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

Authors: Priscilla Turelli[1], Christopher Playfoot[1], Dephine Grun[1†], Charlène Raclot[1†], Julien Pontis[1], Alexandre Coudray[1], Christian Thorball[1], Julien Duc[1], Eugenia Pankevich[1], Bart Deplancke[1], Volker Busskamp[2,3] and Didier Trono[1]*

Affiliations:

[1]School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland.

[2]Center for Regenerative Therapies, Technische Universität Dresden, Germany.

[3]Faculty of Medicine, Department of Ophthalmology, University of Bonn, Germany.

[†]Equal contribution

*Correspondence to: didier.trono@epfl.ch

Abstract:

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 coopted 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 (ESC). 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.

One Sentence Summary: Young transposable elements and their protein controllers team up to regulate the differentiation and function of human neurons.

Main Text:

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, which dock KAP1 (KRAB-associated protein 1, a.k.a. TRIM28) and associated heterochromatin inducers to 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.

We previously determined that a 35bp-long TE sequence encompassing the HERVK14C primer binding site (PBS)-encoding region (coined HERVK-R) confers KAP1-induced repression to a nearby PGK promoter in 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 lentivector (LV) (**Fig. 1A**), while producing them in murine ESCs silenced this vector, demonstrating the sequence-specific repressor potential of ZNF417 and ZNF587 and the likely absence of mouse orthologue (fig. S1A). Phylogenetic analyses confirmed that *ZNF417* emerged in the human ancestral genome 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, the series of amino acids trios predicted to dictate the sequence specificity of their DNA binding (**Fig. 1B** and fig. S1, B and C). Only rare individuals harbor homozygous loss of function (LoF) mutations in *ZNF417* or *ZNF587* in the gnomAD repertoire (https://gnomad.broadinstitute.org/) (**Fig. 1B**), and the two genes exhibit fairly comparable patterns of expression across tissues according to GTEX (https://gtexportal.org/) 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 in pre-natal compared to post-natal brain structures (**Fig. 1C** and fig. S1D).

Chromatin immunoprecipitation- sequencing (ChIP-seq) of H1 hESCs overexpressing HA-tagged versions of ZNF417 and ZNF587 identified 321 and 451 peaks, respectively, including 171 in common. About 85% mapped to primate-restricted LTR/ERVK, SVAs and LTR/ERV1 (**Fig. 1D** and fig. S2, A to C), amongst which 12 human-specific LTR/ERVK, and 4 out of 8 HML-2 HERVK previously noted to be polymorphic in the population (*11*) (table S1). KAP1, which binds both KZFPs (fig. S2D), and H3K9me3, the repressive mark instated by the KAP1-associated histone methyltransferase SETDB1, were enriched at the PBS-coding sequence of ZNF417/ZNF587-bound LTR/ERVKs in hESCs (fig. S2E). 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 (*12*) revealed that ZNF417 and ZNF587 had a higher affinity for methylated than unmethylated versions of this sequence (fig. S2F). Correspondingly, these KZFPs inefficiently repressed the HERVK-R-PGK-GFP LV in hESCs depleted for the *de novo* DNA methyltransferases DNMT3A and 3B, although this might also reflect indirect effects (fig. S2G).

The knockdown (KD) of *ZNF417/ZNF587* in hESC resulted in upregulating (fold change>2, 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 knockdown ESC (**Fig. 2B**, top). Expression of 854 genes was also altered (**Fig. 2A**, right, and table S2), a majority upregulated (fold change>2, FDR< 0.05), some (P=0.039, one-sided Fisher's exact test) with a ZNF417/587 peak within 100kb of the transcription start site (TSS). Those within 20kb 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-sensitive genes (ISGs) and the putative targets of 31 upregulated 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 the EnhancerAtlas 2.0 (*13*). Accordingly, several genes normally expressed in the brain stood out amongst transcriptional units upregulated in hESC depleted for ZNF417/587. For instance, *AADAT*, the product of which facilitates the synthesis of the neuroprotective kynurenic acid (*14*), and *PRODH*, a gene highly expressed in the brain where it influences GABAergic neurotransmission and previously linked to schizophrenia (*15*, *16*),

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).

