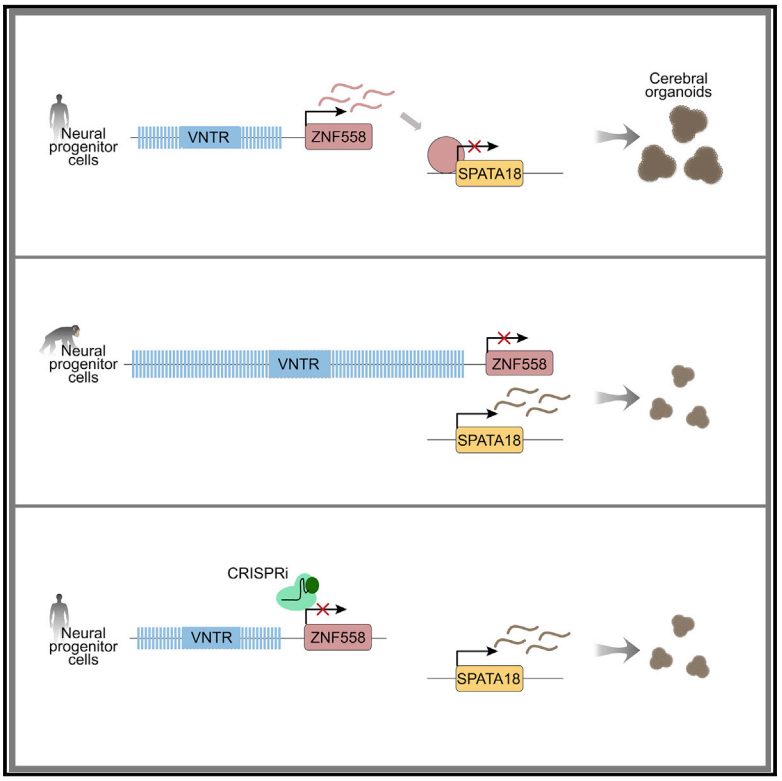
Cell Stem Cell

Article

A *cis*-acting structural variation at the *ZNF558* locus controls a gene regulatory network in human brain development

Graphical abstract



Highlights

- ZNF558 is uniquely expressed in human but not chimpanzee forebrain progenitors
- ZNF558 has been co-opted to control the expression of a single gene, *SPATA18*
- ZNF558 plays a role in mitochondrial homeostasis and brain development
- ZNF558 expression is controlled by the size of a downstream VNTR

Authors

Pia A. Johansson, Per Ludvik Brattås, Christopher H. Douse, ..., Didier Trono, Evan E. Eichler, Johan Jakobsson

Correspondence

johan.jakobsson@med.lu.se

In brief

Johansson et al. identify ZNF558, a KRAB-ZFP expressed in human but not chimpanzee forebrain progenitors, where it regulates its target *SPATA18.* The expression of ZNF558 is controlled by the size of a VNTR that is longer in chimpanzees compared to humans, demonstrating a role for structural variations in human brain evolution.

Johansson et al., 2022, Cell Stem Cell *29*, 1–18
January 6, 2022
https://doi.org/10.1016/j.stem.2021.09.008

CellPress

Please cite this article in press as: Johansson et al., A *cis*-acting structural variation at the *ZNF558* locus controls a gene regulatory network in human brain development, Cell Stem Cell (2021), https://doi.org/10.1016/j.stem.2021.09.008

Cell Stem Cell

CellPress
OPEN ACCESS

Article

A *cis*-acting structural variation at the *ZNF558* locus controls a gene regulatory network in human brain development

Pia A. Johansson,[1,7] Per Ludvik Brattås,[1,7] Christopher H. Douse,[1] PingHsun Hsieh,[2] Anita Adami,[1] Julien Pontis,[3] Daniela Grassi,[1] Raquel Garza,[1] Edoardo Sozzi,[4] Rodrigo Cataldo,[5] Marie E. Jönsson,[1] Diahann A.M. Atacho,[1] Karolina Pircs,[1] Feride Eren,[1] Yogita Sharma,[1] Jenny Johansson,[1] Alessandro Fiorenzano,[4] Malin Parmar,[4] Malin Fex,[5] Didier Trono,[3] Evan E. Eichler,[2,6] and Johan Jakobsson[1,8,*]

[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 Genome Sciences, University of Washington School of Medicine, Seattle, WA 98195, USA
[3]School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland
[4]Developmental and Regenerative Neurobiology, Department of Experimental Medical Science, Wallenberg Neuroscience Center and Lund Stem Cell Center, Lund University, Lund, Sweden
[5]Lund University Diabetes Centre, Department of Clinical Sciences, Lund University, Lund, Sweden
[6]Howard Hughes Medical Institute, University of Washington, Seattle, WA, USA
[7]These authors contributed equally
[8]Lead contact
*Correspondence: johan.jakobsson@med.lu.se
https://doi.org/10.1016/j.stem.2021.09.008

SUMMARY

The human forebrain has expanded in size and complexity compared to chimpanzees despite limited changes in protein-coding genes, suggesting that gene expression regulation is an important driver of brain evolution. Here, we identify a KRAB-ZFP transcription factor, ZNF558, that is expressed in human but not chimpanzee forebrain neural progenitor cells. ZNF558 evolved as a suppressor of LINE-1 transposons but has been co-opted to regulate a single target, the mitophagy gene *SPATA18*. ZNF558 plays a role in mitochondrial homeostasis, and loss-of-function experiments in cerebral organoids suggests that ZNF558 influences developmental timing during early human brain development. Expression of ZNF558 is controlled by the size of a variable number tandem repeat that is longer in chimpanzees compared to humans, and variable in the human population. Thus, this work provides mechanistic insight into how a *cis*-acting structural variation establishes a regulatory network that affects human brain evolution.

INTRODUCTION

The human forebrain has increased in size and complexity after the split between the human and chimpanzee lineages, giving rise to a new level of cognitive functions during hominid evolution (Hill and Walsh, 2005; Lui et al., 2011; Rakic, 2009; Sousa et al., 2017). Cellular and anatomical adaptations have been driven by genetic changes in the human lineage (Enard, 2016), but the actual genetic modifications responsible for this evolutionary process are mostly not understood. Protein-coding genes are highly conserved between human and chimpanzees (Kronenberg et al., 2018) and aside from the well-studied transcription factor FOXP2 (Lai et al., 2001), there remains limited evidence for a wider impact of amino acid substitutions on human brain evolution. Recently, larger structural variations resulting in gene duplication were implicated in human forebrain function and evolution. *NOTCH2NL*, a human-specific paralog of *NOTCH2*, contributes to cortical development (Fiddes et al., 2018; Suzuki et al., 2018), and duplication of *TBC1D3* and the mitochondrial protein *ARHGAP11B* affected cortical expansion via the basal progenitor populations (Dennis and Eichler, 2016; Florio et al., 2015; Ju et al., 2016; Namba et al., 2020). Changes in *cis*-regulatory regions have also long been thought to contribute to species-specific differences (King and Wilson, 1975), and several studies have revealed divergent gene expression patterns in developing primate brains, although their evolutionary impact is unclear (Johnson et al., 2009; Khaitovich et al., 2006; Mora-Bermúdez et al., 2016; Prescott et al., 2015).

One gene family of particular interest in human brain evolution is the Krüppel-associated box (KRAB) domain-containing zinc finger proteins (KZFPs), the largest individual family of transcription factors in mammalian genomes. KZFPs have undergone a rapid expansion during mammalian and primate evolution, and the human genome encodes for at least 350 KZFPs, ~170 of which are primate specific (Imbeault et al., 2017; Jacobs et al., 2014). Many KZFPs are expressed in the human brain and

Please cite this article in press as: Johansson et al., A *cis*-acting structural variation at the *ZNF558* locus controls a gene regulatory network in human brain development, Cell Stem Cell (2021), https://doi.org/10.1016/j.stem.2021.09.008

integrated into neuronal gene regulatory networks (Farmiloe et al., 2020; Imbeault et al., 2017). Notably, these expression patterns are different between human and chimpanzee (Nowick et al., 2009). The majority of KZFPs are thought to be transcriptional repressors. Their conserved N-terminal KRAB domain interacts with the epigenetic co-repressor TRIM28, which induces heterochromatin formation and transcriptional repression of targets (Ayyanathan et al., 2003; Emerson and Thomas, 2009; Matsui et al., 2010; Rowe et al., 2010; Sripathy et al., 2006). KZFPs differ mainly in the number and sequence of the DNA-binding ZF domains, with the number of ZFs ranging between 2 and 40 in humans (Imbeault et al., 2017). The KZFP family has expanded and diversified through repeated cycles of segmental duplications, giving rise to novel KZFP genes with new targets and biological functions (Nowick et al., 2010).

Several studies have demonstrated an important role for KZFPs in the repression of transposable elements (TEs) in a variety of cell types, including embryonic stem cells (ESCs) and neural progenitor cells (Brattås et al., 2017; Ecco et al., 2016; Fasching et al., 2015; Najafabadi et al., 2015; Pontis et al., 2019; Rowe et al., 2010; Rowe and Trono, 2011; Turelli et al., 2014; Wolf et al., 2015; Zhang et al., 2019). The rapid expansion of KZFPs in mammalian genomes is correlated with the expansion of TEs, where KZFPs are thought to evolve to target new TE insertions and sequences (Jacobs et al., 2014; Thomas and Schneider, 2011). The majority of KZFPs bind to specific families of TEs in human cells (Imbeault et al., 2017; Najafabadi et al., 2015), but it has also been proposed that both KZFPs and TEs have been co-opted by host genomes for the broader regulation of transcriptional networks (Brattås et al., 2017; Ecco et al., 2016; Friedli and Trono, 2015; Imbeault et al., 2017).

These observations make KZFPs promising candidates to mediate evolutionary differences between the human and the chimpanzee brain, but there is a lack of experimental data directly addressing this hypothesis. In this study, we investigate the co-option of KZFPs in transcriptional regulation during human and chimpanzee brain development. To this end, we established an *in vitro* differentiation protocol allowing for quantitative comparisons between human and chimpanzee forebrain neural progenitor cells (fbNPCs). We discovered several KZFP transcription factors that are highly expressed in human but not chimpanzee fbNPCs. One of these, ZNF558, is a conserved gene that originally evolved to control the expression of long interspersed element-1 (LINE-1) elements ~100 million years ago (mya). Our data show that ZNF558 no longer suppresses TEs but has been co-opted in fbNPCs to regulate a single gene, the mitophagy regulator *SPATA18* (Kitamura et al., 2011). CRISPR inhibition (CRISPRi)-mediated silencing of ZNF558 in human cerebral organoids suggests a role for ZNF558 in the control of developmental timing in early human brain development. Mechanistically, we provide evidence that ZNF558 expression is controlled by a downstream variable number tandem repeat (VNTR) that is extended in chimpanzees relative to humans and variable in the human population. Epigenetic manipulation of the human VNTR was sufficient to switch off ZNF558 expression and thereby increase SPATA18 expression. Our data reveal the co-option of a TE-controlling KZFP to regulate a protein-coding gene and how this regulatory network is controlled between species and human individuals by a *cis*-acting structural variation.

RESULTS

Derivation of human and chimpanzee fbNPCs

Comparative transcriptomic and epigenetic analyses of chimpanzee and human brain development have been limited by both the availability of material from these species and tissue heterogeneity. As induced pluripotent stem cells (iPSCs) from chimpanzees and other primates have become available, it is now possible to establish *in vitro* cell culture models (Gallego Romero et al., 2015; Marchetto et al., 2013; Mora-Bermúdez et al., 2016; Wunderlich et al., 2014). To directly compare human and chimpanzee fbNPCs, we optimized a defined, feeder-free, 2-dimensional (2D) differentiation protocol based on dual-SMAD inhibition (Grassi et al., 2019). Chimpanzee and human iPSCs could be maintained *in vitro* under identical conditions (Figure 1A), and upon differentiation, we observed a rapid switch to a fbNPC-like morphology in cells from both species (Figure 1B). After 2 weeks of differentiation, we found that both human and chimpanzee fbNPCs expressed high levels of *FOXG1*, a key forebrain marker, while the pluripotency marker *NANOG* was not expressed (Figures 1B and S1A).

Both human and chimpanzee fbNPCs expressed appropriate neuronal and forebrain markers, while genes related to other brain regions or other tissues were close to undetectable (Figure 1C). To investigate whether human and chimpanzee iPSCs differentiate into fbNPCs with different temporal trajectories, we performed bulk RNA sequencing (RNA-seq) at 13, 14, 15, and 16 days of differentiation and analyzed the covariance of gene expression between these time points and the 2 species. At the selected time points, the fbNPCs corresponded to a differentiation stage just before neuronal commitment, demonstrated by the gradual increase in basal progenitor marker *EOMES* and neuronal markers *DCX* and *TBR1* from days 13 to 16 (Figure 1C). We found no upregulation of glial markers, in line with the stepwise generation of neurons and glia during human brain development (Figure 1C). Globally, a similar set of genes were up- and downregulated between days 13 and 16 in human and chimpanzee fbNPCs, indicating that the temporal dynamics of the protocol were closely matched (Figure 1D). Overall, human and chimpanzee fbNPCs display a very similar transcriptome (Figure S1B). We confirmed a limited batch-to-batch variation in the differentiation protocol, and the results were consistent in cell lines from different individuals (Figure S1C).

To investigate the heterogeneity of human and chimpanzee fbNPC cultures, we performed single-cell RNA-seq analysis at day 14 for 4,355 human and 3,620 chimpanzee cells. Principal-component analysis (PCA) showed that 95% of the cells clustered into 1 major population in both species (Figures S1D–S1F). Transcriptional variation within the major population was mainly explained by differences in cell-cycle state rather than cell identity, as these cells clustered into a dense population inseparable on PC1 and PC2 after regressing out cell-cycle effects (Figure S1E). t-Distributed stochastic neighbor embedding (tSNE) confirmed the presence of a large major population of cells homogeneously expressing the forebrain progenitor markers *FOXG1* and *PAX6* (Figure 1E). The tSNE analysis also

Please cite this article in press as: Johansson et al., A *cis*-acting structural variation at the *ZNF558* locus controls a gene regulatory network in human brain development, Cell Stem Cell (2021), https://doi.org/10.1016/j.stem.2021.09.008

CellPress
OPEN ACCESS

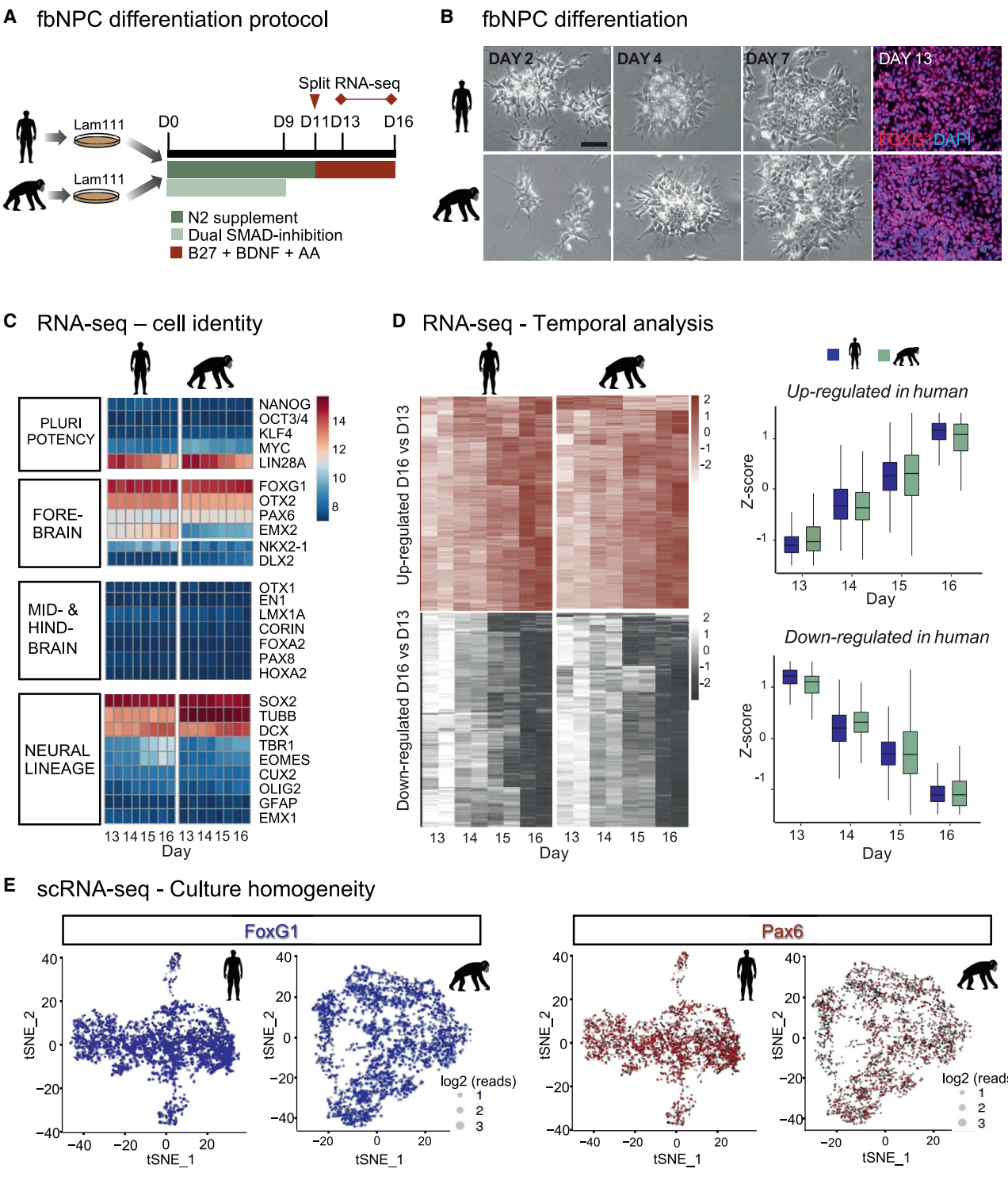


Figure 1. Differentiation of human and chimpanzee iPSCs to fbNPCs

(A) Schematics of the dual-SMAD inhibition-based differentiation protocol from seeding iPSCs at day 0 to harvesting the fbNPCs at days 13–16 of differentiation.
(B) Bright-field images of human and chimpanzee cells during the first week of differentiation, and FOXG1 immunocytochemistry at day 13, scale bar: 100 μm.
(C) Heatmap showing marker expression at days 13–16 of differentiation (n = 1 human, n = 1 chimpanzee, 2 differentiation batches at each time point).
(D) Heatmaps (left) displaying genes that are significantly (p adjusted < 0.01) up- and downregulated over time (day 16/day 13, respectively) in humans, the same set of genes mapped for both species. Boxplots (right) showing the same set of genes. The lower and upper hinges correspond to the first and third quartiles.
(E) tSNE-analysis of single-cell RNA-seq data of the forebrain markers PAX6 and FOXG1 in human and chimpanzee fbNPCs (n = 1 human, n = 1 chimpanzee, 2 different differentiation batches at each time point).
See also Figure S1.

