sequencing depth was done for both genes and TEs using the Trimmed Mean of *v*-values (TMM) method as implemented in the limma package of Bioconductor (41) and using the counts on genes as library size. Differential gene expression analysis was performed using Voom (42) as it has been implemented in the limma package of Bioconductor. A moderated *t* test (as implemented in the limma package of R) was used to test significance. *P* values were adjusted for multiple testing using the Benjamini-Hochberg's method. A gene (or TE) was considered to be DE when the absolute fold change between groups was bigger than 2 and the FDR (adjusted *P* value) was smaller than 0.05 ($***P < 0.001$, $**P < 0.01$, and $*P < 0.05$). Interspecies RNA-seq normalization was performed as previously described (8).

Repeat merging and aging

The RepeatMasker 4.0.5 (Library 20140131, <http://www.repeatmasker.org/species/hg.html>) was used to generate an in-house repeats list, where fragmented annotated ERVs have been reassembled in one single element when relevant and according to the following procedure: ERV fragments are annotated as either fragmented internal parts (ERV-int) or LTRs. Starting from each annotated ERV-int element, we first computed a frequency table of the surrounding retroelement fragments. We then computed the frequency distributions of each LTR type for each ERV-int subfamily and performed a Wald test to assign a *P* value to each LTR/ERV-int pair. To merge an LTR to an ERV-int fragment, it had to represent more than 2% of the pairs, the *P* value had to be smaller than 0.001, the two elements had to be in the same orientation, and the distance between the two elements had to be shorter than 100 bp. The name of the internal part was given to the resulting LTR/ERV-int merged element. For example, an element resulting from LTR7 and HERVH-int fragment merging is identified as HERVH-int. Relevance of LTR and internal piece merging was verified when possible in the literature and in the Dfam database of repetitive DNA based on profile hidden Markov models (dfam.org). Fragmented ERV-int or fragmented LTRs from the same subfamily (same name in Repbase database) with the same orientation and closer than 100 bp were also merged.

We used RNA-seqs with a depth superior to 30 to 40 million reads and at least 75 bp in length as we determined, by comparing various RNA-seq approaches for their respective yield in mappable TE transcripts, that it allowed us to assign more than 91% of repeat-derived reads to specific unique genomic loci through our TE-optimized analytical pipeline.

TE age was estimated by measuring substitution density of TE loci by comparison with their respective subfamily consensus. Sub-

stitutions were then separated between transitions and transversions to feed a Kimura two-parameter model (in-house script) (43). To estimate an evolutionary age from Kimura substitution levels, we assumed a constant rate of substitution per million years of 2.2×10^{-3} as described previously (44), which allowed us to obtain divergence time in million years for each TE locus.