To test functionally whether ZNF417/ZNF587-controlled TEeRS act as neuron-specific enhancers, we first used an *in vitro* neuronal differentiation system where the doxycycline-inducible expression of Neurogenin-1 and -2 in human induced pluripotent stem cells (iPSC) triggers their high-efficiency differentiation into bipolar neurons (*17*) with TEs 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) (*18*). ZNF417/587-depleted iPSCs displayed a dysregulation of genes and TEs very reminiscent of that observed in hESCs (**Fig. 4A**). Neurons derived therefrom were characterized by the aberrant expression of non-neuronal genes (e.g. endothelium) (*19*) and the upregulation 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 and calcium channel-associated RNAs and a drop in GABAergic-related transcripts (**Fig. 4, B and C** and fig. S4, B and C). Furthermore, amongst 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 upregulated in cortical and anterior horn neurons of patients with sporadic amyotrophic lateral sclerosis (ALS) (*20*), we found fifteen to be targeted by ZNF417/587 and six of these to be upregulated 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 upregulation of the *np9* and *rec* alternative transcripts, which

encode for proteins with oncogenic potential and promoting expression of the IFITM1 antiviral factor (*21–25*). As well, KZFP-depleted iN were characterized by the upregulation of IFNγ and ISGs such as TNF, IFITMs, APOBEC3B, IRF1, IFIH1 (a.k.a. MDA5), IFI44L, MOV10, RTP4, and Bst2 (**Fig. 4, F and G** and table S3) (*26*). This phenomenon was only partly abrogated by inhibiting the cytoplasmic DNA sensor STING (a.k.a.TMEM173) (fig. S4D), suggesting that it was not just due to ERV or SVA reverse transcripts but likely to additional TE-derived products as observed in *TREX1*- or *ADAR1*-inactivated astrocytes or neuronal progenitor cells, respectively, and upon *Rec* overexpression or treatment with inhibitors of DNA methyltransferases (*23*, *27–31*). Reciprocally, levels of several ISG transcripts were decreased in HERVK-CRISPRi iN (**Fig. 4G**) or in ZNF417/587-overepressing iPSCs (fig. S4E). Finally, brain organoids derived from *ZNF417/587*-knockdown hESCs were smaller in size and displayed a greater abundance of PAX6-expressing immature cells than controls (**Fig. 4, H and I**), as well as increased levels of LTR/ERVK, SVAs 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 (**Fig. 4, J and K** and fig. S4F).

In sum, these results demonstrate 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. They further indicate that the evolutionary selection of some KZFP genes was key to the domestication of evolutionarily recent TEeRS towards the genesis of transcription networks active in the human brain. They finally imply that inter-individual 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.

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 Genetics*. **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. *PNAS*. **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. 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).

12. 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).

13. 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*, btw495 (2016).

14. B. M. Campbell, E. Charych, A. W. Lee, T. Möller, Kynurenines in CNS disease: regulation by inflammatory cytokines. *Front Neurosci*. **8** (2014), doi:10.3389/fnins.2014.00012.

15. 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 Genetics*. **4**, e1000252 (2008).

16. 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 Reports*. **17**, 570–582 (2016).

17. 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. *Molecular Systems Biology*. **10** (2014), doi:10.15252/msb.20145508.

18. 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. *Nature Methods*. **12**, 1143–1149 (2015).

19. 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).

20. 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. *Science Translational Medicine*. **7**, 307ra153-307ra153 (2015).

21. M. Garcia-Montojo, T. Doucet-O'Hare, L. Henderson, A. Nath, Human endogenous retrovirus-K (HML-2): a comprehensive review. *Critical Reviews in Microbiology*. **44**, 715–738 (2018).