Please cite this article in press as: Johansson et al., A *cis*-acting structural variation at the *ZNF558* locus controls a gene regulatory network in human brain development, Cell Stem Cell (2021), https://doi.org/10.1016/j.stem.2021.09.008

revealed 2 minor populations (<5%), 1 of which expressed markers associated with early-committed neurons such as *NEUROG1* and *NEUROD1*, while the other related to the endothelial lineage (e.g., *ANKRD1*, *CTGF*) (Figure S1F). Taken together, immunocytochemistry, bulk, and single-cell RNA-seq demonstrated that our 2D differentiation protocol was reproducible and gave rise to temporally and phenotypically matched homogeneous cultures of human and chimpanzee fbNPCs, making it a suitable model system for direct comparative analysis.

Human-specific expression of KZFP transcription factors

Next, we queried our transcriptomic datasets for differentially expressed genes, choosing to focus on KZFP transcription factors because their evolutionary and biochemical characteristics make them prime candidates for governing species-specific differences. We identified 312 KZFPs that were expressed in at least 1 sample and 35 that were significantly different between species (27 higher in human, 8 higher in chimp; Wald's test, $p < 1.0 \times 10^{-15}$, with Benjamini-Hochberg correction and $\log_2$ fold change [fc] > 1) (Figure 2A). Seven candidates (*ZNF138*, *ZNF248*, *ZNF439*, *ZNF557*, *ZNF558*, *ZNF596*, and *ZNF626*) were highly expressed in human fbNPCs, with nearly no expression in chimpanzee fbNPCs (Figures 2B and S2A).

The massive expansion of the KZFP family in mammals has arisen through rounds of segmental duplication events, making them challenging to study with comparative genomic analyses. To confirm that human-specific expression was not caused by biases in mapping, reference genome builds, or gene annotation, we mapped all of the samples to both human (GRCh38) and chimpanzee (PanTro6) assemblies. *ZNF138* expression was only detected when mapping human samples to GRCh38, but not when mapping human samples to PanTro6, nor in chimp samples mapping to either assembly. Pairwise alignment of human and chimpanzee *ZNF138* coding sequences revealed only 34% sequence identity, with point mutations and several deletions in the human sequence, explaining why reads from human samples do not map to the chimpanzee genome. We conclude that *ZNF138* has diverged in both sequence and expression pattern in forebrain progenitors. *ZNF557* and *ZNF626* were detected in chimpanzee samples only when mapping to GRCh38. This suggested that issues with the PanTro6 assembly prevented mapping of these transcripts, invalidating attempts to infer human-specific expression of these genes. Four candidate genes (*ZNF248*, *ZNF439*, *ZNF558*, and *ZNF596*) were exclusively expressed in human samples, despite high mapability for both human and chimpanzee orthologs (Figure 2B).

To further dissect the expression divergence for the 5 candidate KZFPs (*ZNF138*, *ZNF248*, *ZNF439*, *ZNF558*, and *ZNF596*) with human-specific expression in the developing forebrain, we used our single-cell RNA-seq analysis of human and chimpanzee fbNPCs. This analysis confirmed that all 5 candidates were highly expressed in human samples (Figure 2C). *ZNF138*, *ZNF248*, and *ZNF558* were exclusively expressed in human cells, while *ZNF439* and *ZNF596* were also detected in a rare number of chimpanzee cells (Figures 2C and S2B). The candidates were expressed throughout the cell population, excluding the possibility that the human-specific expression was due to a specific subpopulation in the human samples. We also analyzed publicly available expression patterns in human and chimpanzee cerebral organoids, which confirmed that *ZNF138*, *ZNF248*, and *ZNF558* also display human-specific expression in 3D models of neural development (Field et al., 2019; Kanton et al., 2019; Mora-Bermúdez et al., 2016) (Figures 2D, S2C, and S2D). CUT&RUN epigenomic profiling in human and chimpanzee fbNPCs demonstrated a striking human-specific enrichment of the activating epigenetic mark H3K4me3 over the promoters of the candidate genes, confirming that the observed differences in RNA levels result from differences in transcriptional activity (Figure 2F).

qRT-PCR analysis of iPSCs demonstrated that *ZNF138*, *ZNF248*, and *ZNF558* were expressed at low levels in human iPSCs and that their expression increased during the course of fbNPC differentiation (Figure S2E). RNA-seq data from human fetal forebrain samples confirmed their expression during human forebrain development (Figure 2E). We further verified that our results were not limited to the 2 chimpanzee individuals used to derive the iPSCs used in this study; a dataset with 8 other individuals confirmed human-specific expression of ZNF558, the candidate with the most robust expression in iPSCs (Figure S2F).

We next asked whether the differences in *ZNF138*, *ZNF248*, and *ZNF558* expression were limited to brain development or also found in other tissues. Publicly available transcriptome data (GTEx) indicate that *ZNF138*, *ZNF248*, and *ZNF558* are widely expressed in human tissues (Figure S3A). However, we found no apparent difference in expression between humans and chimpanzees in the adult brain, heart, or fibroblasts (Figure S3B). Thus, the difference in the expression of *ZNF138*, *ZNF248*, and *ZNF558* in humans and chimpanzees appears to be limited to specific organs and developmental stages, including brain development.

Our analysis identified 3 KZFPs expressed during human but not chimpanzee brain development (*ZNF138*, *ZNF248*, and *ZNF558*). We selected *ZNF558* for further analysis as it was exclusively and highly expressed in human samples and because this difference was detected among multiple individuals.

ZNF558 has evolved under constraint in the mammalian lineage

Many KZFPs have undergone rapid evolution in primate evolution, in which new genes arise and diversify through segmental duplications (Nowick et al., 2010). However, *ZNF558* is estimated to have originated before the common ancestor of afrotherian mammals, ~105 mya, with orthologs in a range of placental mammals (Figure 3A). A duplication event occurred in the common ancestor with new world monkeys, giving rise to a paralog (*ZNF557*), which was expressed in both human and chimpanzee fbNPCs. *ZNF557* encodes 1 more ZF domain (10 in total) than *ZNF558*, and 8 of the 9 common ZF domains have ≥1non-synonymous mutations in DNA binding residues relative to ZNF558, suggesting that positive selection acted to diversify the function of these paralogs, resulting in divergent DNA targets.

Pairwise alignments revealed high conservation among ZNF558 orthologs (Figure 3B). Analysis of the ratio between non-synonymous and synonymous mutations (dN:dS) suggests that ZNF558 has evolved under constraint in the mammalian lineage, with a median dN:dS ratio of 0.159 (Figure 3C). KZFP

Please cite this article in press as: Johansson et al., A *cis*-acting structural variation at the *ZNF558* locus controls a gene regulatory network in human brain development, Cell Stem Cell (2021), https://doi.org/10.1016/j.stem.2021.09.008

CellPress
OPEN ACCESS

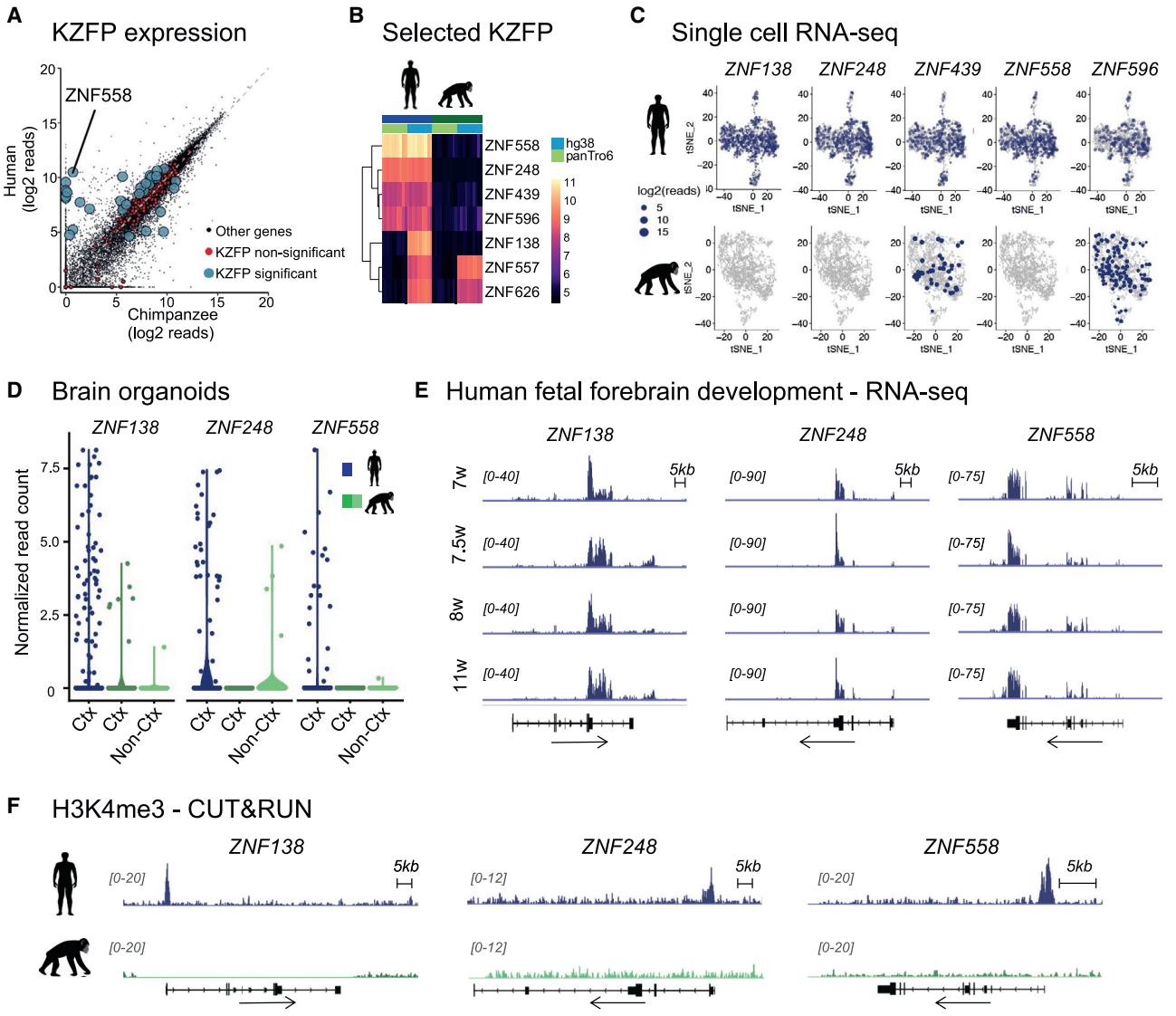


Figure 2. Identification of KZFPs with human-specific expression

(A) Expression levels of genes shown as normalized read counts in human (y axis) and chimp (x axis) in fbNPCs. ($p < 1e{-}15$; Benjamini-Hochberg corrected, and log2 fold change [fc] >1).

(B) Heatmap of 7 KZFPs with high expression in human and no expression in chimp, mapping of all samples to both the human and chimp reference genome.

(C) tSNE plots of KZFP expression in single-cell sequencing data of human and chimp fbNPCs.

(D) KZFP expression in single-cell sequencing data of human (blue) and chimp (green) cortical organoid data (Mora-Bermúdez et al., 2016). Ctx, cortical-like cells; non-Ctx, non-cortical cells, based on marker expression and clustering.

(E) RNA-seq analysis of KZFP expression in dissected forebrain tissue from aborted human embryos of different gestational age (w, weeks). Scale is reads per kilobase per million mapped reads (RPKM).

(F) Normalized CUT&RUN tracks illustrating H3K4me3 levels over KZFP genes in human (top, blue) and chimp (bottom, green) fbNPCs. Data were obtained in biological duplicate with similar results. CUT&RUN data were normalized based on spike-in DNA.

See also Figure S2.

DNA binding is mediated by tandem C2H2 ZF domains, each of which interacts with contiguous stretches of 2–4 nt per finger (Patel et al., 2018; Persikov and Singh, 2014). In particular, the identities of 4 residues in each ZF are responsible for sequence specificity, with multiple ZF domains combining to encode a particular "zinc fingerprint" (Imbeault et al., 2017). To test whether the DNA-contacting residues of its 9 ZFs have evolved during recent mammalian evolution, we aligned sequences of the ZNF558 fingerprint among 8 mammals and found that these residues are almost completely conserved. All primates display identical binding residues, while single substitutions in mouse, whale, and panda proteins are conservative (Ile to Val; Ser to Thr) and thus are unlikely to have a major effect on DNA binding (Figure 3D).

Please cite this article in press as: Johansson et al., A *cis*-acting structural variation at the *ZNF558* locus controls a gene regulatory network in human brain development, Cell Stem Cell (2021), https://doi.org/10.1016/j.stem.2021.09.008

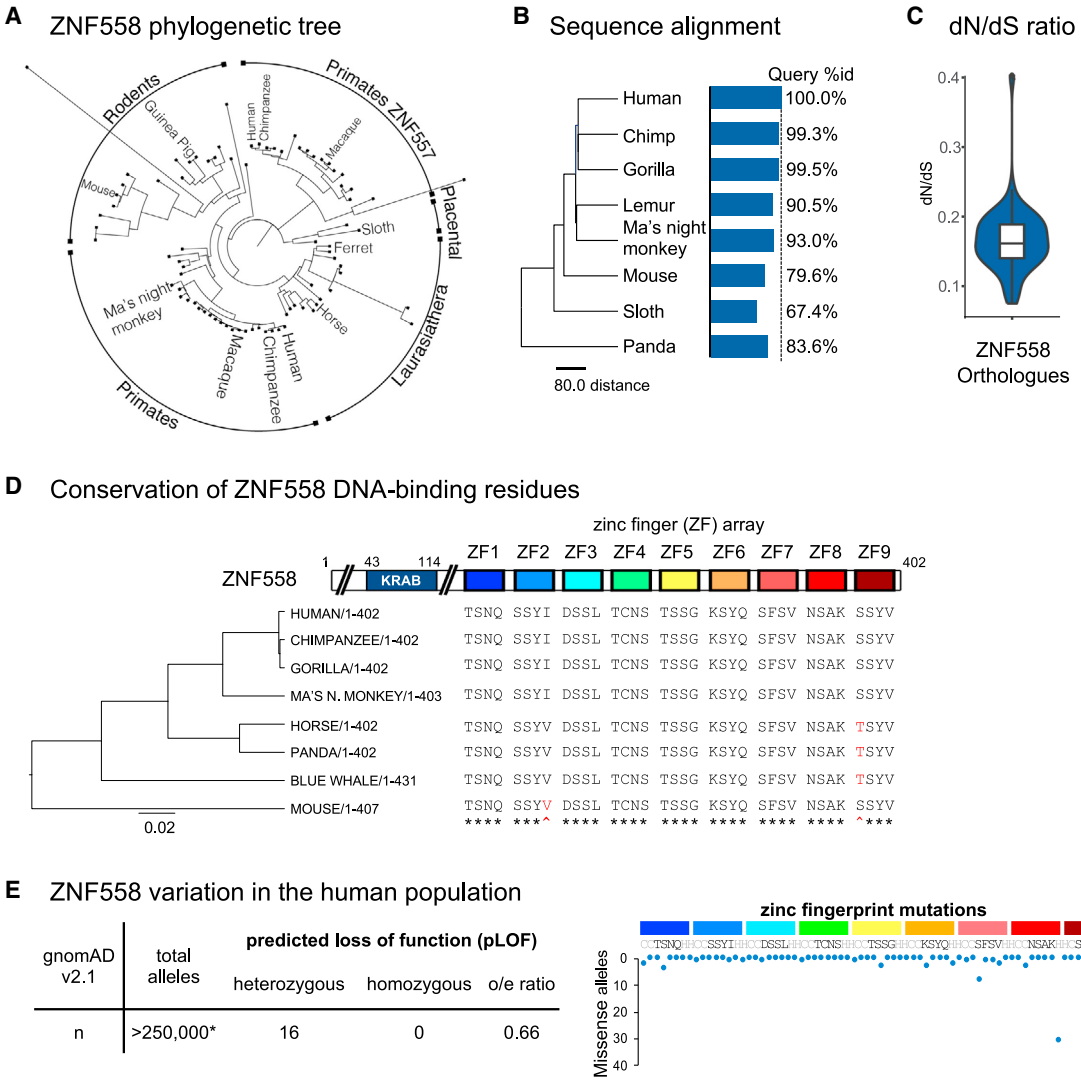


Figure 3. Evolutionary history of ZNF558

(A) Phylogenetic tree of ZNF558 orthologs and paralogs (ZNF557).

(B) Percentage of orthologous protein sequence matching human ZNF558 protein sequence in selected species.

(C) dN:dS ratio of all ZNF558 orthologs identified in Ensembl.