Chromatin immunoprecipitation sequencing

KAP1 ChIP-seq has been described earlier (10). Chromatin for ZNF-HA ChIPs was prepared in triplicates from 3×10^7 cells and chromatin for histone ChIPs in at least duplicates from two independent experiments with 10^7 cells, as followed: Adherent stem cells were detached with TrypLE, diluted in Dulbecco's minimum essential medium (DMEM) KO medium, and washed with phosphate-buffered saline. Cells were fixed with methanol-free formaldehyde (333 mM final) for 10 min at room temperature, diluted with tris (pH 8) (200 mM final), and washed twice with cold phosphate-buffered saline, and pellets were frozen at -80°C . Cross-linked cells were incubated for 10 min at 4°C in buffer I [50 mM HEPES-KOH (pH 7.4), 140 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 10% glycerol, 0.5% NP-40, and 0.25% Tx100, supplemented with protease inhibitors], spun, incubated for 10 min at 4°C in buffer II [10 mM tris (pH 8), 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, and protease inhibitors], spun, washed three times at 4°C in buffer III [buffer II supplemented with 0.1% NaDOC and 0.25% N-Lauroylsarcosine sodium salt (NLS)], and resuspended in 1 ml of buffer III before being transferred to a TC12X12 AFA fiber tube (Covaris). Sonication was performed with an automated Covaris Focused-ultrasonicator with the following parameters: 20 min, 5% duty cycle, 140 W, and 200 cycles, and extracts were cleared by centrifugation for 10 min at 4°C . Chromatin sonication was evaluated on a Bioanalyzer DNA chip (Agilent) with one aliquot of material treated with ribonuclease A overnight at 65°C and purified with Qiagen MinElute columns. Immunoprecipitations were performed in low DNA binding tubes using Dynabeads (Thermo Fisher Scientific) after a pre-clearing step for histone ChIPs and as recommended by the manufacturer. Antibodies specific for HA (#901503, BioLegend), H3K9me3 (#C15410056, Diagenode), H3K4me1 (#pAb-037-050, Diagenode), and H3K27ac (#ab4729, Abcam) were used. Immunoprecipitated material was eluted and treated with proteinase K overnight at 65°C in EB (TE 1× and 1% SDS) and purified on Qiagen MinElute columns. Total inputs and chromatin immunoprecipitated samples were prepared for libraries using 10 ng of material. Aliquots of the samples were tested in qPCR using the KAPA SYBR FAST qPCR Master Mix (Kapa Biosystems, KK4604) to determine the optimal number of PCR cycles needed to amplify each library without reaching saturation. Libraries were amplified with KAPA HiFi HotStart Readymix polymerase (Kapa Biosystems, KK2602), size-selected using Ampure XP beads (Beckman Coulter), and always quality-checked on a Bioanalyzer DNA high sensitivity chip (Agilent) and quantified with a Qubit dsDNA HS assay kit (Qubit 2.0 Fluorometer, Invitrogen). ChIP enrichment was always confirmed by qPCR with positive versus negative controls before and after library preparation. Libraries were sequenced with 75-bp single or paired-end reads on an Illumina Hi-Seq machine with standard Illumina protocols.

ChIP-seq analysis

The previously published KAP1 ChIP-seq (10) was remapped with our updated pipeline. Reads were mapped to the human genome

(hg19) using the short read aligner program Bowtie2, with the sensitive local mode (the exact parameters are as follows: bowtie2 -p 6 -t --sensitive-local -x \$index -U \$reads). Only reads with a mapping quality score higher than 10 were retained. Before peak calling, BAM files were filtered removing reads aligning to ENCODE blacklist regions and to regions of ChIP experiments with high signal in the input using the Bioconductor package GreyListChIP (<https://bioconductor.org/packages/GreyListChIP>). Peaks for histone marks were called via EPIC [reimplementation of SICER (45)], keeping only peaks with an FDR below 1%. Peaks for ZNF417, ZNF587, and KAP1 were called using MACS2 (46) (with --BAMPE option as from paired end reads), keeping only peaks with score higher than 80. Peaks from biological replicates were merged using the Bioconductor package DiffBind (<https://bioconductor.org/packages/DiffBind>), and when triplicates were made, only peaks that overlapped in at least two peak sets were kept and merged into consensus peaks. The intersectBed tool (with default parameters and unique option) from the BEDTools suite (47) was used to calculate intersections, and the genomeCoverageBed tool was used to generate coverage files, which were converted into bigWig files with the bedGraphToBigWig tool (provided by University of California Santa Cruz).

Multiple sequence alignment plot

FASTA sequences from KZFP overlapping ERVK integrants were extracted from the hg19 genome assembly and processed as described previously (8), except that regions in the alignment consisting of more than 90% of gaps were trimmed out. For each aligned integrant, the ZNF417, ZNF587, KAP1, or H3K9me3 ChIP-seq signal was extracted from the BAM alignment files using the python pysam library and scaled to the interval [0,1] before being superimposed to the alignment. The average ChIP-seq signals were plotted on top.

SMILE-seq with methylated probes

SMILE-seq device and analysis was previously described (14). Motif enrichment analysis was performed using HOMER (48). Methylated and unmethylated probes were used here.

Human genetic data

Information on human genetic data and LoF mutations was obtained from gnomAD (release-2.0.2) (49, 50) containing exome and whole-genome sequencing data for 123,136 and 15,496 individuals, respectively.