22. 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).

23. 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).

24. M. Denne, M. Sauter, V. Armbruester, 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. *Journal of Virology*. **81**, 5607–5616 (2007).

25. 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).

26. 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).

27. C. H. 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 reports*. **19**, e45000 (2018).

28. 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. Buhu, 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).

29. D. Roulois, H. Loo Yau, R. Singhania, 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).

30. 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).

31. 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).

32. S. M. Haag, M. F. Gulen, L. Reymond, 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).

33. D. Kim, B. Langmead, S. L. Salzberg, HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*. **12**, 357–360 (2015).

34. S. Anders, P. T. Pyl, W. Huber, HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*. **31**, 166–169 (2015).

35. 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).

36. 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. Yang, J. Zhang, Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology*. **5**, R80 (2004).

37. C. W. Law, Y. Chen, W. Shi, G. K. Smyth, voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biology*. **15**, R29 (2014).

38. 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).

39. S. Kumar, S. Subramanian, Mutation rates in mammalian genomes. *PNAS*. **99**, 803–808 (2002).

40. 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).

41. J. Feng, T. Liu, B. Qin, Y. Zhang, X. S. Liu, Identifying ChIP-seq enrichment using MACS. *Nat Protoc*. **7**, 1728–1740 (2012).

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

43. 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).

44. 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).

45. 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. Deflaux, 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).

46. K. J. Karczewski, L. C. Francioli, G. Tiao, B. B. Cummings, J. Alföldi, 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, T. G. A. D. 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 (2019).

47. 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 Evarist Planet for the bioinformatics guidance, the EPFL genomics, Flow Cytometry and Histology Core Facilities for help with sequencing, immunohistochemistry and sorting, respectively. We thank Andrea Ablasser for the STING inhibitor, Raquel Fueyo and Joanna Wysocka for the NCCIT cell line and useful tips for the ENV detection, Riccardo Dainese for the help with the SMILE-seq experiment and the Tronolab for helpful discussions. **Funding:** This study was supported by grants from the Personalized Health and Related Technologies (PHRT-508), the European Research Council (KRABnKAP, No. 268721; Transpos-X, No. 694658) and the Swiss National Science Foundation (310030_152879 and 310030B_173337) to D.T; **Author contributions:** T.P. and D.T. conceived the study, interpreted the data and wrote the manuscript. T.P. designed, performed and analyzed experiments. D.G. performed most of the bioinformatics analyses. C.R. performed most wet experiments; organoids-related experiments were performed by C.P. A.C., C.T. and J.D. developed bioinformatics tools. J.P. and P.E. conducted experiments. B.D. and V.B. gave intellectual input. All authors reviewed the manuscript; **Competing interests:** Authors declare no competing interests; and **Data and materials**

availability: Sequencing data and processed files have been submitted to the NCBI Gene Expression Omnibus (GEO).

Supplementary Materials:

Materials and Methods

Figures S1-S4

Tables S1-S4

References (*31-47*)

Figures:

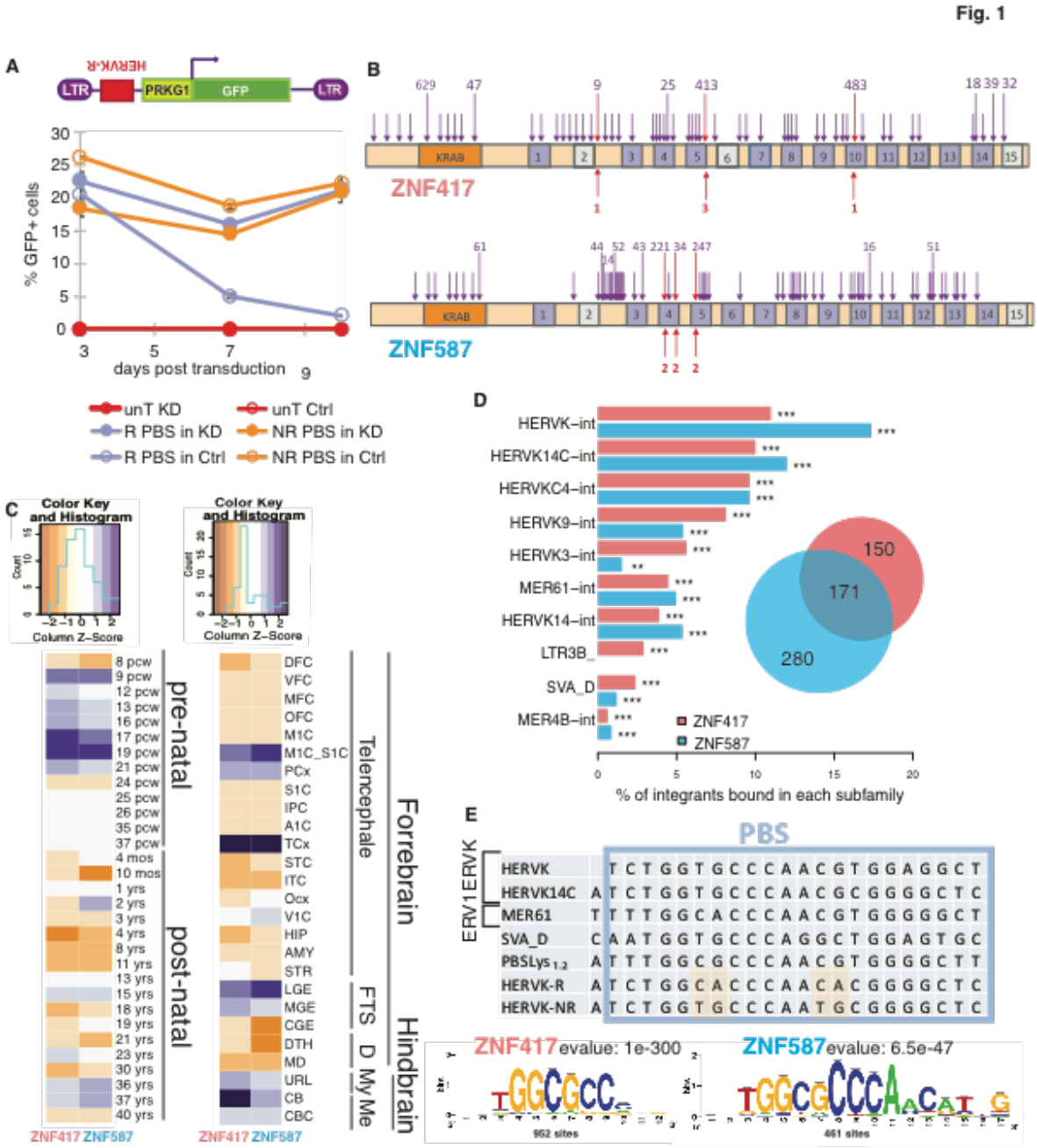

Fig. 1: Genomic characterization of ZNF417/587.

(A) Expression from a PGK-GFP cassette cloned downstream of KAP1-restricted (R) or -non restricted (NR) HERVK PBS sequences in control (Ctrl) or ZNF417/587 KD hESCs (average and s.d. values of duplicates). UnT: untransduced. **(B)** LoF mutations for *ZNF417/587* with numbers of most frequent alleles amongst >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 week. FTS: forebrain fetal transient structures, D: diencephale, My: myelencephale, Me: metencephale, 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)** ZNF417/587-bound TEs in hESCs, with histogram representing percentage of integrants from each TE subfamily (*P* values, hypergeometric test) and Venn diagram total number of ChIP-Seq peaks. **(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.