(D) Domain structure of ZNF558 (top). Multi-sequence alignment of DNA-contacting residues in the zinc finger (ZF) array of selected ZNF558 orthologs (bottom). The 4 amino acids shown per ZF are defined as the −1, −4, −5, and −7 positions relative to the first histidine of each C2H2 ZF.

(E) Genetic variation of *ZNF558* in the human population. Shown are the number of high-quality genotypes annotated as predicted loss of function (pLOF, left) or causing missense substitutions in zinc- (gray) or DNA-binding (black) resides from ZF domains (right). o/e, observed:expected ratio of pLOF variants. Data from gnomAD version 2.1 (Karczewski et al., 2020).

See also Figure S3.

We next investigated the presence of *ZNF558* gene variants in the human population using a database (gnomAD version 2) of 125,748 exome and 15,708 whole-genome sequences from unrelated individuals (Karczewski et al., 2020). Heterozygous predicted loss of function (pLoF) alleles were found at slightly below the expected frequency (depth-corrected observed:expected [oe] ratio = 0.66; 90% confidence interval [CI] 0.48–1.12), and no individuals harboring homozygous pLoF variants were identified (Figure 3E). Analysis of missense variants showed that the KRAB domain is strictly conserved in putative TRIM28-interacting residues (Friedman et al., 1996; Murphy et al., 2016). We also found low variant counts in the zinc fingerprints and zinc-coordinating residues among the human population, with 1 exception, a H370P substitution in ZF8 found in ~1 in 1,000 Finnish individuals (Figure 3E). Tthis evolutionary and population-wide analysis of ZNF558 conservation and variation indicates that the protein has been under stringent evolutionary constraint for ~100 million years, which is in line with an important role in mammalian physiology.

Please cite this article in press as: Johansson et al., A *cis*-acting structural variation at the *ZNF558* locus controls a gene regulatory network in human brain development, Cell Stem Cell (2021), https://doi.org/10.1016/j.stem.2021.09.008

CellPress
OPEN ACCESS

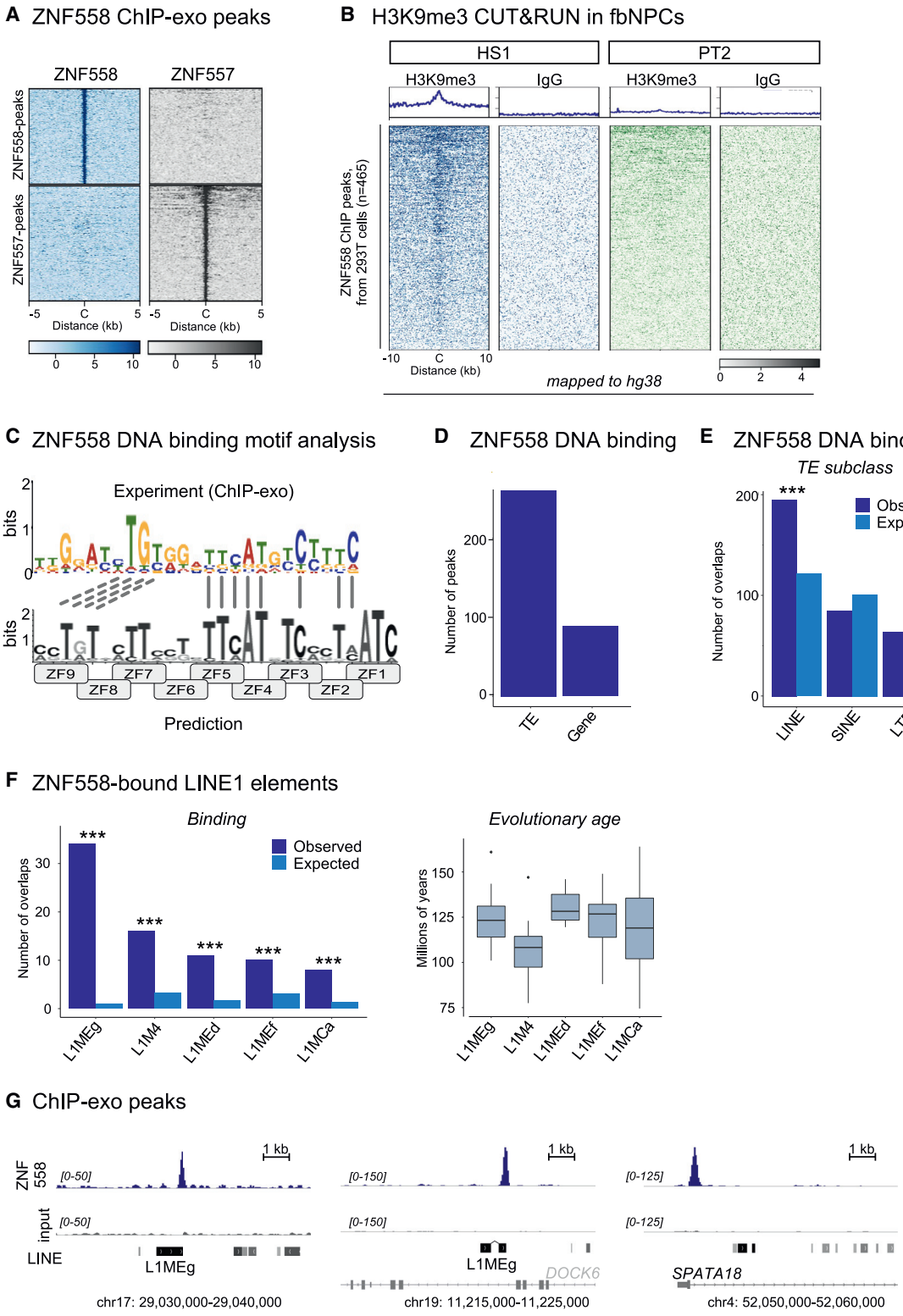


Figure 4. ZNF558 genome binding analysis

(A) Heatmaps of ZNF558 and ZNF557 ChIP-exo signal plotted over high-confidence ZNF558 (top) and ZNF557 (bottom) peaks.

(B) Heatmaps of H3K9me3 CUT&RUN signal in human (HS1) and chimpanzee (PT2) fbNPC over ZNF558 binding sites shown in (A) (n = 465). Above each heatmap, the mean signal is plotted.

(C) ZNF558 binding motif analysis based on experimental ChIP-exo data and computational prediction.

(legend continued on next page)

Please cite this article in press as: Johansson et al., A *cis*-acting structural variation at the *ZNF558* locus controls a gene regulatory network in human brain development, Cell Stem Cell (2021), https://doi.org/10.1016/j.stem.2021.09.008

ZNF558 binds evolutionarily old LINE-1 elements and protein-coding genes

The expansion of KZFPs in mammalian genomes is thought to be driven by adaptation to rapidly evolving TEs (Jacobs et al., 2014; Thomas and Schneider, 2011). It was therefore striking that our candidate gene for human/chimpanzee divergence, *ZNF558*, is highly conserved. This suggests that it may have been co-opted for regulating non-TE targets early in mammalian evolution. To investigate the function of ZNF558, we considered the DNA binding preferences of its 9-membered ZF array. Chromatin immunoprecipitation with exonuclease digestion (ChIP-exo) data on hemagglutinin (HA)-tagged ZNF558-expressing 293T cells showed robust signal enrichment and no overlap in target specificity between ZNF558 and its paralog ZNF557 (Imbeault et al., 2017) (Figure 4A). CUT&RUN profiling in human and chimpanzee fbNPCs demonstrated a human-specific enrichment of the repressive epigenetic mark H3K9me3, which is associated with KZFP binding, over the ZNF558 binding sites, suggesting that the sites identified in 293T cells are also occupied in human fbNPCs (Figures 4B and S4A). Computational prediction of the ZNF558 binding motif with overlapping 4-nt subsites suggested a T-rich binding sequence of 28 nt. The experimental binding motif partly matched the prediction, with some gaps (Figure 4C). Longer arrays are known to deviate more from the canonical binding model and A/T-rich sites are less predictable, so such discrepancies were not surprising (Patel et al., 2018).

We next investigated the features of genomic regions bound by ZNF558. We found that 54% of the 465 high-confidence ZNF558 binding sites were located in TEs, which were enriched for LINE-1 elements (Figures 4D–4F). Of these ZNF558-bound LINE-1 elements, the majority of the sites were in evolutionary-old LINE-1 families such as L1MEg, L1M4, and L1Med (Figure 4F). These LINE-1 families are remnants of old transposition events and have since degenerated and lost transposition capacity. Estimation of the evolutionary age of the ZNF558-bound elements revealed that the majority were dated at ~100 million years old, which correlates well with the age of *ZNF558* (Figure 4F). This observation is in line with a model in which ZNF558 originally evolved to repress then-active LINE-1 elements. These LINE-1s would then have degenerated and ZNF558 could have been co-opted to control other non-TE genomic targets. In agreement with this model, we noted that several of the ZNF558 target sites overlap with protein-coding genes. For example, the fourth highest-scoring ChIP-exo peak was located just downstream of the first exon of *SPATA18*, a gene involved in mitochondrial homeostasis (Figure 4G). These co-opted ZNF558 targets may be important for mammalian physiology, since ZNF558 has been under stringent positive selection for a long time. If this scenario is true, then the LINE-1 binding that remains today likely represents a genomic fossil.

CRISPRi-mediated transcriptional silencing reveals SPATA18 as the sole functional target of ZNF558

To investigate the functional relevance of ZNF558 in human fbNPCs, we designed a CRISPRi strategy to silence ZNF558 expression (Figure 5A). We targeted 2 distinct guide RNAs (gRNA) to a genomic region located next to the *ZNF558* transcription start site (TSS) and co-expressed these with a transcriptional repressor domain fused to catalytically dead Cas9 (dCas9). The transduction of human iPSCs resulted in the efficient silencing of *ZNF558* upon differentiation to fbNPCs (Figures 5A, 5B, and 5E), but no difference in differentiation capacity or expression of cell fate markers compared to controls (Figures S5A and S5B). Silencing of ZNF558 also resulted in an almost complete loss of H3K9me3 at ZNF558 binding sites (Figure S4B), confirming the requirement for maintained ZNF558 binding to retain local heterochromatin.

Next, we performed RNA-seq on *ZNF558*-silenced fbNPCs and analyzed the transcriptome for alterations in TE and gene expression. Remarkably we found a single protein-coding gene, *SPATA18*, to be upregulated upon ZNF558 silencing (Figures 5C, 5D, and S5D), but no alterations in TE expression either when querying expression of unique elements or entire families. In accordance with this, H3K4me3 CUT&RUN in ZNF558-CRISPRi fbNPCs revealed a total loss of H3K4me3 over the promoter of *ZNF558* and that the sole ZNF558 binding site to acquire this histone mark was the binding site at *SPATA18* (Figures 5E and S4C). These data demonstrate that ZNF558 has been co-opted in human fbNPCs to regulate *SPATA18*, and that its binding sites in old LINE-1 elements do not have detectable consequences on transcription of those TEs.

ZNF558 regulates SPATA18 expression in human and chimpanzee fbNPCs

Next, we considered the species specificity of *SPATA18* repression by ZNF558. Since transcription factors and their DNA target sites co-evolve, we analyzed the conservation of the ZNF558 binding site in a range of mammalian *SPATA18* genes and found that it is conserved only in primates (Figure S4D). The conservation of the binding site in chimpanzee *SPATA18* suggests that the lack of ZNF558 expression in chimpanzee fbNPCs should result in the increased expression of SPATA18 in these cells. In accordance with this model, epigenomic profiles over the *SPATA18* promoter in chimpanzee fbNPCs revealed enrichment of H3K4me3 but not H3K9me3 with concomitant expression of SPATA18 mRNA (Figure 5F). RNA-seq analysis showed a 2.4-fold increase in chimpanzee compared to human fNPCs, and increased *SPATA18* expression in chimpanzee was also observed in cerebral organoids (Field et al., 2019) (Figure S4E). To further test that ZNF558 has the capacity to regulate *SPATA18* in chimpanzee fbNPCS, we transduced chimpanzee iPSCs with a ZNF558 overexpression vector (Figures 5G–5I). Following differentiation to fbNPCs, *SPATA18* was significantly downregulated upon ZNF558 overexpression (Figures 5H, 5I,

(D–F) ZNF558 binding sites in the human genome (D), TEs (E), and LINE-1-subfamilies (F). (F) Evolutionary age of the ZNF558-bound elements on the top 5 most-bound LINE-1 subfamilies.

(G) Example 10-kb screenshots of ZNF558 ChIP-exo peaks. Shown are instances of intergenic and intronic L1MEg (left, center) and gene targets (right). Scales are RPKM.

See also Figure S4.

Please cite this article in press as: Johansson et al., A *cis*-acting structural variation at the *ZNF558* locus controls a gene regulatory network in human brain development, Cell Stem Cell (2021), https://doi.org/10.1016/j.stem.2021.09.008

CellPress
OPEN ACCESS

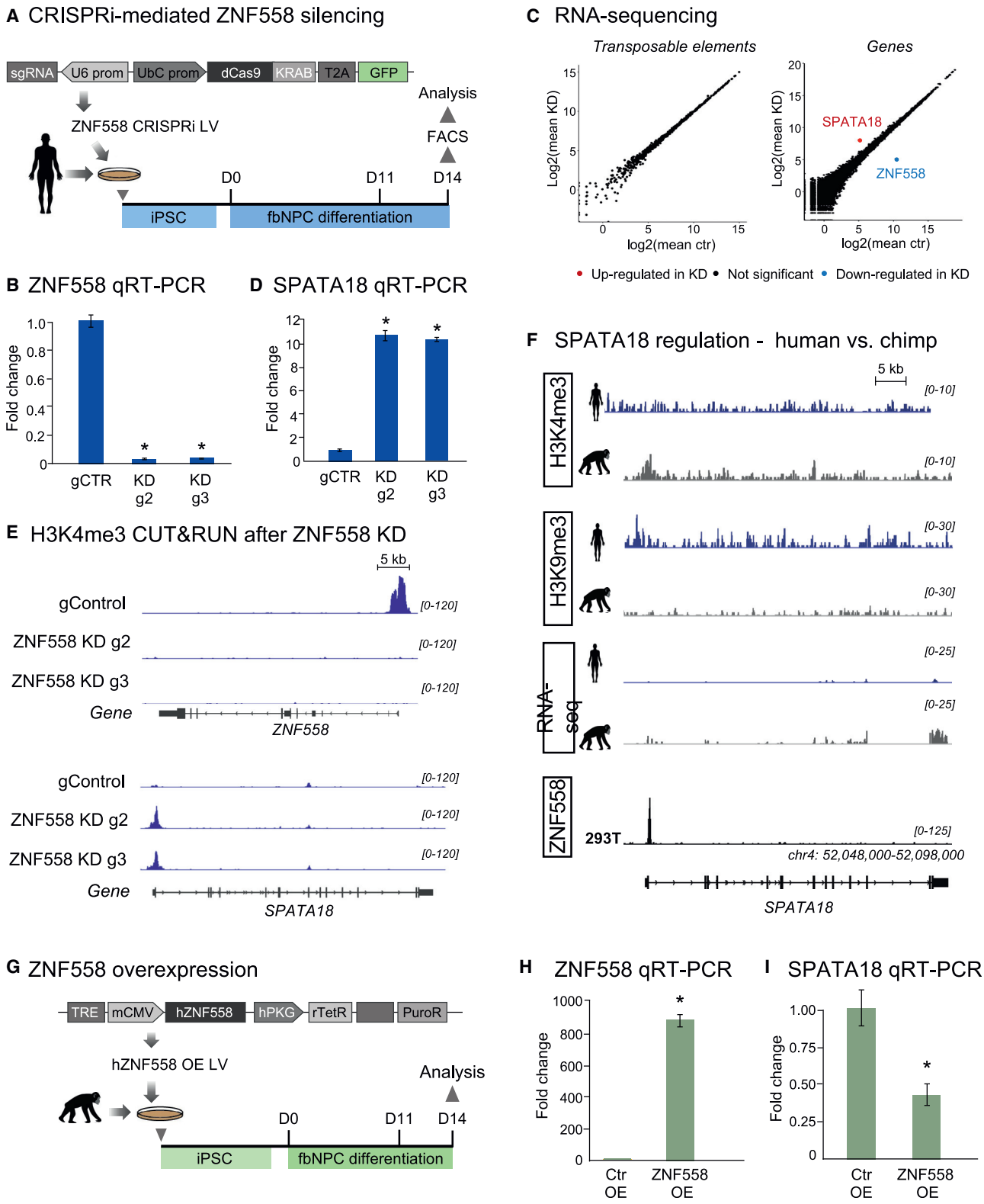


Figure 5. Functional analysis of ZNF558 in fbNPCs

(A) Schematics illustrating the CRISPRi-based strategy to silence ZNF558 in human fbNPCs. iPSCs were transduced with a lentiviral CRISPRi construct and then differentiated for 14 days.

(B) qRT-PCR analysis of ZNF558 expression in fbNPCs after CRISPRi silencing (n = 4)..

(legend continued on next page)

Please cite this article in press as: Johansson et al., A *cis*-acting structural variation at the *ZNF558* locus controls a gene regulatory network in human brain development, Cell Stem Cell (2021), https://doi.org/10.1016/j.stem.2021.09.008

and S5E). These results demonstrate that the functional consequence of differential *ZNF558* expression in human and chimpanzee brain development is the repression of a single gene, *SPATA18*.

ZNF558 contributes to human-specific developmental timing in organoids

The product of the ZNF558 target gene *SPATA18* is mitochondrial eating protein (MIEAP), which contributes to the degradation of mitochondria via mitophagy (Kitamura et al., 2011). In line with this, we found that ZNF558 silencing resulted in a small but significant decrease in mitochondrial content, as measured with qPCR for mitochondrial DNA, in human fbNPCs (Figure 6A). In addition, we found an increase in the levels of ATPase5A in respiratory chain complex V (Figure 6B), suggesting an adaptive change in ATPase levels in response to the decreased number of mitochondria. Interestingly, recent observations directly link mitochondrial function to speciation and the expansion of the human cerebral cortex (Namba et al., 2020). Thus, to investigate the functional role of ZNF558 in human brain development, we generated ZNF558-deficient cerebral organoids (Figure 6C), since this model system allows the study of human-specific developmental process in a 3D setting (Benito-Kwiecinski et al., 2021; Kanton et al., 2019; Pollen et al., 2019).