Immunofluorescence on brain organoids

Day 20 organoids were fixed with 4% paraformaldehyde (pH 7.4) for 15 min, washed twice with phosphate-buffered saline, and placed in 30% sucrose solution overnight at 4°C. Fixed organoids were placed in cryomolds, covered with cryomatrix embedding resin (#6769006, Thermo Fisher Scientific), snap-frozen with isopentane, and followed by cutting 8- μ m sections with a cryostat (Leica) to Superfrost Plus slides (Thermo Fisher Scientific). For immunofluorescence staining, tissue was permeabilized by 0.25% Triton X-100 for 10 min, washed three times with phosphate-buffered saline, blocked for 30 min in 1% bovine serum albumin in phosphate-buffered saline, incubated with primary antibody [PAX6, #561462, BD Biosciences (1:200); SOX2, #AB5603, Chemicon (1:300); TUJ1, #MMS-435P, Covance (1:750); Ki67, #M3060, Spring AMBIO

(1:250); cleaved caspase 3, #9661, Cell Signaling Technology (1:400)] overnight at 4°C, and then incubated with secondary Alexa Fluor antibody (Thermo Fisher Scientific) for 40 min. Nuclear staining was performed with 4',6-diamidino-2-phenylindole, and slides were mounted with Fluoromount-G (#00-4958-02, Thermo Fisher Scientific). Fluorescence imaging was performed with a Leica DM5500 microscope or a Leica SP8 microscope. For wide-field organoid images, a Leica stereoscope was used and the outer area of individual organoids was measured with FIJI using automatic thresholding. To determine the contribution of immature neuroectoderm or rosette-like structures to organoids, FIJI was used with the following criteria. Neuroectodermal-like regions with continual PAX6-positive staining were manually drawn around, and rosette-like structures were determined via the presence of an obvious lumen, flanked by cells of apical basal polarity. The percentage overall contribution of both structures was measured as a proportion of the whole organoid, excluding any large internal cavities. Quantification of Ki67-positive cells was performed with FIJI. Channels were split and converted to 8-bit images and an intensity threshold of 160, and “watershed” was applied to separate touching cells. The tool “Analyze Particles” was used with “show outlines” to demarcate and record the number of Ki67-positive cells. To obtain representative cell counts throughout the organoids, this was repeated on three serial sections separated by approximately 80 to 100 μ m, resulting in an average number of Ki67-positive cells per organoid.

Allen brain transcriptome mapping

Only uniquely mapped reads and TEs with more than 50 reads were retained for analysis. The “cerebroViz” R package was used to generate the brain spatiotemporal gene expression plots (51).

Analysis of signature gene expression in organoids

Signature genes were identified from (36). In each category, the fold change in expression in KD versus control organoids of each expressed signature gene was plotted in yellow, blue, or orange when not DE, down-regulated, or up-regulated, respectively (adjusted $P < 0.05$; fold change, 2). Fold change in expression of the complete set of signature genes is plotted in gray in each category. P values were computed with hypergeometric test (phyper of R).

Statistical analyses

R version 1.1.447 and Wilcoxon test were used for statistical analyses of boxplots. Details of the other statistical tests are given in the figure legends or in the respective method subsections. In all analyses, *** $P < 0.001$, ** $P < 0.01$, and * $P < 0.05$.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/6/35/eaba3200/DC1>

REFERENCES AND NOTES

1. A. P. J. de Koning, W. Gu, T. A. Castoe, M. A. Batzer, D. D. Pollock, Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet.* **7**, e1002384 (2011).
2. G. Bourque, B. Leong, V. B. Vega, X. Chen, Y. L. Lee, K. G. Srinivasan, J.-L. Chew, Y. Ruan, C.-L. Wei, H. H. Ng, E. T. Liu, Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Res.* **18**, 1752–1762 (2008).
3. V. Sundaram, Y. Cheng, Z. Ma, D. Li, X. Xing, P. Edge, M. P. Snyder, T. Wang, Widespread contribution of transposable elements to the innovation of gene regulatory networks. *Genome Res.* **24**, 1963–1976 (2014).