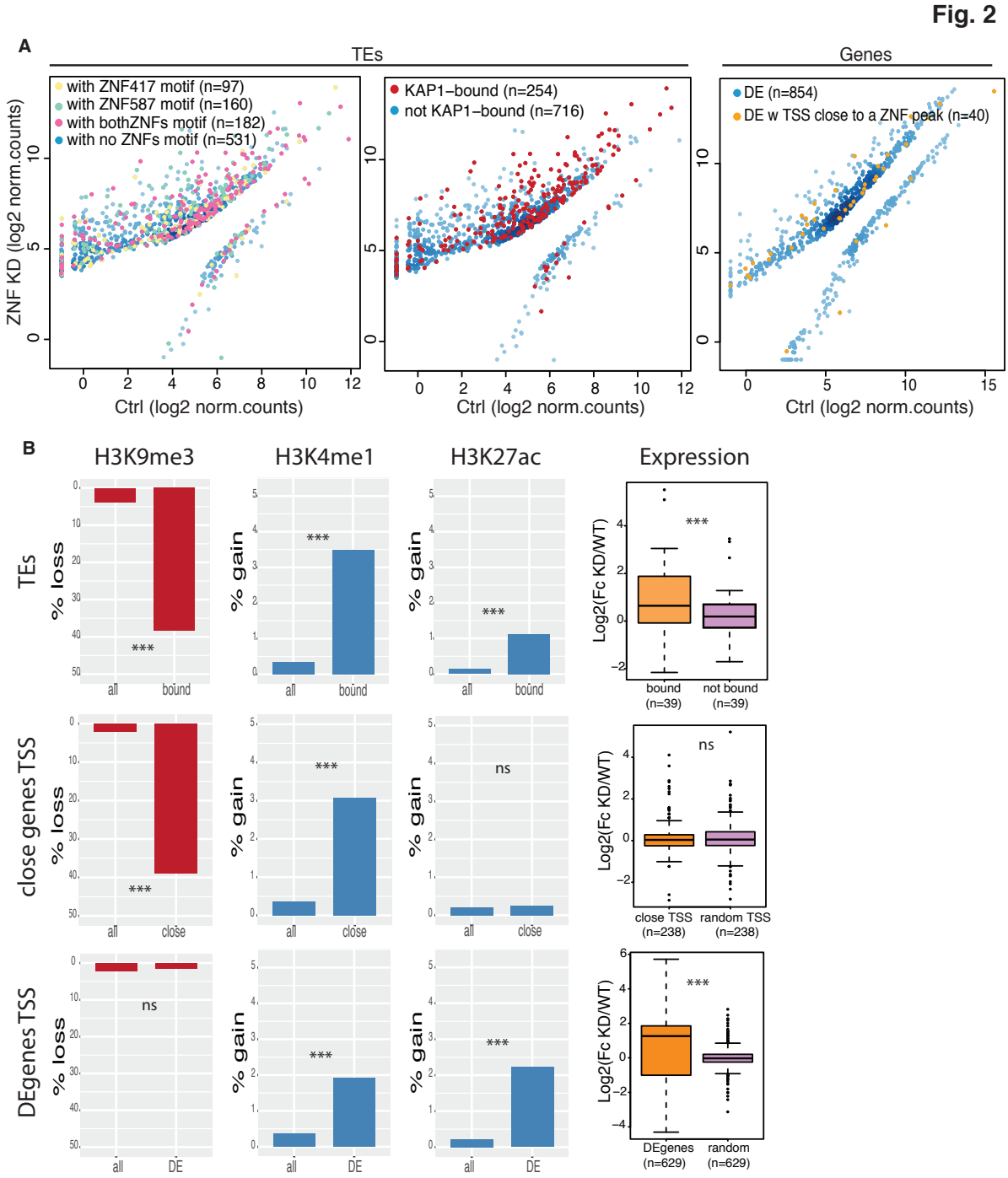

Fig. 2: Impact of ZNF417/587 depletion in pluripotent stem cells.