We found that CRISPRi-mediated ZNF558 silencing using 2 distinct gRNAs did not impair cerebral organoid formation and that the resulting organoids displayed characteristic neural rosettes after 1 month of differentiation, as visualized with Pax6/ZO1 staining, in a similar pattern when compared to organoids generated from control iPSCs (lacZ-gRNA; Figure 6D). However, we noted that ZNF558-CRISPRi organoids appeared smaller than control organoids during the first month of differentiation (Figure 6E), which corresponds to a period when the neuroepithelial buds have formed and are undergoing expansion.

To further assess the long-term consequences of ZNF558-inhibition on human brain development, we analyzed organoids at later stages of maturation (2 and 4 months) using single-nuclei RNA-seq. High-quality data were generated from a total of 22,923 cells, including 15,768 from ZNF558-CRISPRi organoids (2 gRNAs) and 7,155 from control organoids (lacZ-gRNA). We performed an unbiased clustering analysis to identify and quantify the different cell types present in the organoids. Eight separate clusters (Figures 6F and 6G) were identified, including cerebral cells of different stages of maturation, such as neural progenitor cells, newborn neurons, and mature neurons (Figures 6F, 6G, and S6A). All of the clusters contained cells from both the 2- and 4-month time points, and we found no apparent difference in the contribution to the different clusters by ZNF558-CRISPRi organoids (Figures 6F, 6G, and S6B), suggesting that ZNF558 does not influence developmental fate during early human brain development.

Next, we analyzed the transcriptional difference between control and ZNF558-KD organoids. We confirmed the transcriptional silencing of *ZNF558* in all cell populations at both time points (Figure S6C) and the upregulation of *SPATA18* after ZNF558 inhibition in NPCs (Figure S6D). We found that in the neuron cluster, many genes linked to neuronal differentiation and maturation, such as *MYT1L*, *NCAM1*, or *RBFOX1*, were upregulated in ZNF558-CRISPRi organoids (Figure 6H). In addition, several genes linked to the establishment of neuronal projections, such as *DCC* and *ROBO1*, were upregulated in ZNF558-CRISPRi organoids (Figure 6H). An unbiased Gene Ontology (GO) analysis of genes upregulated in organoid neurons lacking ZNF558 confirmed a significant enrichment of genes in categories linked to neuronal differentiation and maturation as well as neuron projection and synapse formation (Figure 6I). These results indicate that neurons present in organoids that lack ZNF558 display a more mature transcriptional profile than those found in control organoids.

These results demonstrate that silencing of ZNF558 results in cerebral organoids that contain the same cell types as control organoids, suggesting that ZNF558 does not influence developmental fate. However, we noted that ZNF558 organoids were smaller during early differentiation and displayed a transcriptome indicative of the presence of more mature neurons at later stages of differentiation. These observations are in line with a role for ZNF558 in developmental timing during early brain development.

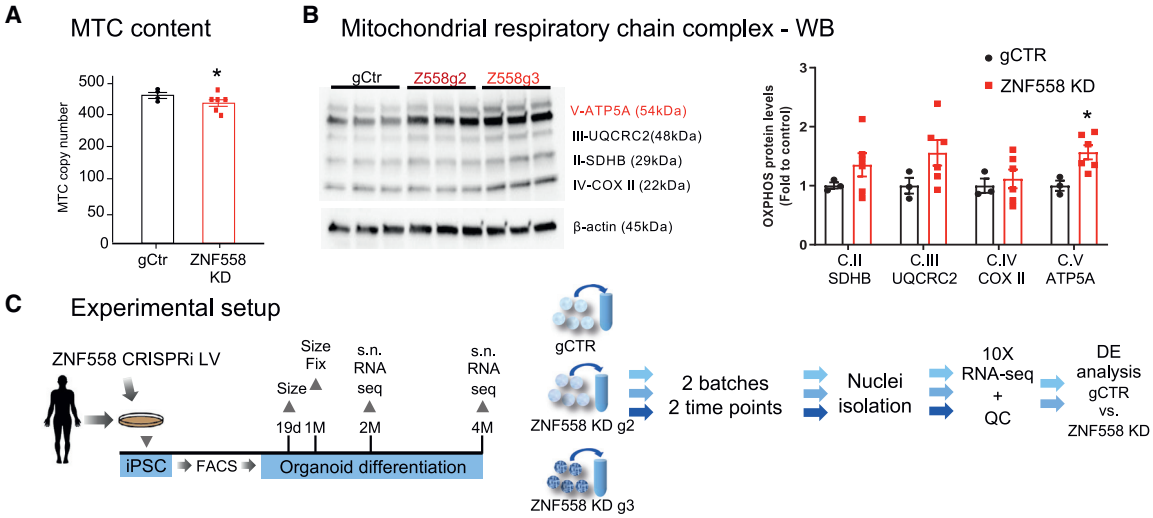
Regulation of ZNF558 is controlled by a *cis*-acting VNTR

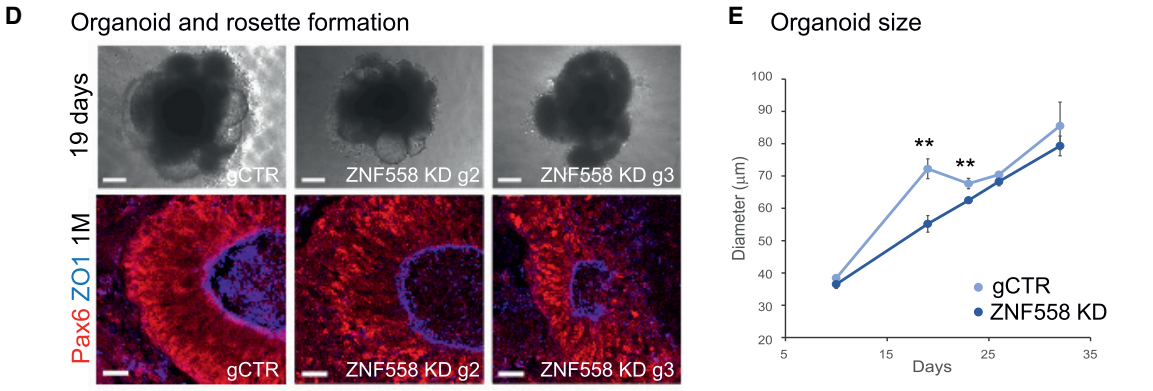
Given that *ZNF558* is conserved between human and chimpanzee, we next explored the underlying genetic basis for differential expression between the 2 species. We found limited variation in the human and chimpanzee genomic context upstream of *ZNF558*, suggesting that alterations of promoter structure are unlikely to be responsible. However, downstream of *ZNF558* we noted a large repetitive region in the human genome assembly (chr19:8735220-8794192, GRCh38; Figure 7A). We investigated copy-number variation at this locus by generating read-depth profiles for a collection of 1,112 high-coverage human and non-human great ape genomes from publicly available datasets (Figure 7B). Our analysis identified a VNTR motif with a unit length of 7,460 bp, which showed striking copy-number variation across populations and species (Figures 7B and 7C). In particular, we observed a significant difference in copy number at this VNTR locus between the human and non-human primate

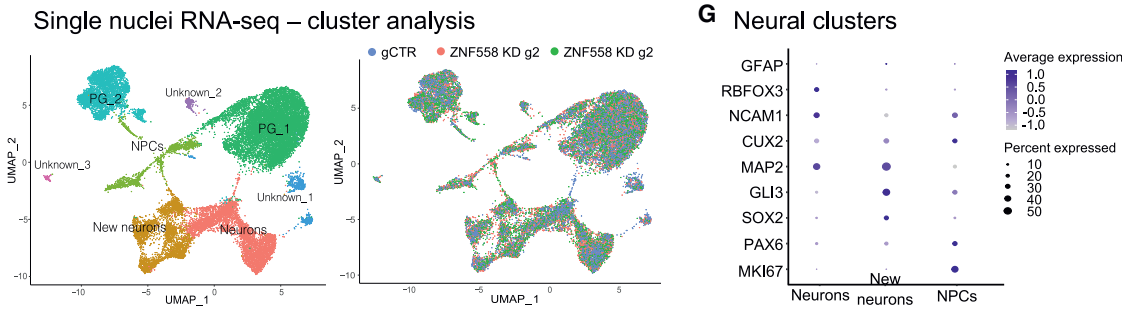
(C) RNA-seq analysis of ZNF558-CRISPRi fbNPCs for expression of TEs (left) and protein coding genes (right), p adjusted < 0.05, |log2(fc) > 1|.
(D) qRT-PCR analysis of SPATA18 expression in fbNPCs after ZNF558 inhibition (n = 4).
(E) H3K4me3 CUT&RUN after ZNF558 inhibition.
(F) Epigenomic (H3K4me3, H3K9me3), transcriptomic (RNA-seq), and ZNF558 binding (ChIP-exo) analysis of the *SPATA18* locus in human and chimp fbNPCs. Apart from the ChIP-exo analysis in HEK293T (Imbeault et al., 2017), data were obtained in at least biological duplicate (i.e., cells derived from 2 individuals for each species) with similar results. CUT&RUN data were normalized based on spike-in DNA. Scales for RNA-seq and ChIP-exo data are RPKM.
(G) Schematics of the experimental strategy to overexpress ZNF558 in chimpanzee fbNPCs.
(H and I) qRT-PCR analysis of ZNF558 and SPATA18 in chimpanzee fbNPCs transduced with a lentiviral vector overexpressing ZNF558 (n = 4). *p < 0.05 (Student's t test); error bars are mean ± SEM.
See also Figures S4 and S5.

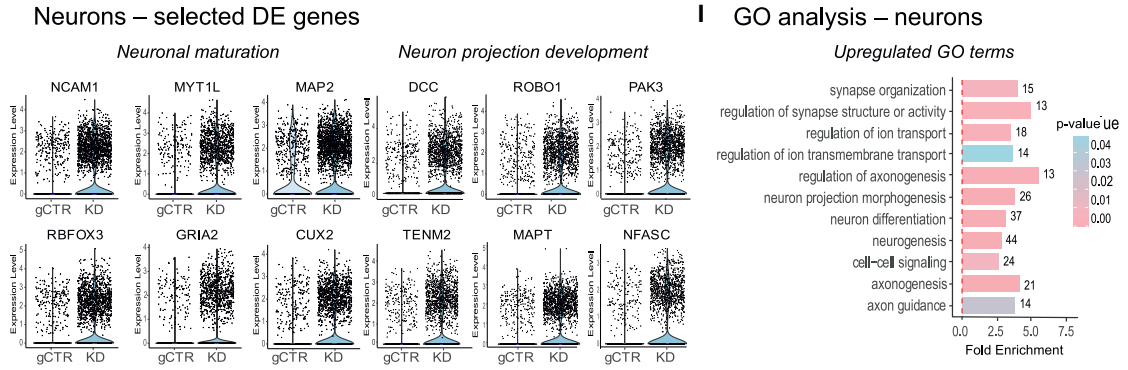
Please cite this article in press as: Johansson et al., A *cis*-acting structural variation at the *ZNF558* locus controls a gene regulatory network in human brain development, Cell Stem Cell (2021), https://doi.org/10.1016/j.stem.2021.09.008

CellPress
OPEN ACCESS









(legend on next page)

Please cite this article in press as: Johansson et al., A *cis*-acting structural variation at the *ZNF558* locus controls a gene regulatory network in human brain development, Cell Stem Cell (2021), https://doi.org/10.1016/j.stem.2021.09.008

(NHP) samples; all non-human great ape samples, except for orangutans, have significantly higher copy numbers (>70) of this VNTR compared to humans (means: 24–43 copies) ($p = 8.61 \times 10^{-21}$, Mann-Whitney *U* test; Figure 7C). This result is consistent with qPCR data using genomic DNA obtained from the human and chimpanzee iPSC lines used in the present study (Figure 7D). One interesting outlier in this analysis is the orangutan, which has a VNTR copy number similar to that of humans (mean 36, standard deviation 8.1) compared to that of other non-human great apes. Published transcriptome data showed high ZNF558 transcript levels in orangutan forebrain organoids (Field et al., 2019) (Figure S2D). We noted that the VNTR copy number is variable in the human population. To investigate whether this variation influences ZNF558 expression, we analyzed publicly available genomic and expression data from 409 human lymphoblastoid cell lines (Geuvadis consortium, Lappalainen et al., 2013). In line with our hypothesis, we found that ZNF558 expression was negatively correlated with VNTR copy number ($R = -0.24$, $p = 0.0000013$; Figure 7E). This effect was specific to ZNF558 as the expression of other genes located within 250 kb of the VNTR was not correlated with copy number (Figure S7A). These results demonstrate that a lower VNTR copy number correlates with higher ZNF558 expression, both across hominoid species and within the human population and suggest that the repetitive genomic region downstream of *ZNF558* is responsible for its differential expression.

Epigenetic manipulation of the human VNTR switches the ZNF558-SPATA18 regulatory network

Finally, we investigated the mechanism of *cis*-regulation by the VNTR. We considered that the longer VNTR in chimpanzees could have established a repressive hub leading to *ZNF558* silencing. Profiling of H3K27me3 and H3K9me3 levels over the human and chimpanzee VNTRs is challenging due to the low mapability of this genomic region. However, exploiting a spike-in normalization strategy for CUT&RUN data, we confirmed that relative read depths for control immunoglobulin G (IgG) tracks matched the relative VNTR copy numbers (Figure S7C), thereby validating our mapping strategy. H3K9me3 but not H3K27me3 enrichment was detected over the VNTR in both human and chimpanzee samples using this approach (Figures S7B and S7D). These data are consistent with a model in which the VNTR attracts H3K9me3, the functional consequence of which is controlled by the size of the VNTR. RNA-seq analysis showed evidence of transcription within the VNTR in human but not chimpanzee fbNPCs (Figure S7B), and we found evidence of a human-specific long non-coding RNA (lncRNA) originating in the VNTR transcribed in the antisense orientation relative to *ZNF558* (Figure S7E). These observations suggest that the shorter human VNTR is not fully silenced.

To experimentally test the model of VNTR-mediated *cis*-repression, we used a multi-guide CRISPRi strategy to impose silencing over the shorter human VNTR. We designed 3 different gRNAs to target different parts of the VNTR unit to co-express with the dCas9 repressor fusion (Figures 7F and 7G). Upon iPSC transduction, we observed reduced expression of the VNTR-associated lncRNA (Figure 7H), in line with targeted silencing. Strikingly, we observed robust inhibition of ZNF558 transcription upon VNTR silencing, which in turn led to increased levels of SPATA18 transcripts (Figures 7I and 7J). These data support the model of a brain-specific gene-regulatory network in which a downstream VNTR controls *ZNF558* expression in *cis*, which in turn affects the repression of its target gene *SPATA18*.

DISCUSSION

In this study, we show how genetic variations in non-coding regions of the genome can control the activity of conserved protein-coding genes, resulting in the establishment of species-specific transcriptional networks. To date, there is limited evidence that changes in *cis*-acting regulatory elements are important for human brain evolution. Classically, non-coding genetic changes are thought to result in gene expression differences along a gradient, leading to slightly more or less mRNA product of a nearby gene. It remains debatable whether such differences in mRNA levels correspond to changes in protein levels, since there is evidence of compensatory buffering at the translational level (Khan et al., 2013). It also remains unclear how differences in the level of a protein affects species fitness. In contrast, our results demonstrate that non-coding regions have the capacity to mediate an on/off switch of a conserved protein-coding gene.

KZFPs are the largest family of mammalian transcription factors and have rapidly expanded and evolved in the primate lineage. It has been proposed that KZFPs are engaged in an "arms race" with TEs, in which KZFPs evolve to bind specific TEs and silence their activity (Jacobs et al., 2014). In this model, TEs mutate their ZFP-targeted sites, thereby escaping from silencing

Figure 6. Functional analysis of ZNF558 in cerebral organoids

(A) Estimation of mitochondrial copy number via mitochondrial and nuclear DNA ratio in fbNPCs. *$p < 0.05$. Student's t test, n = 4–6, bars are mean ± SEM.

(B) Western blot analysis and quantifications of mitochondrial respiratory chain complexes. n = 3–6; data are represented as mean ± SEM; *$p < 0.05$, Student's t test.

(C) Schematics of the workflow for the generation and analysis of human cerebral organoids grown from ZNF558-CRISPRi (g2 and g3) and control iPSCs.

(D) Pictures of iPSCs-derived cerebral organoids in the culture dish (top, day 19; scale bar, 200 μm) and after immunohistochemistry for Pax6/ZO1 (bottom, 1 month; scale bar, 25 μm).

(E) Quantification of organoid diameter between days 10 and 32. n = 7–27, dots are mean ± SEM, **$p < 0.01$. Student's t test.

(F) Single nuclei-RNA seq: UMAP clustering (left) and the distribution of control and the 2 ZNF558-CRISPRi-derived nuclei over the clusters (right).

(G) Dot blot showing expression of neuronal and neural progenitor cell markers in the NPC, New neurons, and Neurons clusters.