4. M. Trizzino, Y. Park, M. Holsbach-Beltrame, K. Aracena, K. Mika, M. Caliskan, G. H. Perry, V. J. Lynch, C. D. Brown, Transposable elements are the primary source of novelty in primate gene regulation. *Genome Res.* **27**, 1623–1633 (2017).
5. E. B. Chuong, N. C. Elde, C. Feschotte, Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science* **351**, 1083–1087 (2016).
6. H. M. Rowe, J. Jakobsson, D. Mesnard, J. Rougemont, S. Reynard, T. Aktas, P. V. Maillard, H. Layard-Liesching, S. Verp, J. Marquis, F. Spitz, D. B. Constam, D. Trono, KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* **463**, 237–240 (2010).
7. D. Wolf, K. Hug, S. P. Goff, TRIM28 mediates primer binding site-targeted silencing of Lys1,2 tRNA-utilizing retroviruses in embryonic cells. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 12521–12526 (2008).
8. J. Pontis, E. Planet, S. Offner, P. Turelli, J. Duc, A. Coudray, T. W. Theunissen, R. Jaenisch, D. Trono, 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 (2019).
9. M. Imbeault, P.-Y. Helleboid, D. Trono, KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. *Nature* **543**, 550–554 (2017).
10. P. Turelli, N. Castro-Diaz, F. Marzetta, A. Kapopoulou, C. Raclot, J. Duc, V. Tieng, S. Quenneville, D. Trono, Interplay of TRIM28 and DNA methylation in controlling human endogenous retroelements. *Genome Res.* **24**, 1260–1270 (2014).
11. K. Nowick, A. T. Hamilton, H. Zhang, L. Stubbs, Rapid sequence and expression divergence suggest selection for novel function in primate-specific KRAB-ZNF genes. *Mol. Biol. Evol.* **27**, 2606–2617 (2010).
12. J. H. Wildschutte, Z. H. Williams, M. Montesion, R. P. Subramanian, J. M. Kidd, J. M. Coffin, Discovery of unfixed endogenous retrovirus insertions in diverse human populations. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E2326–E2334 (2016).
13. D. C. Schultz, K. Ayyanathan, D. Negorev, G. G. Maul, F. J. Rauscher, 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 (2002).
14. A. Isakova, R. Groux, M. Imbeault, P. Rainer, D. Alpern, R. Dainese, G. Ambrosini, D. Trono, P. Bucher, B. Deplancke, SMILE-seq identifies binding motifs of single and dimeric transcription factors. *Nat. Methods* **14**, 316–322 (2017).
15. T. Gao, B. He, S. Liu, H. Zhu, K. Tan, J. Qian, EnhancerAtlas: A resource for enhancer annotation and analysis in 105 human cell/tissue types. *Bioinformatics* **32**, 3543–3551 (2016).
16. B. M. Campbell, E. Charych, A. W. Lee, T. Möller, Kynurenes in CNS disease: Regulation by inflammatory cytokines. *Front Neurosci.* **8**, 12 (2014).
17. L. Kempf, K. K. Nicodemus, B. Kolachana, R. Vakkalanka, B. A. Verchinski, M. F. Egan, R. E. Straub, V. A. Mattay, J. H. Callicott, D. R. Weinberger, A. Meyer-Lindenberg, Functional polymorphisms in PRODH are associated with risk and protection for schizophrenia and fronto-striatal structure and function. *PLOS Genet.* **4**, e1000252 (2008).
18. G. W. Crabtree, A. J. Park, J. A. Gordon, J. A. Gogos, Cytosolic accumulation of L-proline disrupts GABA-ergic transmission through GAD blockade. *Cell Rep.* **17**, 570–582 (2016).
19. M. Suntsova, E. V. Gogvadze, S. Salozhin, N. Gaifullin, F. Eroshkin, S. E. Dmitriev, N. Martynova, K. Kulikov, G. Malakhova, G. Tukhbatova, A. P. Bolshakov, D. Ghilarov, A. Garzha, A. Aliper, C. R. Cantor, Y. Solokhin, S. Roumiantsev, P. Balaban, A. Zhavoronkov, A. Buzdin, Human-specific endogenous retroviral insert serves as an enhancer for the schizophrenia-linked gene PRODH. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 19472–19477 (2013).
20. V. Busskamp, N. E. Lewis, P. Guye, A. H. Ng, S. L. Shipman, S. M. Byrne, N. E. Sanjana, J. Murn, Y. Li, S. Li, M. Stadler, R. Weiss, G. M. Church, Rapid neurogenesis through transcriptional activation in human stem cells. *Mol. Syst. Biol.* **10**, 760 (2014).
21. P. I. Thakore, A. M. D'Ippolito, L. Song, A. Safi, N. K. Shivakumar, A. M. Kabadi, T. E. Reddy, G. E. Crawford, C. A. Gersbach, Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. *Nat. Methods* **12**, 1143–1149 (2015).
22. K. W. Kelley, H. Nakao-Inoue, A. V. Molofsky, M. C. Oldham, Variation among intact tissue samples reveals the core transcriptional features of human CNS cell classes. *Nat. Neurosci.* **21**, 1171–1184 (2018).
23. W. Li, M.-H. Lee, L. Henderson, R. Tyagi, M. Bachani, J. Steiner, E. Campanac, D. A. Hoffman, G. von Geldern, K. Johnson, D. Maric, H. D. Morris, M. Lentz, K. Pak, A. Mammen, L. Ostrow, J. Rothstein, A. Nath, Human endogenous retrovirus-K contributes to motor neuron disease. *Sci. Transl. Med.* **7**, 307ra153 (2015).
24. M. Garcia-Montojo, T. Doucet-O'Hare, L. Henderson, A. Nath, Human endogenous retrovirus-K (HML-2): A comprehensive review. *Crit. Rev. Microbiol.* **44**, 715–738 (2018).
25. S. M. Chan, T. Sapir, S.-S. Park, J.-F. Rual, R. Contreras-Galindo, O. Reiner, D. M. Markovitz, The HERV-K accessory protein Np9 controls viability and migration of teratocarcinoma cells. *PLOS ONE* **14**, e0212970 (2019).
26. E. J. Grow, R. A. Flynn, S. L. Chavez, N. L. Bayless, M. Wossidlo, D. J. Wesche, L. Martin, C. B. Ware, C. A. Blish, H. Y. Chang, R. A. Reijo Pera, J. Wysocka, Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. *Nature* **522**, 221–225 (2015).