(A) Dot-plots of RNAseq from ZNF417/587 KD *vs.* control hESCs, outlining differentially expressed (DE) TEs and genes (fold change>2, FDR<0.05). TEs with predicted ZNF417 (yellow), ZNF587 (green) or both (pink) binding motifs (left panel) or bound by KAP1 in hESC (middle panel) or genes with a TSS closer than 100kb from a ZNF binding site (right panel) are highlighted. **(B)** Bar plots depicting loss of H3K9me3 or gain in H3K4me1 or H3K27ac at indicated loci. Upper panel: ZNF417/587-bound *vs*. -unbound TEs; middle panel: TSS of coding genes close to (<20kb) *vs.* distant from a KZFP peak; lower panel: TSS of coding DE genes *vs.* all genes (*P* values, hypergeometric test). Right: fold change in expression in KD *vs.* WT hESC of loci illustrated on left (*P* values, Wilcoxon test).

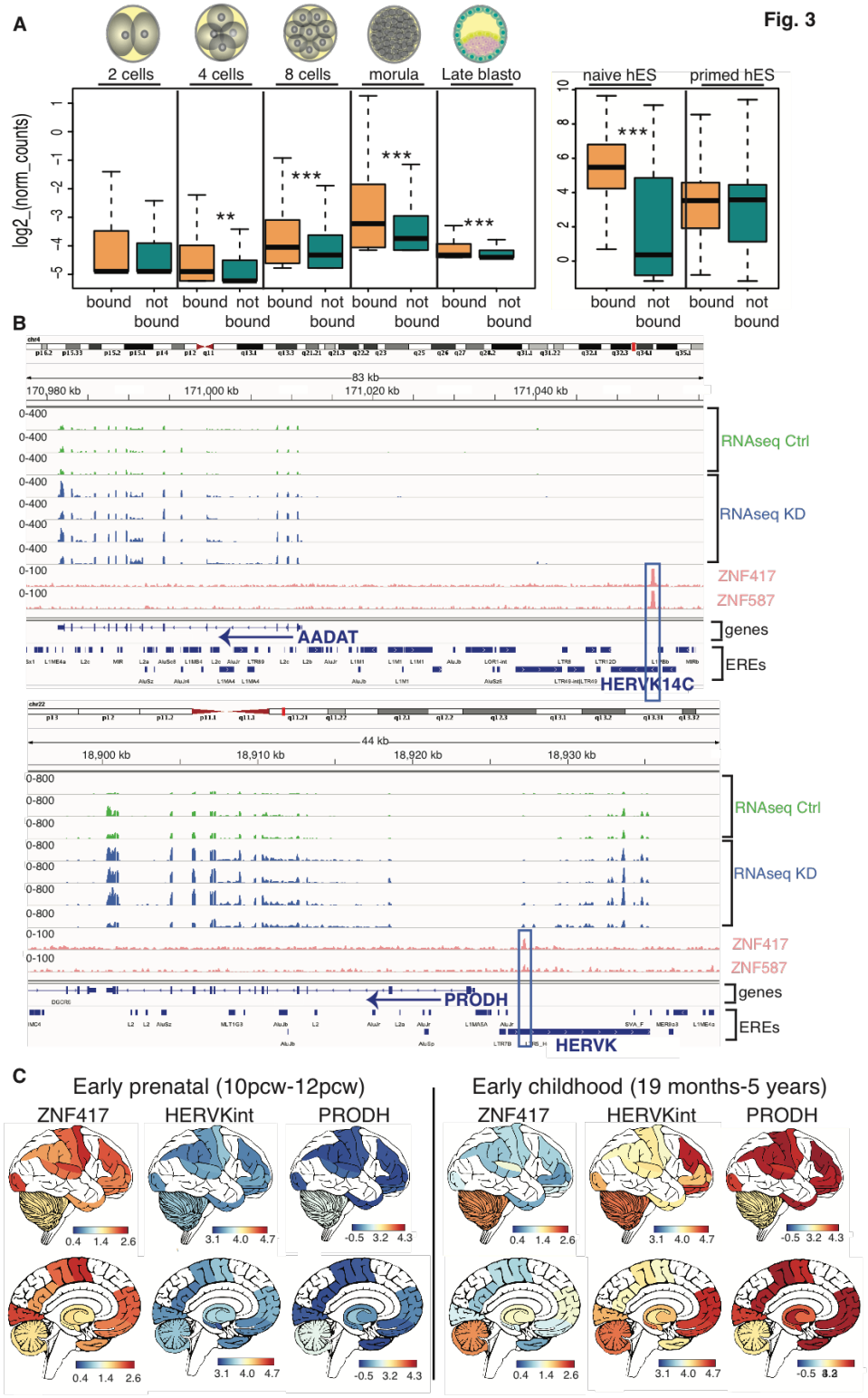

Fig. 3: ZNF417/587-mediated repression of TEeRs and neuron-specific genes.

(A) Expression of TEs bound or not bound at indicated stages of human development using single-cell (left panel) or in naïve versus primed (right panel) hESC RNAseq data. **(B)** IGV screenshots of independent RNAseq replicates from control (Ctrl) and KZFP KD hESCs with boxed ZNF417/587 peaks at HERVK integrants upstream of *AADAT* (upper panel) and *PRODH* (lower panel). **(C)** Spatial representation of *ZNF417*, *HERVK* and *PRODH* expression in early prenatal and childhood brains, using RNAseqs from the Brain Span Atlas of the Developing Human Brain. Pre-natal brain is depicted as anatomically adult for consistency.