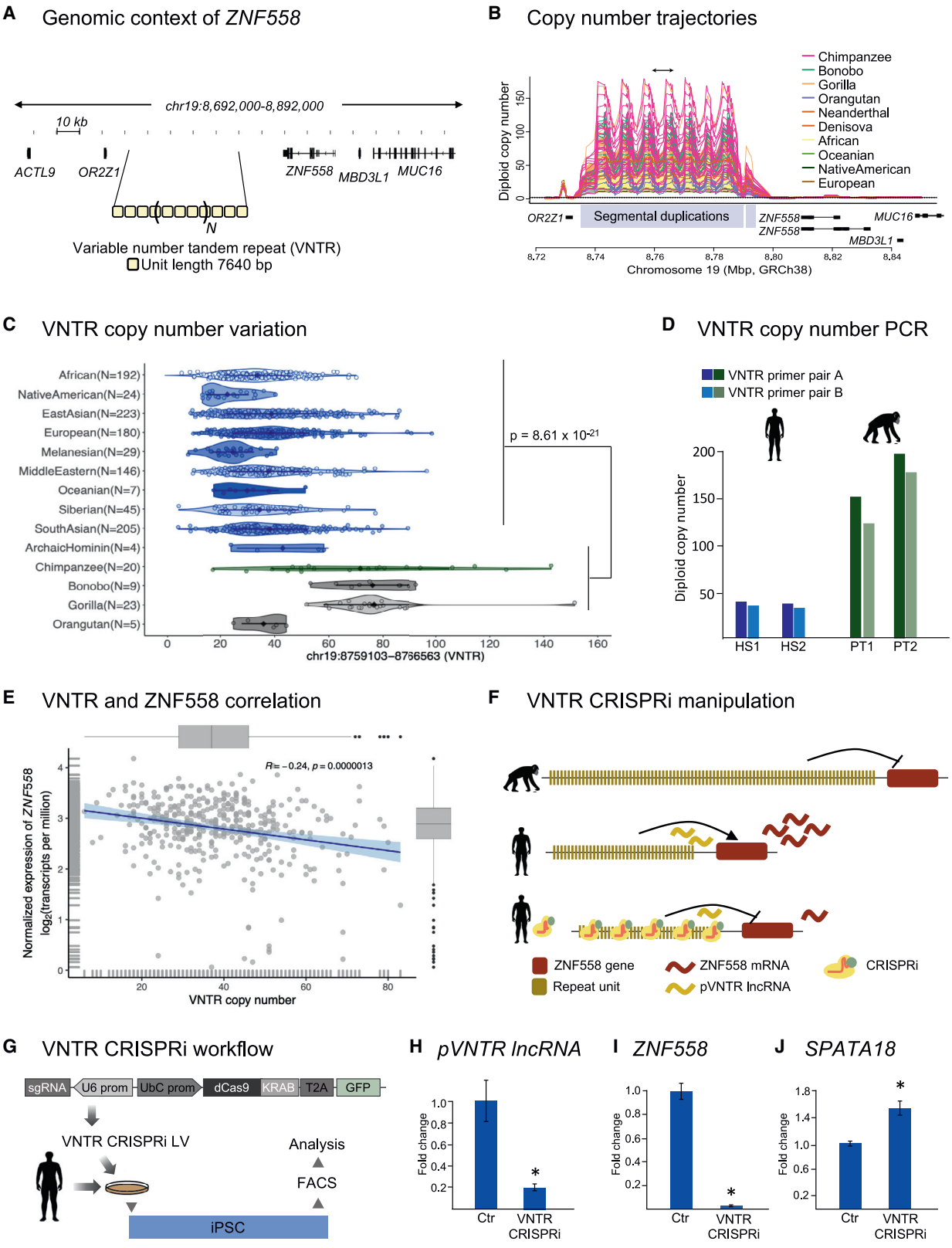
(H) Examples of significant cell-type-specific changes in gene expression between control (gCTR) and ZNF558-CRISPRi organoids in the Neurons cluster as revealed by single-nuclei RNA-seq (p adjusted < 0.01).

(I) Bar graph showing significant GO terms in the upregulated genes from the Neurons cluster after ZNF558-CRISPRi. Number after each bar represent number of differentially expressed genes in each category.

See also Figure S6.

Please cite this article in press as: Johansson et al., A *cis*-acting structural variation at the *ZNF558* locus controls a gene regulatory network in human brain development, Cell Stem Cell (2021), https://doi.org/10.1016/j.stem.2021.09.008

CellPress
OPEN ACCESS



(legend on next page)

Please cite this article in press as: Johansson et al., A *cis*-acting structural variation at the *ZNF558* locus controls a gene regulatory network in human brain development, Cell Stem Cell (2021), https://doi.org/10.1016/j.stem.2021.09.008

and regaining activity. The KZFP would then further evolve to once again silence the TE, resulting in a dynamic competition between KZFPs and TEs that drive their evolution. This evolutionary mechanism has been proposed to have been used by host genomes to drive evolutionary mechanisms (Trono, 2015). In this study, we provide evidence of human-specific expression of several KZFPs in brain development. One example, *ZNF558*, arose >100 mya to repress then-active LINE-1 transposons. Although its TE targets have degenerated, ZNF558 is highly conserved among mammals, in line with functional biological roles independent of TE suppression. ZNF558 now binds to both ancient L1s—likely representing a genomic fossil—and several gene targets in human cells (Imbeault et al., 2017). During human brain development, we found that ZNF558 regulates a single protein-coding gene, *SPATA18.*

The genetic basis for the unusual on/off switch of ZNF558 expression between human and chimpanzee fbNPCs resides in a downstream VNTR. The length of the VNTR correlates with the repression of *ZNF558*; in humans, in whom the repeat is 30–40 units long, the locus is linked to active transcription, while in NHPs, the VNTR is much larger, and this is associated with transcriptional repression. We do not fully understand the molecular mechanisms that underlie how differences in VNTR unit copy number affects ZNF558 expression. The VNTR is decorated by the repressive histone mark H3K9me3 both in human and chimpanzee fbNPCs. Since the chimpanzee VNTR is longer, a VNTR-based heterochromatin region is substantially larger. We also found evidence that the short human VNTR unit is transcribed, giving rise to a lncRNA, which may influence in *cis* the expression of nearby genes such as *ZNF558*. The most likely explanation for the observed phenomenon is that the extended VNTR is a heterochromatic region and that its shortening reveals a brain-specific enhancer and/or lncRNA in the repeat. Such a copy-number-dependent VNTR-based epigenetic regulatory mechanism of nearby protein coding gene expression is reminiscent of what has been previously observed at the *FSHD* locus (Cabianca et al., 2012).

Our observations are in line with a model in which the common ancestor of humans and NHPs carried a long VNTR downstream of *ZNF558*. This VNTR contracted after the human/chimpanzee evolutionary split, ultimately resulting in differences in SPATA18 levels and altered mitochondria homeostasis in the developing human brain. However, this is a simplified model, and the functional role of *ZNF558* is likely multifaceted. We note that both orangutans and macaques express ZNF558 during brain development (Field et al., 2019; Kanton et al., 2019) and have a conserved ZNF588 binding site in SPATA18. Thus, the ZNF558-mediated repression of SPATA18 during brain development is not only a human-specific mechanism but is also present in other primate species. In addition, *ZNF558* is expressed in most adult tissues in humans, and we find no evidence of a divergence of expression between human and chimpanzee in most of these tissues. Thus, ZNF558 is likely to perform additional roles in adult tissues. While ZNF558 is a highly conserved gene, the ZNF558-binding site in *SPATA18* is conserved only in primates, suggesting that ZNF558 plays a different role in other species, to regulate alternative co-opted targets. It will be interesting to chart the functional roles of ZNF558 in other more distantly related mammals, such as mice, in which innovation in target sequences may have led to co-option of this transcription factor for other functional roles.

The downstream consequences of a shortened VNTR and activated ZNF558 expression during human brain development is transcriptional repression of *SPATA18*. The product of *SPATA18* is MIEAP, which clears mitochondria via mitophagy (Kitamura et al., 2011). The emerging literature indicates that mitochondria are central in neural stem cell fate decisions and the process of neurogenesis partly through the control of a metabolic switch from glycolysis to OXPHOS occurring when NPCs differentiate to neurons (Arrázola et al., 2019; Khacho et al., 2016, 2019). ARHGAP11B, a human-specific gene (Florio et al., 2015), is localized to the mitochondria, where it induces a cancer-like metabolism to promote the proliferation of basal neural progenitors (Namba et al., 2020). This recent observation directly links mitochondrial function to speciation and the expansion of the human cerebral cortex. Exactly how SPATA18, regulated by ZNF558 uniquely in humans, feeds into this metabolic pathway, remains to be determined, but our results using cerebral organoids indicate that ZNF558 plays a role in developmental timing during brain development. Organoids in which ZNF558 expression was silenced appeared smaller in size during the early phase of differentiation and displayed neurons with a

Figure 7. Variations in copy number of a downstream VNTR regulates ZNF558 expression in *cis*

(A) Schematic illustrating the presence of a VNTR downstream of the *ZNF558* gene on chromosome 19 (GRCh38).

(B) Copy-number trajectories illustrating an increase in copy number for a 7.46-kbp VNTR motif (chr19:8,759,103−8,766,563, GRCh38) in chimpanzee, bonobo, and gorilla relative to humans. Each line is the inferred copy-number trajectory of a given sample over this region. Black arrows at the top indicate a unit of the VNTR motif. The dashed lines indicate diploid copy number 2.

(C) Copy-number variation for the 7.46-kbp VNTR motif among archaic and modern humans and non-human great ape samples. Each dot is the overall (diploid) copy number for a given sample. Black diamonds and bars represent the mean and 1 standard deviation of the VNTR copy numbers in individual populations and species (Mann-Whitney *U* test).

(D) qPCR analysis of the VNTR copy number in the human and chimp iPSC lines used in this study.

(E) Significantly negative correlation between the normalized expression of ZNF558 and the copy number of the 7.46-kbp VNTR motif (chr19: 8,759,103−8,766,563) near ZNF558 in humans. RNA-seq data generated by the Geuvadis consortium, which includes samples from the 1000 Genomes project. Each dot is 1 of the 409 samples that have data for both ZNF558 gene expression and copy-number estimates for the VNTR motif. The blue line represents linear regression, and the shaded area indicates the 95% confidence interval.

(F and G) Schematics of the CRISPRi-based multi-guide strategy to epigenetically silence the human VNTR. pVNTR lncRNA, human-specific putative lncRNA originating from within in the VNTR extending toward ZNF558.

(H–J) qRT-PCR analysis of VNTR-lncRNA, ZNF558, and SPATA18 in human iPSCs (HS1 and HS2) transduced with VNTR-CRISPRi lentiviral vectors (n = 5–6). *$p < 0.05$, (Student's t test); data represented as mean ± SEM in (H).

See also Figure S7.

more mature transcriptome during later growth stages. These findings are reminiscent of previously observed differences when comparing human cerebral organoids to those derived from non-human great apes (Benito-Kwiecinski et al., 2021; Kanton et al., 2019; Marchetto et al., 2019).

Our data also demonstrate that the length of the VNTR at the *ZNF558* locus is variable in the human population and that there are individuals with a repeat length similar to that of chimpanzees. We found that the length of the VNTR negatively correlates with ZNF558 expression in a panel of human lymphoblastoid cell lines. Unfortunately, it was not possible to resolve the actual composition of VNTR alleles for each individual from this dataset. However, the molecular and phenotypical analysis of human individuals carrying bi-allelic long VNTRs would be interesting and may uncover a role for VNTRs in human phenotypical variation. We also note that DNA duplications of the *SPATA18* locus can lead to intellectual disability and delayed language and speech development (https://www.deciphergenomics.org), suggesting that *SPATA18* and its partner transcription factor ZNF558 may be linked to neurodevelopmental disorders.

In summary, our results illustrate how one KRAB-ZFP, ZNF558, was co-opted and subsequently contributed to human brain evolution. Future studies of KRAB-ZFPs and VNTRs in human evolution and human disease will be interesting and rewarding.

Limitations of the study

The fbNPC and organoid model systems used in this project recapitulates some aspects of human brain development. However, there are limitations to the accuracy of the cell models in comparison to actual human brain development, and they are associated with experimental variation which limits the strength of some conclusions. To validate the key findings in our study, we have relied on human fetal tissue. However, such material is not available from chimpanzees, and it was therefore not possible to validate the lack of ZNF558 expression in actual chimpanzee brain development.

Furthermore, our iPSC-modeling relies on a limited number of different chimpanzee iPSC lines. Where possible, we have cross-validated our results with published data from other iPSC lines, but the number of available iPSCs from different chimpanzee individuals are still very few (~10) and associated with shipment restrictions. This represents a limitation for our understanding of how ZNF558 expression is variable in the chimpanzee population.

Finally, our data demonstrate a *cis*-acting role for a downstream VNTR in controlling *ZNF558* gene expression. Additional molecular investigations are needed to fully resolve the mechanism of how this non-coding region exerts its effect.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.stem.2021.09.008.

ACKNOWLEDGMENTS

We would like to thank Fred H. Gage, Wieland Huttner, Svante Pääbo, Hindrik Mulder, Barbara Treutlein, Steve Henikoff, He Zhisong, Mitchell Vollger, and Yafei Mao for their comments and support. We also thank M. Persson Vejgården and A. Hammarberg for technical assistance. We are grateful to all of the members of the Jakobsson lab. The work was supported by grants from the Swedish Research Council (2018-02694, to J.J.), the Swedish Brain Foundation (FO2019-0098, to J.J.), Cancerfonden (190326, to J.J.), Barncancerfonden (PR2017-0053, to J.J.), the Swedish Society for Medical Research (S19-0100, to C.H.D.) and the Swedish Government Initiative for Strategic Research Areas (MultiPark & StemTherapy). E.E.E. is an investigator of the Howard Hughes Medical Institute. M.P. is a New York Stem Cell Foundation Robertson Investigator.

AUTHOR CONTRIBUTIONS

All of the authors took part in designing the study and interpreting the data. P.A.J., P.L.B., and J.J. conceived and designed the study. P.A.J., C.H.D., D.G., J.J., D.A.M.A., M.E.J., A.A., E.S., R.C., and K.P. performed the experimental research. P.L.B., C.H.D., P.H., R.G., Y.S., and F.E. performed the bioinformatic analyses. J.P., D.T., M.P., A.F., M.F., and E.E.E. contributed reagents and expertise. P.A.J., P.L.B., C.H.D., and J.J. wrote the manuscript, and all of the authors reviewed the final version.

DECLARATIONS OF INTEREST

The authors declare no competing interests.

Received: August 20, 2020
Revised: July 20, 2021
Accepted: September 13, 2021
Published: October 7, 2021

REFERENCES

Arrázola, M.S., Andraini, T., Szelechowski, M., Mouledous, L., Arnauné-Pelloquin, L., Davezac, N., Belenguer, P., Rampon, C., and Miquel, M.C. (2019). Mitochondria in Developmental and Adult Neurogenesis. Neurotox. Res. *36*, 257–267.

Ayyanathan, K., Lechner, M.S., Bell, P., Maul, G.G., Schultz, D.C., Yamada, Y., Tanaka, K., Torigoe, K., and Rauscher, F.J., 3rd (2003). Regulated recruitment of HP1 to a euchromatic gene induces mitotically heritable, epigenetic gene silencing: a mammalian cell culture model of gene variegation. Genes Dev. *17*, 1855–1869.

Bailey, T.L., Johnson, J., Grant, C.E., and Noble, W.S. (2015). The MEME Suite. Nucleic Acids Res. *43* (W1), W39–W49.

Benito-Kwiecinski, S., Giandomenico, S.L., Sutcliffe, M., Riis, E.S., Freire-Pritchett, P., Kelava, I., Wunderlich, S., Martin, U., Wray, G.A., McDole, K., et al. (2021). An early cell shape transition drives evolutionary expansion of the human forebrain. Cell *184*, 2084–2102.e19.

Bergström, A., McCarthy, S.A., Hui, R., Almarri, M.A., Ayub, Q., Danecek, P., Chen, Y., Felkel, S., Hallast, P., Kamm, J., et al. (2020). Insights into human genetic variation and population history from 929 diverse genomes. Science *367*, eaay5012.

Brattås, P.L., Jönsson, M.E., Fasching, L., Nelander Wahlestedt, J., Shahsavani, M., Falk, R., Falk, A., Jern, P., Parmar, M., and Jakobsson, J. (2017). TRIM28 Controls a Gene Regulatory Network Based on Endogenous Retroviruses in Human Neural Progenitor Cells. Cell Rep. *18*, 1–11.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. *36*, 411–420.

Cabianca, D.S., Casa, V., Bodega, B., Xynos, A., Ginelli, E., Tanaka, Y., and Gabellini, D. (2012). A long ncRNA links copy number variation to a polycomb/trithorax epigenetic switch in FSHD muscular dystrophy. Cell *149*, 819–831.

Dennis, M.Y., and Eichler, E.E. (2016). Human adaptation and evolution by segmental duplication. Curr. Opin. Genet. Dev. *41*, 44–52.

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.

Emerson, R.O., and Thomas, J.H. (2009). Adaptive evolution in zinc finger transcription factors. PLoS Genet. *5*, e1000325.

Enard, W. (2016). The Molecular Basis of Human Brain Evolution. Curr. Biol. *26*, R1109–R1117.

Farmiloe, G., Lodewijk, G.A., Robben, S.F., van Bree, E.J., and Jacobs, F.M.J. (2020). Widespread correlation of KRAB zinc finger protein binding with brain-developmental gene expression patterns. Philos. Trans. R. Soc. Lond. B Biol. Sci. *375*, 20190333.

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.

Fiddes, I.T., Lodewijk, G.A., Mooring, M., Bosworth, C.M., Ewing, A.D., Mantalas, G.L., Novak, A.M., van den Bout, A., Bishara, A., Rosenkrantz, J.L., et al. (2018). Human-Specific NOTCH2NL Genes Affect Notch Signaling and Cortical Neurogenesis. Cell *173*, 1356–1369.e22.

Field, A.R., Jacobs, F.M.J., Fiddes, I.T., Phillips, A.P.R., Reyes-Ortiz, A.M., LaMontagne, E., Whitehead, L., Meng, V., Rosenkrantz, J.L., Olsen, M., et al. (2019). Structurally Conserved Primate LncRNAs Are Transiently Expressed during Human Cortical Differentiation and Influence Cell-Type-Specific Genes. Stem Cell Reports *12*, 245–257.