27. M. Denne, M. Sauter, V. Armbruster, J. D. Licht, K. Roemer, N. Mueller-Lantzsch, Physical and functional interactions of human endogenous retrovirus proteins Np9 and Rec with the promyelocytic leukemia zinc finger protein. *J. Virol.* **81**, 5607–5616 (2007).
28. K. Heyne, K. Kölsch, M. Bruand, E. Kremmer, F. A. Grässer, J. Mayer, K. Roemer, Np9, a cellular protein of retroviral ancestry restricted to human, chimpanzee and gorilla, binds and regulates ubiquitin ligase MDM2. *Cell Cycle* **14**, 2619–2633 (2015).
29. X. Wu, V. L. Dao Thi, Y. Huang, E. Billerbeck, D. Saha, H.-H. Hoffmann, Y. Wang, L. A. V. Silva, S. Sarbanes, T. Sun, L. Andrus, Y. Yu, C. Quirk, M. Li, M. R. MacDonald, W. M. Schneider, X. An, B. R. Rosenberg, C. M. Rice, Intrinsic immunity shapes viral resistance of stem cells. *Cell* **172**, 423–438.e25 (2018).
30. A. Ablasser, Z. J. Chen, cGAS in action: Expanding roles in immunity and inflammation. *Science* **363**, eaat8657 (2019).
31. C. H. C. Tie, L. Fernandes, L. Conde, L. Robbez-Masson, R. P. Sumner, T. Peacock, M. T. Rodriguez-Plata, G. Mickute, R. Gifford, G. J. Towers, J. Herrero, H. M. Rowe, KAP1 regulates endogenous retroviruses in adult human cells and contributes to innate immune control. *EMBO Rep.* **19**, e45000 (2018).
32. K. B. Chiappinelli, P. L. Strissel, A. Desrichard, H. Li, C. Henke, B. Akman, A. Hein, N. S. Rote, L. M. Cope, A. Snyder, V. Makarov, S. Budhu, D. J. Slamon, J. D. Wolchok, D. M. Pardoll, M. W. Beckmann, C. A. Zahnow, T. Merghoub, T. A. Chan, S. B. Baylin, R. Strick, Inhibiting DNA methylation causes an interferon response in cancer via dsRNA including endogenous retroviruses. *Cell* **162**, 974–986 (2015).
33. D. Roulois, H. Loo Yau, R. Singhanian, Y. Wang, A. Danesh, S. Y. Shen, H. Han, G. Liang, P. A. Jones, T. J. Pugh, C. O'Brien, D. D. De Carvalho, DNA-demethylating agents target colorectal cancer cells by inducing viral mimicry by endogenous transcripts. *Cell* **162**, 961–973 (2015).
34. C. A. Thomas, L. Tejwani, C. A. Trujillo, P. D. Negraes, R. H. Herai, P. Mesci, A. Macia, Y. J. Crow, A. R. Muotri, Modeling of TREX1-dependent autoimmune disease using human stem cells highlights L1 accumulation as a source of neuroinflammation. *Cell Stem Cell* **21**, 319–331.e8 (2017).
35. H. Chung, J. J. A. Calis, X. Wu, T. Sun, Y. Yu, S. L. Sarbanes, V. L. Dao Thi, A. R. Shilvock, H.-H. Hoffmann, B. R. Rosenberg, C. M. Rice, Human ADAR1 prevents endogenous RNA from triggering translational shutdown. *Cell* **172**, 811–824.e14 (2018).
36. S. Kanton, M. J. Boyle, Z. He, M. Santel, A. Weigert, F. Sanchis-Calleja, P. Gujjarro, L. Sidow, J. S. Fleck, D. Han, Z. Qian, M. Heide, W. B. Huttner, P. Khaïtovich, S. Pääbo, B. Treutlein, J. G. Camp, Organoid single-cell genomic atlas uncovers human-specific features of brain development. *Nature* **574**, 418–422 (2019).
37. S. M. Haag, M. F. Gulen, L. Raymond, A. Gibelin, L. Abrami, A. Decout, M. Heymann, F. G. van der Goot, G. Turcatti, R. Behrendt, A. Ablasser, Targeting STING with covalent small-molecule inhibitors. *Nature* **559**, 269–273 (2018).
38. D. Kim, B. Langmead, S. L. Salzberg, HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
39. S. Anders, P. T. Pyl, W. Huber, HTSeq—A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
40. Y. Liao, G. K. Smyth, W. Shi, featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
41. R. C. Gentleman, V. J. Carey, D. M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, K. Hornik, T. Hothorn, W. Huber, S. Iacus, R. Irizarry, F. Leisch, C. Li, M. Maechler, A. J. Rossini, G. Sawitzki, C. Smith, G. Smyth, L. Tierney, J. Y. H. Yang, J. Zhang, Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biol.* **5**, R80 (2004).
42. C. W. Law, Y. Chen, W. Shi, G. K. Smyth, voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).
43. M. Kimura, A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**, 111–120 (1980).
44. S. Kumar, S. Subramanian, Mutation rates in mammalian genomes. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 803–808 (2002).
45. C. Zang, D. E. Schones, C. Zeng, K. Cui, K. Zhao, W. Peng, A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics* **25**, 1952–1958 (2009).
46. J. Feng, T. Liu, B. Qin, Y. Zhang, X. S. Liu, Identifying ChIP-seq enrichment using MACS. *Nat. Protoc.* **7**, 1728–1740 (2012).
47. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
48. S. Heinz, C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh, C. K. Glass, Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
49. M. Lek, K. J. Karczewski, E. V. Minikel, K. E. Samocha, E. Banks, T. Fennell, A. H. O'Donnell-Luria, J. S. Ware, A. J. Hill, B. B. Cummings, T. Tukiainen, D. P. Birnbaum, J. A. Kosmicki, L. E. Duncan, K. Estrada, F. Zhao, J. Zou, E. Pierce-Hoffman, J. Berghout, D. N. Cooper, N. DeLafuente, M. DePristo, R. Do, J. Flannick, M. Fromer, L. Gauthier, J. Goldstein, N. Gupta, D. Howrigan, A. Kiezun, M. I. Kurki, A. L. Moonshine, P. Natarajan, L. Orozco, G. M. Peloso, R. Poplin,