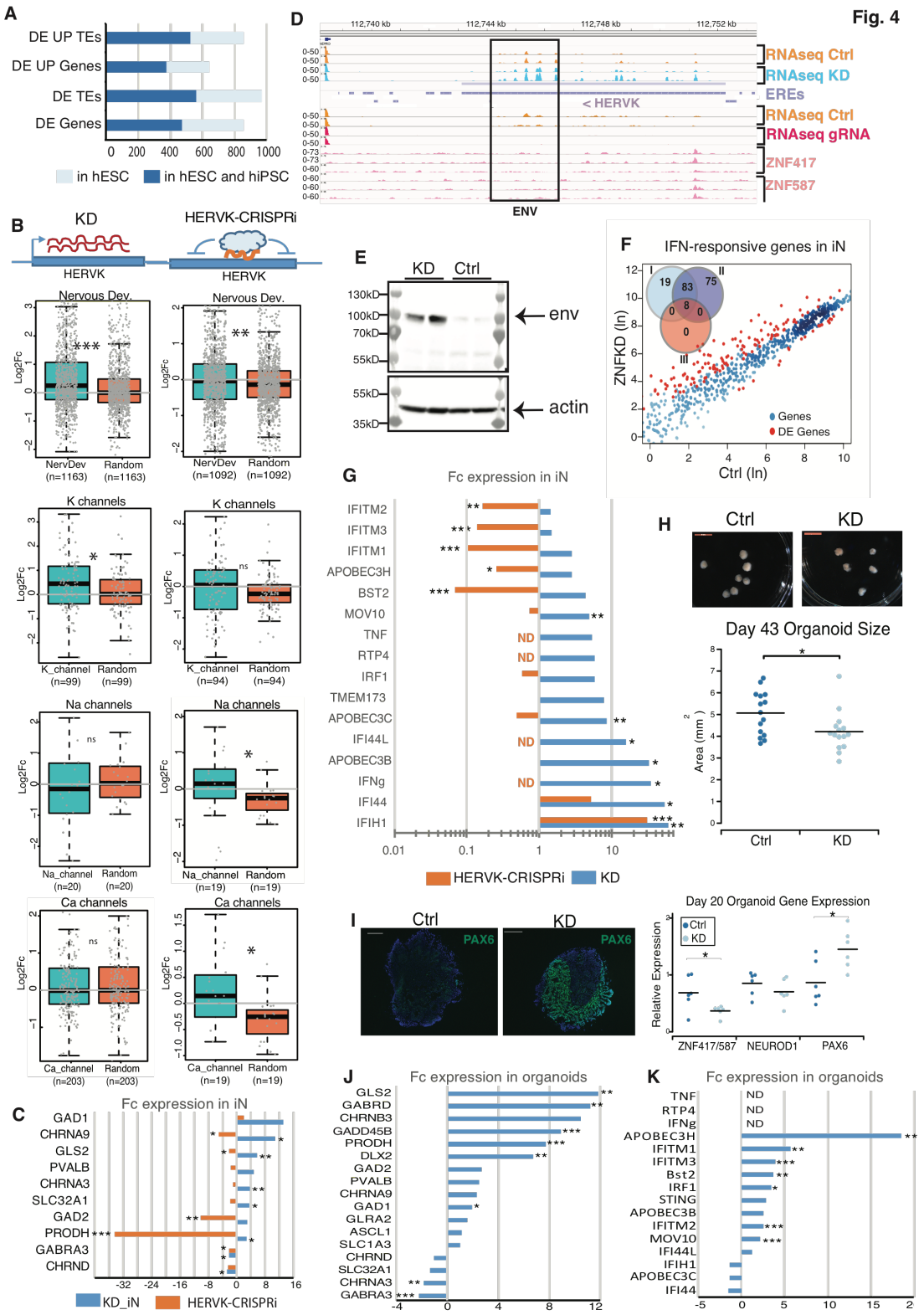

Fig. 4: Impact of ZNF417/587 on neuronal differentiation, function and homeostasis.

(A) Number of DE elements in *ZNF417/587* KD hESCs (light blue) and in common between hESCs+hiPSCs (dark blue). **(B)** Fold change in expression in KZFPs KD (left panels) or

HERVK-silenced (right panels) *vs.* control induced neurons (iN) of genes related to each indicated category (as defined in Allen Brain Atlas GOs) compared to same number of random genes in each case. **(C)** Examples of genes DE in KZFPs KD and HERVK-CRISPRi iN and related to GABA-ergic pathway. **(D)** IGV screenshot of a region encompassing a full length HERVK encoding a consensus envelope with (top) RNAseq duplicates of control *vs.* KZFPs KD and control *vs.* HERVK-CRISPRi iN, and (bottom) ZNF417 and ZNF587 ChIPseq triplicates in H1 hESCs. **(E)** Western blot of independent duplicates of NCCIT-control or -KZFPs depleted cells lysates probed with HERVK anti-ENV antibody. Actin is used as loading control. **(F)** Dot-plot depicting expression of IFN-responsive genes (at least upregulated 5 times by IFN treatment in normal tissues or cells in Interferome database) in KZFPs KD *vs.* control iN. DE genes (Fc>2, FDR< 0.05) are highlighted in red. Venn diagram indicates number of DE genes stimulated by IFNs Type I, II or III. **(G)** Fold change expression of antiviral IFN-responsive genes in KZFPs KD and HERVK-CRISPRi *vs.* control iN (ND: not detected). **(H)** Brain organoids obtained 43 days post-differentiation of control or KZFPs KD hESCs, with size quantification below (Mann Whitney U-Test). **(I)** PAX6 immunostaining and quantification by RT-qPCR of indicated genes in organoids at 20 days post-induction of differentiation of indicated hESCs (two tailed T-test). **(J)** Fold change expression of neuronal function-related genes in KZFPs KD *vs.* control 20-days organoids. **(K)** Fold change expression of IFN-responsive genes in KZFPs KD *vs.* control 20-days organoids. For panels C, G, J, K, the average fold change of independent RNA-seq duplicates is shown (***FDR<0.001, **FDR<0.01, *FDR<0.05).