Florio, M., Albert, M., Taverna, E., Namba, T., Brandl, H., Lewitus, E., Haffner, C., Sykes, A., Wong, F.K., Peters, J., et al. (2015). Human-specific gene ARHGAP11B promotes basal progenitor amplification and neocortex expansion. Science *347*, 1465–1470.

Friedli, M., and Trono, D. (2015). The developmental control of transposable elements and the evolution of higher species. Annu. Rev. Cell Dev. Biol. *31*, 429–451.

Friedman, J.R., Fredericks, W.J., Jensen, D.E., Speicher, D.W., Huang, X.P., Neilson, E.G., and Rauscher, F.J., 3rd (1996). KAP-1, a novel corepressor for the highly conserved KRAB repression domain. Genes Dev. *10*, 2067–2078.

Gallego Romero, I., Pavlovic, B.J., Hernando-Herraez, I., Zhou, X., Ward, M.C., Banovich, N.E., Kagan, C.L., Burnett, J.E., Huang, C.H., Mitrano, A., et al. (2015). A panel of induced pluripotent stem cells from chimpanzees: a resource for comparative functional genomics. eLife *4*, e07103.

Grassi, D.A., Brattås, P.L., Jönsson, M.E., Atacho, D., Karlsson, O., Nolbrant, S., Parmar, M., and Jakobsson, J. (2019). Profiling of lincRNAs in human pluripotent stem cell derived forebrain neural progenitor cells. Heliyon *6*, e03067.

Hach, F., Hormozdiari, F., Alkan, C., Hormozdiari, F., Birol, I., Eichler, E.E., and Sahinalp, S.C. (2010). mrsFAST: a cache-oblivious algorithm for short-read mapping. Nat. Methods *7*, 576–577.

Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S., et al. (2012). GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res *22,* 1860–1774. https://doi.org/10.1101/gr.135350.111.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

Hill, R.S., and Walsh, C.A. (2005). Molecular insights into human brain evolution. Nature *437*, 64–67.

Imbeault, M., Helleboid, P.Y., and Trono, D. (2017). KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. Nature *543*, 550–554.

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.

Jin, Y., Tam, O.H., Paniagua, E., and Hammell, M. (2015). TEtranscripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets. Bioinformatics *31*, 3593–3599. https://doi.org/10.1093/bioinformatics/btv422.

Johnson, M.B., Kawasawa, Y.I., Mason, C.E., Krsnik, Z., Coppola, G., Bogdanović, D., Geschwind, D.H., Mane, S.M., State, M.W., and Sestan, N. (2009). Functional and evolutionary insights into human brain development through global transcriptome analysis. Neuron *62*, 494–509.

Ju, X.C., Hou, Q.Q., Sheng, A.L., Wu, K.Y., Zhou, Y., Jin, Y., Wen, T., Yang, Z., Wang, X., and Luo, Z.G. (2016). The hominoid-specific gene TBC1D3 promotes generation of basal neural progenitors and induces cortical folding in mice. eLife *5*, e18197.

Kanton, S., Boyle, M.J., He, Z., Santel, M., Weigert, A., Sanchís-Calleja, F., Guijarro, P., Sidow, L., Fleck, J.S., Han, D., et al. (2019). Organoid single-cell genomic atlas uncovers human-specific features of brain development. Nature *574*, 418–422.

Kapusta, A., Kronenberg, Z., Lynch, V.J., Zhuo, X., Ramsay, L., Bourque, G., Yandell, M., and Feschotte, C. (2013). Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. PLoS Genet. *9*, e1003470.

Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al.; Genome Aggregation Database Consortium (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. Nature *581*, 434–443.

Kent, W.J., Zweig, A.S., Barber, G., Hinrichs, A.S., and Karolchik, D. (2010). BigWig and BigBed: enabling browsing of large distributed datasets. Bioinformatics *26*, 2204–2207. https://doi.org/10.1093/bioinformatics/btq351.

Khacho, M., Clark, A., Svoboda, D.S., Azzi, J., MacLaurin, J.G., Meghaizel, C., Sesaki, H., Lagace, D.C., Germain, M., Harper, M.E., et al. (2016).

Mitochondrial Dynamics Impacts Stem Cell Identity and Fate Decisions by Regulating a Nuclear Transcriptional Program. Cell Stem Cell *19*, 232–247.

Khacho, M., Harris, R., and Slack, R.S. (2019). Mitochondria as central regulators of neural stem cell fate and cognitive function. Nat. Rev. Neurosci. *20*, 34–48.

Khaitovich, P., Enard, W., Lachmann, M., and Pääbo, S. (2006). Evolution of primate gene expression. Nat. Rev. Genet. *7*, 693–702.

Khan, Z., Ford, M.J., Cusanovich, D.A., Mitrano, A., Pritchard, J.K., and Gilad, Y. (2013). Primate transcript and protein expression levels evolve under compensatory selection pressures. Science *342*, 1100–1104.

King, M.C., and Wilson, A.C. (1975). Evolution at two levels in humans and chimpanzees. Science *188*, 107–116.

Kitamura, N., Nakamura, Y., Miyamoto, Y., Miyamoto, T., Kabu, K., Yoshida, M., Futamura, M., Ichinose, S., and Arakawa, H. (2011). Mieap, a p53-inducible protein, controls mitochondrial quality by repairing or eliminating unhealthy mitochondria. PLoS ONE *6*, e16060.

Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P.R., and Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. Nat. Methods *16*, 1289–1296.

Kronenberg, Z.N., Fiddes, I.T., Gordon, D., Murali, S., Cantsilieris, S., Meyerson, O.S., Underwood, J.G., Nelson, B.J., Chaisson, M.J.P., Dougherty, M.L., et al. (2018). High-resolution comparative analysis of great ape genomes. Science *360*, eaar6343.

Lai, C.S., Fisher, S.E., Hurst, J.A., Vargha-Khadem, F., and Monaco, A.P. (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. Nature *413*, 519–523.

Lancaster, M.A., Renner, M., Martin, C.A., Wenzel, D., Bicknell, L.S., Hurles, M.E., Homfray, T., Penninger, J.M., Jackson, A.P., and Knoblich, J.A. (2013). Cerebral organoids model human brain development and microcephaly. Nature *501*, 373–379.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357–359. https://doi.org/10.1038/nmeth.1923.

Lappalainen, T., Sammeth, M., Friedländer, M.R., 't Hoen, P.A., Monlong, J., Rivas, M.A., Gonzàlez-Porta, M., Kurbatova, N., Griebel, T., Ferreira, P.G., et al.; Geuvadis Consortium (2013). Transcriptome and genome sequencing uncovers functional variation in humans. Nature *501*, 506–511.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (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.

Lui, J.H., Hansen, D.V., and Kriegstein, A.R. (2011). Development and evolution of the human neocortex. Cell *146*, 18–36.

Lynch, V.J., Nnamani, M.C., Kapusta, A., Brayer, K., Plaza, S.L., Mazur, E.C., Emera, D., Sheikh, S.Z., Grützner, 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. https://doi.org/10.1016/j.celrep.2014.12.052.

Mafessoni, F., Grote, S., de Filippo, C., Slon, V., Kolobova, K.A., Viola, B., Markin, S.V., Chintalapati, M., Peyrégne, S., Skov, L., et al. (2020). A high-coverage Neandertal genome from Chagyrskaya Cave. Proc. Natl. Acad. Sci. USA *117*, 15132–15136.

Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., Zhao, M., Chennagiri, N., Nordenfelt, S., Tandon, A., et al. (2016). The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. Nature *538*, 201–206.

Marchetto, M.C.N., Narvaiza, I., Denli, A.M., Benner, C., Lazzarini, T.A., Nathanson, J.L., Paquola, A.C.M., Desai, K.N., Herai, R.H., Weitzman, M.D., et al. (2013). Differential L1 regulation in pluripotent stem cells of humans and apes. Nature *503*, 525–529.

Marchetto, M.C., Hrvoj-Mihic, B., Kerman, B.E., Yu, D.X., Vadodaria, K.C., Linker, S.B., Narvaiza, I., Santos, R., Denli, A.M., Mendes, A.P., et al. (2019). Species-specific maturation profiles of human, chimpanzee and bonobo neural cells. eLife *8*, e37527.

Matsui, T., Leung, D., Miyashita, H., Maksakova, I.A., Miyachi, H., Kimura, H., Tachibana, M., Lorincz, M.C., and Shinkai, Y. (2010). Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. Nature *464*, 927–931.

Meyer, M., Kircher, M., Gansauge, M.T., Li, H., Racimo, F., Mallick, S., Schraiber, J.G., Jay, F., Prüfer, K., de Filippo, C., et al. (2012). A high-coverage genome sequence from an archaic Denisovan individual. Science *338*, 222–226.

Mora-Bermúdez, F., Badsha, F., Kanton, S., Camp, J.G., Vernot, B., Köhler, K., Voigt, B., Okita, K., Maricic, T., He, Z., et al. (2016). Differences and similarities between human and chimpanzee neural progenitors during cerebral cortex development. eLife *5*, e18683.

Murphy, K.E., Shylo, N.A., Alexander, K.A., Churchill, A.J., Copperman, C., and García-García, M.J. (2016). The Transcriptional Repressive Activity of KRAB Zinc Finger Proteins Does Not Correlate with Their Ability to Recruit TRIM28. PLoS ONE *11*, e0163555.

Najafabadi, H.S., Mnaimneh, S., Schmitges, F.W., Garton, M., Lam, K.N., Yang, A., Albu, M., Weirauch, M.T., Radovani, E., Kim, P.M., et al. (2015). C2H2 zinc finger proteins greatly expand the human regulatory lexicon. Nat. Biotechnol. *33*, 555–562.

Namba, T., Dóczi, J., Pinson, A., Xing, L., Kalebic, N., Wilsch-Bräuninger, M., Long, K.R., Vaid, S., Lauer, J., Bogdanova, A., et al. (2020). Human-Specific ARHGAP11B Acts in Mitochondria to Expand Neocortical Progenitors by Glutaminolysis. Neuron *105*, 867–881.e9.

Nolbrant, S., Heuer, A., Parmar, M., and Kirkeby, A. (2017). Generation of high-purity human ventral midbrain dopaminergic progenitors for in vitro maturation and intracerebral transplantation. Nat. Protoc. *12*, 1962–1979.

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.

Nowick, K., Hamilton, A.T., Zhang, H., and Stubbs, L. (2010). Rapid sequence and expression divergence suggest selection for novel function in primate-specific KRAB-ZNF genes. Mol. Biol. Evol. *27*, 2606–2617.

Patel, A., Yang, P., Tinkham, M., Pradhan, M., Sun, M.A., Wang, Y., Hoang, D., Wolf, G., Horton, J.R., Zhang, X., et al. (2018). DNA Conformation Induces Adaptable Binding by Tandem Zinc Finger Proteins. Cell *173*, 221–233.e12.

Persikov, A.V., and Singh, M. (2014). De novo prediction of DNA-binding specificities for Cys2His2 zinc finger proteins. Nucleic Acids Res. *42*, 97–108.

Pollen, A.A., Bhaduri, A., Andrews, M.G., Nowakowski, T.J., Meyerson, O.S., Mostajo-Radji, M.A., Di Lullo, E., Alvarado, B., Bedolli, M., Dougherty, M.L., et al. (2019). Establishing Cerebral Organoids as Models of Human-Specific Brain Evolution. Cell *176*, 743–756.e17.

Pontis, J., Planet, E., Offner, S., Turelli, P., Duc, J., Coudray, A., Theunissen, T.W., Jaenisch, R., and Trono, D. (2019). Hominoid-Specific Transposable Elements and KZFPs Facilitate Human Embryonic Genome Activation and Control Transcription in Naive Human ESCs. Cell Stem Cell *24*, 724–735.e5.

Prado-Martinez, J., Sudmant, P.H., Kidd, J.M., Li, H., Kelley, J.L., Lorente-Galdos, B., Veeramah, K.R., Woerner, A.E., O'Connor, T.D., Santpere, G., et al. (2013). Great ape genetic diversity and population history. Nature *499*, 471–475.

Prescott, S.L., Srinivasan, R., Marchetto, M.C., Grishina, I., Narvaiza, I., Selleri, L., Gage, F.H., Swigut, T., and Wysocka, J. (2015). Enhancer divergence and cis-regulatory evolution in the human and chimp neural crest. Cell *163*, 68–83.

Prüfer, K., de Filippo, C., Grote, S., Mafessoni, F., Korlević, P., Hajdinjak, M., Vernot, B., Skov, L., Hsieh, P., Peyrégne, S., et al. (2017). A high-coverage Neandertal genome from Vindija Cave in Croatia. Science *358*, 655–658.

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

Rakic, P. (2009). Evolution of the neocortex: a perspective from developmental biology. Nat. Rev. Neurosci. *10*, 724–735.

Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. *42*, W187–W191.

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

Rooney, J.P., Ryde, I.T., Sanders, L.H., Howlett, E.H., Colton, M.D., Germ, K.E., Mayer, G.D., Greenamyre, J.T., and Meyer, J.N. (2015). PCR based determination of mitochondrial DNA copy number in multiple species. Methods Mol Biol *1241*, 34–38. https://doi.org/10.1007/978-1-4939-1875-1_3.

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.

Skene, P.J., Henikoff, J.G., and Henikoff, S. (2018). Targeted in situ genome-wide profiling with high efficiency for low cell numbers. Nat. Protoc. *13*, 1006–1019.

Södersten, E., Toskas, K., Rraklli, V., Tiklova, K., Björklund, A.K., Ringnér, M., Perlmann, T., and Holmberg, J. (2018). A comprehensive map coupling histone modifications with gene regulation in adult dopaminergic and serotonergic neurons. Nat. Commun. *9*, 1226.

Sousa, A.M.M., Meyer, K.A., Santpere, G., Gulden, F.O., and Sestan, N. (2017). Evolution of the Human Nervous System Function, Structure, and Development. Cell *170*, 226–247.

Sripathy, S.P., Stevens, J., and Schultz, D.C. (2006). The KAP1 corepressor functions to coordinate the assembly of de novo HP1-demarcated microenvironments of heterochromatin required for KRAB zinc finger protein-mediated transcriptional repression. Mol. Cell. Biol. *26*, 8623–8638.

Sudmant, P.H., Kitzman, J.O., Antonacci, F., Alkan, C., Malig, M., Tsalenko, A., Sampas, N., Bruhn, L., Shendure, J., and Eichler, E.E.; 1000 Genomes Project (2010). Diversity of human copy number variation and multicopy genes. Science *330*, 641–646.

Suzuki, I.K., Gacquer, D., Van Heurck, R., Kumar, D., Wojno, M., Bilheu, A., Herpoel, A., Lambert, N., Cheron, J., Polleux, F., et al. (2018). Human-Specific NOTCH2NL Genes Expand Cortical Neurogenesis through Delta/Notch Regulation. Cell *173*, 1370–1384.e16.

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

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res *22*, 4573–4680. https://doi.org/10.1093/nar/22.22.4673.

Trono, D. (2015). Transposable Elements, Polydactyl Proteins, and the Genesis of Human-Specific Transcription Networks. Cold Spring Harb. Symp. Quant. Biol. *80*, 281–288.

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.

Wagih, O. (2017). ggseqlogo: a versatile R package for drawing sequence logos. Bioinformatics *33*, 3645–3647.

Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., and Barton, G.J. (2009). Jalview Version 2–a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*, 1189–1191. https://doi.org/10.1093/bioinformatics/btp033.

Wolf, G., Yang, P., Füchtbauer, A.C., Füchtbauer, E.M., Silva, A.M., Park, C., Wu, W., Nielsen, A.L., Pedersen, F.S., and Macfarlan, T.S. (2015). The KRAB zinc finger protein ZFP809 is required to initiate epigenetic silencing of endogenous retroviruses. Genes Dev. *29*, 538–554.

Wunderlich, S., Kircher, M., Vieth, B., Haase, A., Merkert, S., Beier, J., Göhring, G., Glage, S., Schambach, A., Curnow, E.C., et al. (2014). Primate iPS cells as tools for evolutionary analyses. Stem Cell Res. (Amst.) *12*, 622–629.

Yates, A.D., Achuthan, P., Akanni, W., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M.R., Armean, I.M., Azov, A.G., Bennett, R., et al. (2020). Ensembl 2020. Nucleic Acids Res *48*, D682–D688. https://doi.org/10.1093/nar/gkz966.

Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS *16*, 284–287.

Zerbino, D.R., Achuthan, P., Akanni, W., Amode, M.R., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., Girón, C.G., et al. (2018). Ensembl 2018. Nucleic Acids Res. *46* (D1), D754–D761.

Zhang, Y., Li, T., Preissl, S., Amaral, M.L., Grinstein, J.D., Farah, E.N., Destici, E., Qiu, Y., Hu, R., Lee, A.Y., et al. (2019). Transcriptionally active HERV-H retrotransposons demarcate topologically associating domains in human pluripotent stem cells. Nat. Genet. *51*, 1380–1388.

Zheng, G.X.Y., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., et al. (2017). Massively parallel digital transcriptional profiling of single cells. Nat Commun *8*, 14049.

Zufferey, R., Nagy, D., Mandel, R.J., Naldini, L., and Trono, D. (1997). Multiply attenuated lentiviral vector achieves efficient gene delivery in vivo. Nat. Biotechnol. *15*, 871–875.