- M. A. Rivas, V. Ruano-Rubio, S. A. Rose, D. M. Ruderfer, K. Shakir, P. D. Stenson, C. Stevens, B. P. Thomas, G. Tiao, M. T. Tusie-Luna, B. Weisburd, H.-H. Won, D. Yu, D. M. Altshuler, D. Ardissino, M. Boehnke, J. Danesh, S. Donnelly, R. Elosua, J. C. Florez, S. B. Gabriel, G. Getz, S. J. Glatt, C. M. Hultman, S. Kathiresan, M. Laakso, S. McCarroll, M. I. McCarthy, D. McGovern, R. McPherson, B. M. Neale, A. Palotie, S. M. Purcell, D. Saleheen, J. M. Scharf, P. Sklar, P. F. Sullivan, J. Tuomilehto, M. T. Tsuang, H. C. Watkins, J. G. Wilson, M. J. Daly, D. G. MacArthur; Exome Aggregation Consortium, Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
50. K. J. Karczewski, L. C. Francioli, G. Tiao, B. B. Cummings, J. Alfoldi, Q. Wang, R. L. Collins, K. M. Laricchia, A. Ganna, D. P. Birnbaum, L. D. Gauthier, H. Brand, M. Solomonson, N. A. Watts, D. Rhodes, M. Singer-Berk, E. G. Seaby, J. A. Kosmicki, R. K. Walters, K. Tashman, Y. Farjoun, E. Banks, T. Poterba, A. Wang, C. Seed, N. Whiffin, J. X. Chong, K. E. Samocha, E. Pierce-Hoffman, Z. Zappala, A. H. O'Donnell-Luria, E. V. Minikel, B. Weisburd, M. Lek, J. S. Ware, C. Vittal, I. M. Armean, L. Bergelson, K. Cibulskis, K. M. Connolly, M. Covarrubias, S. Donnelly, S. Ferriera, S. Gabriel, J. Gentry, N. Gupta, T. Jeandet, D. Kaplan, C. Llanwarne, R. Munshi, S. Novod, N. Petrillo, D. Roazen, V. Ruano-Rubio, A. Saltzman, M. Schleicher, J. Soto, K. Tibbetts, C. Tolonen, G. Wade, M. E. Talkowski; Genome Aggregation Database Consortium, B. M. Neale, M. J. Daly, D. G. MacArthur, Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv* 531210 [Preprint]. 13 August 2019. <https://doi.org/10.1101/531210>.
51. E. Bahl, T. Koomar, J. J. Michaelson, *cerebroViz*: An R package for anatomical visualization of spatiotemporal brain data. *Bioinformatics* **33**, 762–763 (2017).