Please cite this article in press as: Johansson et al., A *cis*-acting structural variation at the *ZNF558* locus controls a gene regulatory network in human brain development, Cell Stem Cell (2021), https://doi.org/10.1016/j.stem.2021.09.008

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
rabbit anti-FOXG1, polyclonal	Abcam	Cat# ab18259, RRID:AB_732415
rabbit anti-Nanog, polyclonal	Abcam	Cat# ab21624, RRID:AB_446437
rabbit anti-H3K9me3, polyclonal	Abcam	Cat# ab8898, RRID:AB_306848
rabbit anti-H3K27me3, monoclonal	Cell Signaling Technology	Cat# 9733, RRID:AB_2616029
rabbit anti-H3K4me3, polyclonal	Active Motif	Cat# 39159, RRID:AB_2615077
goat anti-rabbit IgG, polyclonal	Abcam	Cat#: ab97047 RRID:AB_10681025
OxPhos Human WB Antibody Cocktail	ThermoFisher	Cat# 45-8199, RRID:AB_2533836)
mouse anti-beta actin, monoclonal	Cell Signaling Technology	Cat# 3700, RRID:AB_2242334)
rabbit anti-PAX6, polyclonal	BioLegend	Cat# 901301, RRID:AB_2565003)
rat anti-ZO1, monoclonal	Novus	NB110-68140, RRID:AB_1111431
Chemicals, peptides, and recombinant proteins		
Ascorbic Acid	Sigma	A5960
B27 (-Vit A) supplement	GIBCO(ThermoFisher)	12587010
B27 Supplement (+Vit. A)	GIBCO(ThermoFisher)	17504044
Beta-mercaptoethanol	GIBCO(ThermoFisher)	31350010
cAMP	Sigma	D0627
DMEM/F-12	GIBCO(ThermoFisher)	21331020
Draq7	BD Biosciences	564904
Heparin	Sigma	H3149
Insulin	Sigma	I9278
Knockout replacement serum (KSR)	GIBCO(ThermoFisher)	10828010
L-Glutamine	GIBCO(ThermoFisher)	25030032
Laminin-111	Biolamina	LN111
Laminin-521	Biolamina	LN521
mTeSR1	StemCell Technologies, Inc.	85850
N2 supplement	GIBCO(ThermoFisher)	17502048
Neurobasal	GIBCO(ThermoFisher)	21103049
Noggin	Miltenyi Biotec	130-103-456
Non-Essential Amino Acids	GIBCO(ThermoFisher)	11140
Penicillin-Streptomycin	GIBCO(ThermoFisher)	15140122
rhBDNF	R&D	248-BD
rhGDNF	R&D	212-GD
Roche cOmplete protease inhibitor	Sigma Aldrich	11697498001
Rock-inhibitor, Y27632	Miltenyi Biotec	130-106-538
SB431542	Axon	1661
StemMACS iPS-Brew XF, human	Miltenyi biotec	130-104-368
StemPro Accutase	GIBCO(ThermoFisher)	n/a
Critical commercial assays		
TruSeq Stranded mRNA LP (48 Spl)	Illumina	20020594
Hyperprep kit (KAPA Biosystems)	Roche	07962347001
ConA-coated magnetic beads (BioMag®Plus Concanavalin A	Bangs Laboratories	BP531
Chromium Single Cell 3′ Library	10X Genomics	PN-120233

(*Continued on next page*)

Please cite this article in press as: Johansson et al., A *cis*-acting structural variation at the *ZNF558* locus controls a gene regulatory network in human brain development, Cell Stem Cell (2021), https://doi.org/10.1016/j.stem.2021.09.008

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Raw and processed data for the single nuclei of organoids and bulk RNaseq of human and chimpanzee cell lines, as well as the cut&run of H3K9me3, H3K4me3, H3K27me3 of them, along with the CRISPRi of ZNF558 have been deposited at GEO under the accession number GSE182224	GEO: GSE182224 SRA: SRP332741	n/a
Code for data analysis and visualization can be found on github	https://github.com/Jakobsson-Lab/ZNF558_CellStemCell_2021	n/a
Experimental models: Cell lines		
Human iPSC line (HS1), female	RIKEN	RBRC-HPS0328 606A1 (RRID:CVCL_DQ11)
Human iPSC line (HS2), male	RIKEN	RBRC-HPS0360 648A1 (RRID:CVCL_DQ30)
Chimpanzee iPSC line (PT1), female	Mora-Bermúdez et al., 2016	Sandra A
Chimpanzee iPSC line (PT2), female	Marchetto et al., 2013	PR00818 PTCL-5 (RRID:CVCL_2Z83)
Oligonucleotides		
Primers for qRT-PCR; see Table S1	This paper	n/a
Primers for genomic PCR: see Table S1	This paper	n/a
PCR: MtDNA; FP CACCCAAGAACAGGGTTTGT	Rooney et al., 2015	n/a
PCR: MtDNA; RP TGGCCATGGGTATGTTGTTA	Rooney et al., 2015	n/a
PCR NucDNA; FP TGCTGTCTCCATGTTTGATGTATCT	Rooney et al., 2015	n/a
PCR NucDNA: RP TCTCTGCTCCCCACCTCTAAGT	Rooney et al., 2015	n/a
sgRNA; ZNF558-g2 GCCAAAAGCGCCGACTCGCG	This paper	n/a
sgRNA: ZNF558-g3 AGTCGGCGCTTTTGGCCCCG	This paper	n/a
sgRNA: LacZ; TGCGAATACGCCCACGCGAT	This paper	n/a
sgRNA: VNTR-g1 GCTGCCCTGAGATATGTGTG	This paper	n/a
sgRNA: VNTR-g2 TACTGGAATGGGTAGGAATG	This paper	n/a
sgRNA: VNTR-g3 CCAGGAGCTGCACATTGAGG	This paper	n/a
Recombinant DNA		
Plasmid: pLV hU6-sgRNA hUbC-dCas9-KRAB-T2a-GFP	Addgene/ C. Gersbach	RRID: Addgene_71237
Plasmid: pEXPpSIN-TRE-puroW ZNF558 −3xHA	Imbeault et al., 2017	n/a
Software and algorithms		
STAR Aligner	Dobin et al., 2013	10.1093/bioinformatics/bts635
Bowtie2	Langmead and Salzberg, 2012	10.1038/nmeth.1923
TEtranscripts	Jin et al., 2015	10.1093/bioinformatics/btv422
Cellranger	Zheng et al., 2017	10.1038/ncomms14049
UCSC Liftover	Kent et al., 2010	10.1093/bioinformatics/btq351
DESeq2	Love et al., 2014	10.1186/s13059-014-0550-8
Seurat	Butler et al., 2018	10.1038/nbt.4096
Deeptools	Ramírez et al., 2014	10.1093/nar/gkw257
ClustalW	Thompson et al., 1994	10.1093/nar/22.22.4673
JalView	Waterhouse et al., 2009	10.1093/bioinformatics/btp033
Ensembl	Yates et al., 2020	https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkz966/5613682
HOMER	Heinz et al., 2010	10.1016/j.molcel.2010.05.004
analysis_Shuffle_bed.pl	Lynch et al., 2015	10.1016/j.celrep.2014.12.052
BEDtools	Quinlan and Hall, 2010	10.1093/bioinformatics/btq033
MEME	Bailey et al., 2015	10.1093/nar/gkv416
clusterProfiler	Yu et al., 2012	10.1089/omi.2011.0118
FeatureCounts	Liao et al., 2014	10.1093/bioinformatics/btt656
SAMtools	Li et al., 2009	10.1093/bioinformatics/btp352

Please cite this article in press as: Johansson et al., A *cis*-acting structural variation at the *ZNF558* locus controls a gene regulatory network in human brain development, Cell Stem Cell (2021), https://doi.org/10.1016/j.stem.2021.09.008

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Johan Jakobsson (johan.jakobsson@med.lu.se).

Materials availability

Plasmids generated in this study are available upon request.

Data and code availability

There are no restrictions in data availability. The RNA and DNA sequencing data presented in this study have been deposited at GEO: GSE182224 and are publicly available. Accession numbers are listed in the Key resources table. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

This paper analyzes existing, publicly available data. These accession numbers for the datasets are listed in the Key resources table.

All original code has been deposited at GitHub and is publicly available as of the date of publication. DOIs are listed in the Key resources table.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Induced pluripotent stem cells (iPSCs)

We used two human iPSC lines generated by mRNA transfection (RBRC-HPS0328 606A1 and RBRC-HPS0360 648A1, both from RIKEN; from here on referred to as HS1 and HS2, respectively). We used two chimpanzee iPSC lines: one generated by mRNA transfection (Sandra A, herein referred to as PT1) and the other with viral vector transduction (PR00818 PTCL-5, herein referred to as PT2) (Marchetto et al., 2013; Mora-Bermúdez et al., 2016). iPSCs were maintained on LN521-coated (0.7 μg/cm^2; Biolamina) Nunc multidishes in iPS media (StemMACS iPS-Brew XF and 0.5% penicillin/streptomycin (GIBCO)). Cells were passaged 1:2-1:6 every 2-4 days by being rinsed once with DPBS (GIBCO) and dissociated using 0.5 mM EDTA (75 μl/cm^2; GIBCO) at 37°C for 7 minutes. Following incubation, EDTA was carefully aspirated from the well and the cells were washed off from the dish using washing medium (9.5 mL DMEM/F-12 (31330-038; GIBCO) and 0.5 mL knockout serum replacement (GIBCO)). The cells were then centrifuged at 400*g* for 5 minutes and resuspended in iPS brew medium supplemented with 10 μM Y27632 (Rock inhibitor; Miltenyi) for expansion. The media was changed daily (Grassi et al., 2019; Nolbrant et al., 2017).

METHOD DETAILS

Differentiation into forebrain neural progenitors (fbNPCs)

iPSCs were grown to approximately 70%–90% confluency and were then dissociated as usual for passaging. After centrifugation, the cells were resuspended in N2 medium (1:1 DMEM/F-12 (21331020; GIBCO) and Neurobasal (21103049; GIBCO) supplemented with 1% N2 (GIBCO), 2 mM L-glutamine (GIBCO), and 0.2% penicillin/streptomycin). The cells were manually counted twice and plated at a density of 10,000 cells/cm^2 in 250 μl medium/cm^2 on LN111 Nunc Δ multidishes (1.14 μg/cm^2; Biolamina). 10 μM SB431542 (Axon) and 100 ng/ml noggin (Miltenyi) for dual SMAD inhibition, and 10 μM Y27632 was added to the medium. The medium was changed every 2-3 days (N2 medium with SB431542 and noggin) until day 9 of differentiation, when N2 medium without SMAD inhibitors was used. On day 11, the cells were replated by washing twice with DPBS followed by adding StemPro accutase (75 μl/cm^2; GIBCO) for 10-20 minutes at 37°C. The dissociated cells were washed off with 10 mL wash medium, centrifuged for 5 minutes at 400*g* and resuspended in B27 medium (Neurobasal supplemented with 1% B27 without vitamin A (GIBCO), 2 mM L-glutamine and 0.2% penicillin/streptomycin Y27632 (10 μM), BDNF (20 ng/ml; R&D), and L-ascorbic acid (0.2 mM; Sigma). The cells were counted and replated at 800,000 cells/cm^2 on LN111- coated plastic in B27 medium (600 μl medium/cm^2). The cells were kept in the same medium until day 14, after which new B27 medium was added (Grassi et al., 2019).

Immunocytochemistry

The cells were washed once with DPBS and fixed for 15 minutes with 4% paraformaldehyde (Merck Millipore), followed by three rinses with DPBS. The fixed cells were then pre-blocked for a minimum of 30 minutes in a blocking solution of KPBS with 0.25% Triton X-100 (Fisher Scientific) and 5% donkey serum. The primary antibody (rabbit anti-FOXG1, 1:50 dilution, Abcam, RRID: AB_732415 and anti-NANOG, 1:100 dilution, Abcam, RRID: AB_446437) was added and incubated overnight. On the following day, the cells were washed twice with KPBS. The secondary antibody (donkey anti-rabbit Cy3; 1:200; Jackson Lab) was added with DAPI (1:1000; Sigma-Aldrich) as a nuclear counterstain and incubated at room temperature for one hour, followed by 2-3 rinses with KPBS. The cells were visualized on a Leica microscope (model DMI6000 B), and images were cropped and adjusted in Adobe Photoshop CC.

Bulk RNA sequencing

On the day of harvest, the cells were washed once with PBS and lysed with 350 μl RLT buffer with 1% mercaptoethanol (Thermo Fisher). RNA was extracted using the RNeasy mini kit (QIAGEN) according to manufacturer's protocol. The quality and concentration of the RNA was analyzed using 2100 Bioanalyzer (RNA nano; Agilent) and Qubit (RNA HS assay kit). Libraries for sequencing were prepared using the TruSeq RNA Library Prep kit v2 (Illumina) and again quality-controlled using the Bioanalyzer (high-sensitivity DNA assay) and Qubit (dsDNA HS assay kit). Finally, the libraries were sequenced using an Illumina NextSeq 500, 150-bp paired-end reads (300 cycles).

RNA-sequencing samples were mapped to the human reference genome (GRCh38) and the chimp reference genome (Clint_PTRvs2/PanTro6) using STAR aligner v2.5.0 (Dobin et al., 2013), allowing 0.03 mismatches per base (–outFilterMismatchNoverLmax 0.03), using Gencode v27 (Harrow et al., 2012) gene models for splice junction annotation. For multimapping, STAR was run on default settings, retaining reads that map at up to 10 loci. For unique mapping, STAR was run with–outFilterMultimapNmax 1. Gene counts were quantified with the Subread package FeatureCounts (Liao et al., 2014), counting reads overlapping Gencode (v27) gene annotations. For mapping and quantification of chimp samples, the Gencode annotation for GRCh38 were lifted to the panTro6 reference genome using the UCSC LiftOver tool. Normalization and differential expression analysis was performed with the R package DESeq2 (Love et al., 2014).

To analyze TE expression, we mapped the samples once again using STAR aligner v2.6.0 (Dobin et al., 2013), allowing reads to map 100 times (–outFilterMultimapNmax 100) with 200 anchors (–winAnchorMultimapNmax 200). The bam files were input to TEtranscripts version 2.0.3 (Jin et al., 2015) in multi-mode (–mode multi) using gencode annotation GRCh38 as the gene annotation (–GTF), as well as the curated GTF file of TEs provided by TEtranscripts' authors (–TE). Differential expression analysis was performed using DESeq2 (Love et al., 2014).

Single-cell RNA sequencing

HS1 and PT1 were differentiated to day 14, washed twice with DPBS, and dissociated with Accutase for approximately 10 minutes, followed by centrifugation at 400 × g for 5 minutes in wash medium. All pipetting was done very gently to avoid cell death. The pellets were resuspended in 1 mL PBS with 0.04% BSA (Sigma) and filtered through 100 μm cell strainers (Falcon) twice. Cells were resuspended in order to yield a concentration of approximately 1000 cells/μl. The single-cell libraries were prepared with Chromium Single Cell A chip kit and Chromium Single Cell 3′ Library & Gel Bead kit v2 (10 × Genomics), quality controlled, and quantified using Qubit ds DNA HS and Bioanalyzer High Sensitivity DNA Assay prior to sequencing. The samples were sequenced for 26 cycles on read 1 and 98 cycles on read 2 using the Illumina NextSeq 500. Raw single-cell RNA-seq data was processed using the Cell Ranger software suite. Raw base call files were converted using cellranger mkfastq before aligning, filtering, barcode count, and UMI counting was performed using cellranger count. Count matrices were further analyzed using the Seurat R package (Butler et al., 2018). The data were filtered on number of genes detected in each cell (2000-6000 genes/cell in human and 2000-5000 genes/cell in chimpanzee were kept for further analysis), and only cells with max 0.05% mitochondrial gene reads. 4553 HS1 cells were sequenced and 4355 were used in the analysis, 5674 PT1 cells were sequenced and 3620 were kept after filtering. The data was further log2-normalized and scaled to total expression in each cell, before further scaling cells on number of UMIs detected and percentage of mitochondrial gene count. PCA was run on variable genes defined using the FindVariableGenes function. tSNE was run using PCA dimensionality reduction.

CUT&RUN

We followed the protocol detailed by the Henikoff lab (Skene et al., 2018). Briefly, 100,000 cells were washed twice (20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM spermidine, 1x Roche cOmplete protease inhibitors) and attached to 10 ConA-coated magnetic beads (Bangs Laboratories) that had been pre-activated in binding buffer (20 mM HEPES pH 7.9, 10 mM KCl, 1 mM $CaCl_2$, 1 mM $MnCl_2$). Bead-bound cells were resuspended in 50 μL buffer (20 mM HEPES pH 7.5, 0.15 M NaCl, 0.5 mM Spermidine, 1x Roche complete protease inhibitors, 0.02% w/v digitonin, 2 mM EDTA) containing primary antibody (rabbit anti-H3K9me3, Abcam ab8898, RRID:AB_306848; rabbit anti-H3K27me3, Cell Signaling Technology C36B11 #9733, RRID:AB_2616029; rabbit anti-H3K4me3 Active Motif 39159, RRID:AB_2615077; or goat anti-rabbit IgG, Abcam ab97047, RRID:AB_10681025) at 1:50 dilution and incubated at 4°C overnight with gentle shaking. Beads were washed thoroughly with digitonin buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM Spermidine, 1x Roche cOmplete protease inhibitors, 0.02% digitonin). After the final wash, pA-MNase (a generous gift from Steve Henikoff) was added in digitonin buffer and incubated with the cells at 4°C for 1 h. Bead-bound cells were washed twice, resuspended in 100 μL digitonin buffer, and chilled to 0-2°C. Genome cleavage was stimulated by addition of 2 mM $CaCl_2$ at 0°C for 30 min. The reaction was quenched by addition of 100 μL 2x stop buffer (0.35 M NaCl, 20 mM EDTA, 4 mM EGTA, 0.02% digitonin, 50 ng/μL glycogen, 50 ng/μL RNase A, 10 fg/μL yeast spike-in DNA (a generous gift from Steve Henikoff)) and vortexing. After 10 min incubation at 37°C to release genomic fragments, cells and beads were pelleted by centrifugation (16,000 g, 5 min, 4°C) and fragments from the supernatant purified. Illumina sequencing libraries were prepared using the Hyperprep kit (KAPA) with unique dual-indexed adapters (KAPA), pooled and sequenced on a Nextseq500 instrument (Illumina). Paired-end reads (2x75) were aligned to the human and yeast genomes (hg38 and R64-1-1 respectively) using bowtie2 (–local –very-sensitive-local –no-mixed –no-discordant –phred33 -I 10 -X 700) and converted to bam files with samtools (Langmead and Salzberg,

2012; Li et al., 2009). Normalized bigwig coverage tracks were made with bamCoverage (deepTools) (Ramírez et al., 2014), with a scaling factor accounting for the number of reads arising from the spike-in yeast DNA (10^4/aligned yeast read number). Tracks were displayed in IGV.