Acknowledgments: We thank E. Planet for guidance with the bioinformatics; the EPFL Genomics, Flow Cytometry, and Histology core facilities for help with the relevant

technologies; A. Ablasser (EPFL) for the STING inhibitor; R. Fueyo and J. Wysocka (Stanford University) for the NCCIT cell line and useful tips for ENV detection; R. Dainese (EPFL) for advice for the SMILE-seq experiment; and other members of the Trono Lab for helpful discussions. **Funding:** This study was supported by grants from the Personalized Health and Related Technologies (PHRT-508), the European Research Council (KRABnKAP, #268721; Transpos-X, #694658), and the Swiss National Science Foundation (310030_152879 and 310030B_173337) to D.T. **Author contributions:** P.T. and D.T. conceived the study, interpreted the data, and wrote the manuscript. P.T. designed, performed, and analyzed experiments. D.G. performed most of the bioinformatics analyses, partly using tools developed by J.D., A.C., and C.T. In vitro and cell-based experiments were performed mostly by C.R. and, for some, by J.P. and E.V.P., while C.P. took care of all organoid-related procedures. B.D. and V.B. gave intellectual input. All authors reviewed the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** Sequencing data and processed files have been submitted to the NCBI Gene Expression Omnibus under accession number GSE144192, and all codes are available upon request to authors.

Submitted 22 November 2019

Accepted 16 July 2020

Published 28 August 2020

10.1126/sciadv.aba3200

Citation: P. Turelli, C. Playfoot, D. Grun, C. Raclot, J. Pontis, A. Coudray, C. Thorball, J. Duc, E. V. Pankevich, B. Deplanck, V. Busskamp, D. Trono, Primate-restricted KRAB zinc finger proteins and target retrotransposons control gene expression in human neurons. *Sci. Adv.* **6**, eaba3200 (2020).