Evolutionary analysis of ZNF558

Phylogenetic tree, pairwise sequence alignment scores and dN/dS ratios of ZNF558 orthologs was downloaded from the Ensembl (Zerbino et al., 2018). Orthologs section for ZNF558 in nexus format and FigTree was used for visualization (http://tree.bio.ed.ac.uk/software/figtree/). For analysis of DNA binding residues of the zinc finger domain of ZNF558 and orthologs, the ClustalW multiple sequence alignment from Ensembl orthologs was visualized in JalView. DNA contacting residues were defined the four amino acids in position −1, −4, −5 and −7 relative to the first histidine residue of the two zinc-coordinating histidine residues. For analysis of ZNF558 TE binding, TE coordinates were downloaded from RepeatMasker.

To check the conservation of the ZNF558 binding site in SPATA18, we used a set of 46 eutherian mammals at the comparative genomics section of Ensembl's webtool. Visualization was done using ggmsa (https://github.com/YuLab-SMU/ggmsa) and geom_logo (Wagih, 2017).

Using the resulting peaks from the *ChIP-exo* of ZNF558 with a score higher than 20 (see *ChIP-exo analysis* section), we classified ZNF558 binding sites between genes and TEs using HOMER's annotatePeaks.pl. Only intronic and intragenic TEs are reported (retrotransposons as well as simple repeats and DNA transposons). A peak was registered to be at a gene when the detailed annotation would be reported as 3′, 5′, exon, non-coding, promoter-TSS, TTS, or CpG.

To further explore the significance of ZNF558 binding sites at the different classes, families, and subfamilies of TEs, we used a script implemented by Kapusta et al. (2013), TE-analysis_Shuffle_bed.pl (version 4.4.2)

(https://github.com/4ureliek/TEanalysis/blob/master/TE-analysis_Shuffle_bed.pl).This script uses bedtools (–shuffle bed) (version 2.27.1) (Quinlan and Hall, 2010) to shuffle the genomic locations of TEs from a given annotation file (–query) (intronic and intergenic TEs) and intersects it with the provided peaks (–feat) (ZNF558 *ChIP-exo* peaks with score > 20). It repeats this operation to calculate the number of binding sites of these peaks binding to a particular class, family, or subfamily of TEs as expected by chance.

To obtain the annotation file for intronic or intragenic TEs, we used bedtools intersect to perform an inner join between introns or intergenic regions as annotated at HOMER's hg38.basic.annotation, and repeatmasker original annotation file for hg38 (version 4.0.5).

For the maximum values of chromosomes (–range), we used UCSC hg38.chrom.sizes and for the excluded areas (–excl) we ran TE-analysis_Shuffle_bed.pl setting up the flag–dogaps and providing GRCh38.p13 fasta file.

ChIP-exo analysis

ZNF558 and ZNF557 ChIP-exo data was downloaded along with input control from GSE78099 (Imbeault et al., 2017). Raw ChIP-exo reads were quality controlled with FastQC (Babraham) and aligned to the human reference genome (GRCh38) using bowtie2 with–sensitive-local (Langmead and Salzberg, 2012). Unique reads were filtered by retaining only alignments with MAPQ > 10 (samtools view -q 10) (Li et al., 2009). Peaks were called with HOMER (Heinz et al., 2010) and those with score > 20 retained. Visualization of ChIP signals was done in deepTools using the computeMatrix and plotHeatmap modules (Ramírez et al., 2014). ZNF558 motif analysis was done with MEME (search for one motif, length 18-30 nt) (Bailey et al., 2015). Prediction of the ZNF558 DNA binding sequence was done as detailed in Persikov and Singh (2014).

CRISPRi

In order to silence the transcription of ZNF558 we used the catalytically inactive Cas9 (deadCas9) fused to the transcriptional repressor KRAB. Single guide sequences were designed to recognize DNA regions just down-stream of the transcription start site (TSS) according to the GPP Portal (Broad Institute). See Key resources table for guide RNA sequences. For the induction of repressive marks on the human VNTR we chose three guide RNAs from the UCSC Gene Browser CRISPR target track with high on-target specificity and only found within the VNTR region and avoiding transposable elements (RepeatMasker). See Key resources table for guide RNA sequences. The guides were inserted into a deadCas9-KRAB-T2A-GFP lentiviral backbone containing both the guide RNA under the U6 promoter and dead-Cas9-KRAB and GFP under the Ubiquitin C promoter (pLV hU6-sgRNA hUbC-dCas9-KRAB-T2a-GFP, a gift from Charles Gersbach, Addgene plasmid #71237 RRID:Addgene_71237). The guides were inserted into the backbone using annealed oligos and the BsmBI cloning site. Lentiviruses were produced as described below yielding titers between 4.9E+08 and 9.3E+09. Control virus with a gRNA sequence not present in the human genome (LacZ) was also produced and used in all experiments. All lentiviral vectors were used with an MOI between 5 and 20. Cells were FACS sorted as described above and knock-down efficiency was validated using standard quantitative real-time RT-PCR techniques.

ZNF558-g2 GCCAAAAGCGCCGACTCGCG; ZNF558-g3 AGTCGGCGCTTTTGGCCCCG; LacZ TGCGAATACGCCCACGCGAT; VNTR-g1 GCTGCCCTGAGATATGTGTG; VNTR-g2 TACTGGAATGGGTAGGAATG; VNTR-g3 CCAGGAGCTGCACATTGAGG

GFP+ cell isolation; At day 14, cells were detached with Accutase, resuspended in differentiation media with Rock inhibitor (10μM, Miltenyi) and Draq7 (BD Bioscience), and strained (70μm, BD Bioscience). Gating parameters were determined by side and forward scatter to eliminate debris and aggregated cells. The GFP-positive gates were set using untransduced cells. The sorting gates and

Please cite this article in press as: Johansson et al., A *cis*-acting structural variation at the *ZNF558* locus controls a gene regulatory network in human brain development, Cell Stem Cell (2021), https://doi.org/10.1016/j.stem.2021.09.008

strategies were validated via reanalysis of sorted cells (> 95% purity cut-off). GFP-positive/Draq7-negative single cells (100,000-200,000 cells/pellet) were collected, spun down (400*g*, 10 min) and snap frozen on dry ice. Cell pellets were kept at −80°C until RNA or DNA was isolated.

Lentiviral production

Lentiviral vectors were produced according to Zufferey et al. (1997) and were in titers of 10^8 - 10^9 TU/ml as determined by qRT-PCR. Briefly HEK293T cells were grown to a confluency of 70 – 90% at the day of transfection for lentiviral production. Third-generation packaging and envelop vectors (pMDL, psRev, and pMD2G) together with Polyethyleneimine (PEI Polysciences PN 23966, in DPBS (GIBCO) were used. The lentivirus was harvested two days after transfection The supernatant was then collected, filtered and centrifuged at 25,000*g* for 1.5 hours at 4°C. The supernatant was removed from the tubes and the virus was resuspended in PBS and left at 4°C. The resulting lentivirus was aliquoted and stored at −80°C.

ZNF558 overexpression in chimpanzee cells

Chimpanzee cells (PT1) was transduced with a lentiviral construct overexpressing ZNF558 and the puromycin resistance cassette. An identical virus without the ZNF588 insert was used as a control. Cells were split and transduced and allowed to recover for 2-3 days, followed by puromycin selection. The cells were in puromycin selection for more than 5 days, then expanded briefly without puromycin and frozen. Cells were thawed and exposed to an additional 1-2 day of puromycin selection prior to start of differentiation. Puromycin was added to the differentiation media at day 0, day 4, day 7 and day 11.

qRT-PCR

Total RNA was first extracted according to the supplier's recommendations using the mini RNeasy kit (QIAGEN). cDNA was generated using the Maxima First Strand cDNA Synthesis Kit (Thermo Scientific) and analyzed with SYBR Green I master (Roche) on a LightCycler 480 (Roche). Data are represented with the ΔΔCt method normalized to the housekeeping genes B-actin, GAPDH and HPRT1. See Primer sequences are listed in Table S1. Expression levels were confirmed using one additional primer pair (data not shown).

Copy number analysis and mitochondrial analysis

Genomic and mitochondrial DNA was extracted using DNeasy Blood and Tissue Kits (QIAGEN). All primers were used together with LightCycler 480 SYBR Green I Master (Roche). For mitochondrial analysis we used previously published primers used for mitochondrial copy number assessment (MtDNA; FP CACCCAAGAACAGGGTTTGT, RP TGGCCATGGGTATGTTGTTA, NucDNA FP TGCTGTCTCCATGTTTGATGTATCT, RP TCTCTGCTCCCCACCTCTAAGT (Rooney et al., 2015).

For VNTR unit copy number assessment we designed primers for different parts of the VNTR repeat unit that was not found in other parts of the genome. We used albumin as a positive control with the albumin copy number set as 2 (see Table S1).

Western blotting

Cells were washed with cold PBS (pH 7.4) and lysed with RIPA buffer containing complete protease inhibitor cocktail (Thermo Scientific). Lysates were centrifuged at 14,000 g for 15 min at 4°C, the supernatants were collected and stored at –20°C. Protein concentration was quantified with Pierce BCA Protein Assay Kit (Thermo Scientific). 20–30 μg of the protein extracts were denatured with Laemmli buffer (SDS 10%), separated on 4%–15% SDS-PAGE (TGX Stain-Free, BioRad) for 40-80min at 100-150V and electrotransferred onto LF-PVDF membranes for 3-6min in a Trans-Blot Turbo equipment (BioRad). Membranes were blocked with 5% fat-free BSA in TBS-T buffer (50 mM Tris-HCl, pH 7.6, 150 mM NaCl, 0.1% Tween 20) for 60 min and incubated overnight at 4°C with primary antibodies (OxPhos Human WB Antibody Cocktail, ThermoFisher, code: 45-8199 (dilution1:400) or β-actin antibody, Cell Signaling Cat# 3700, Cat# 3700, RRID:AB_2242334) (dilution 1:1000)). After rinsing with TBS-T, membranes were incubated with an HRP-conjugated anti-mouse IgG secondary antibody for 60 min. Immunoblots were visualized with Western ECL Substrate (BioRad) and imaged in a ChemiDoc XRS equipment (BioRad). Band density for ATP5A proteins were normalized against β-actin protein content and analysis performed by the Image Lab software (BioRad). Data are shown as mean ± SEM. Unpaired Student's t test was used to compare the ATP5A content in control versus Z588 guides conditions. Statistical significance was set at $p < 0.05$. Prism 8.0 (Graph Pad) was used for statistical analysis and graphics generation.

Human cerebral organoid culture

To generate the human cerebral-like organoids we used three hIPSC6-derived lines, one control line (guide against LacZ) and two ZNF558 KD lines (g2 and g3). We used a protocol based by the one by Lancaster et al. (2013) with minor modifications. Starting from the plated cells, they were first washed once with DPBS−/−, then detached from the wells with 500 μL of Accutase with an incubation at 37°C for 5 minutes. The cells were then transferred in 5 mL of washing medium (5% KSR in DMEM F12 +/+) and centrifuged for 5 minutes at 400 x g (RT; acceleration/deceleration 8). After removing the supernatant, the pellet was resuspended in 1-2 mL of mTeSR1 (StemCell Technologies, Inc.) enriched with ROCK Inhibitor (RI) 10 μM (1:1000 from stock concentration 10 mM). The cells in the cell suspension were then counted, and 8000 cells/well were plated in a 96-wells plate (Costar, Ultra Low Attachment, round bottom; REF 7007) with 250 μL of mTeSR1 and RI 10 μM. This is considered day −5 of the differentiation of the iPSCs-derived hFB organoids. At day −3 and −1 the medium was changed (150 μL and 200 μL of mTeSR1, respectively). At day 0 the cells are fed with

Neural Induction Medium (NIM; DMEM/F12 media, N2 Supplement (1:100), L-Glutamine (2mM), Penicillin/Streptomycin (1:500), Non-Essential Amino acids (1:100) and Heparin (2ug/ml).) enriched with 3% KSR. At day 2, 4, and 6, the organoids were fed with NIM with no added KSR.

At day 7 "cavities" for the organoids embedding were prepared by putting parafilm on an empty tip holder (with sterile surface facing up) and pressing every hole to form a little basin. The structure was sterilized with a 30-minutes UV cycle and left inside the hood until the next day. Matrigel was left to thaw overnight at 4°C. On day 8 the organoids were embedded: the parafilm was cut in half, then each organoid was transferred in one "cavity" with a cut P1000-tip; the media surrounding the organoid was removed, trying to place it in the center of the "cavity." The organoids were then covered with 30-50 μL of Matrigel (Corning) avoiding any bubble and incubated at 37°C for 25 minutes to allow the Matrigel to solidify. The organoids were then transferred in Corning REF 3471 6-wells plates with flat bottom containing 3 ml/well of Cortical Differentiation Medium (CDM; F12 Media (-Glut) (48.5%), Neurobasal (48.5%), N2 Supplement (1:200), B27 Supplement (-Vit.A, 1:100), L-Glutamine (2mM), Penicillin/Streptomycin (1:500), Non-Essential Amino acids (1:200), Beta MercaptoEtOH (50uM) and Insulin (2.5 ug/mL)). To detach the organoids from the parafilm, the latter was removed from the tip holder and the "dimples" were turned inside-out; drops of CDM then were used to let the organoids fall into the wells. In the end, in every well there were about 4 mL of medium.

On day 10 and 12 of the differentiation, the medium was changed exchanging 3 ml/well for 3 mL of fresh CDM. On day 15, 17, 19, 21 and 23, ~4 mL the medium was replaced with 4 mL of Improved Differentiation Medium + A (IDM, F12 Media (-Glut) (48.5%), Neurobasal (48.5%), N2 Supplement (1:200), B27 Supplement (+Vit.A, 1:50), L-Glutamine (2mM), Penicillin/Streptomycin (1:500), Non-Essential Amino acids (1:200), Beta MercaptoEtOH (50uM), Insulin (2.5 ug/mL) and Ascorbic Acid (400uM)). From day 25, the medium was changed every 3 days with 3-4 mL of Cortical Terminal Differentiation Medium (CTDM, F12 Media (-Glut) (48.5%), Neurobasal (48.5%), N2 Supplement (1:200), B27 Supplement (+Vit.A) – (1:50) 800uL, L-Glutamine (2mM), Penicillin/Streptomycin (1:500), Non-Essential Amino acids (1:200), Beta MercaptoEtOH (50uM), Insulin (2.5 ug/mL) and Ascorbic Acid (400uM), BDNF (10ng/uL), cAMP (200uM), GDNF (10ng/uL)).

All the diameter measurements of the organoids were taken with the Measure tool from the software GIMP 2.10. The chosen measuring unit was mm.

Organoid immunostaining

Organoids were fixed in 4% paraformaldehyde for 5 h at 4 °C followed by washing in PBS three times for 10 min and were left to sink in 30% sucrose overnight. Sucrose solution was replaced with 1:1 OCT:30% sucrose mixture for 6 h and then transferred to a cryomold and filled with OCT. The embedded tissue was frozen on dry ice and either cryosectioned at 20 μM or stored at −80°C. For immunohistochemistry, sections were washed in PBS1X for 10 min and then blocked and permeabilized in 0.3% Triton X-100 and 4% normal donkey serum in PBS1X. The primary antibodies used were rabbit anti-PAX6 (1:300; BioLegend, Cat# 3700, RRID:AB_2242334)) and rat anti-ZO1 NB110-68140, RRID:AB_1111431 (1:300; Novus)). After incubation with primary antibodies, the sections were incubated for 1 h with the appropriate secondary antibodies (Alexa Fluor 488, 594, 647 used at 1:400; Molecular Probes) and then mounted on gelatin-coated slides and coverslipped with PVA-DABCO containing DAPI (1:1000).

Single nuclei isolation and RNA-sequencing

Organoids (LacZ, g2 and g3) were collected for single nuclei sequencing at 2 and 4 months. At harvest, organoids were dissected out of the Matrigel, frozen on dry ice and stored at −80°C. In order to reduce bias in the downstream analysis due to organoid heterogeneity, they were frozen 5 by 5 per time point and batch. The nuclei isolation protocol has been described in detail elsewhere (Södersten et al., 2018). In brief, the organoids were thawed and homogenized in cold lysis buffer (0.32 M sucrose, 5 mM $CaCl_2$, 3 mM MgAc, 0.1 mM Na_2EDTA, 10 mM Tris-HCl, pH 8.0, 1 mM DTT) using a tissue douncer (Wheaton). The homogenate was carefully layered on top of a sucrose solution (1.8 M sucrose, 3 mM MgAc, 10 mM Tris-HCl, pH 8.0, and 1 mM DTT) before centrifugation at 30,000 × *g* for 2 hours and 15 min. Pelleted nuclei were softened for 10 min in 100 μL of nuclear storage buffer (15% sucrose, 10 mM Tris-HCl, pH 7.2, 70 mM KCl, and 2 mM $MgCl_2$) before resuspended in 400 μL of dilution buffer (10 mM Tris-HCl, pH 7.2, 70 mM KCl, and 2 mM $MgCl_2$) and run through a cell strainer (70 μm). Cells were run through the FACS (FACS Aria, BD Biosciences) at 4°C with low flowrate using a 100 μm nozzle (reanalysis showed > 99% purity), sorting 8,500 nuclei per sample which were then loaded onto 10X Genomics Single Cell 3' Chip along with the reverse transcription mastermix following the manufacturer's protocol for the Chromium Single Cell 3′ Library (10X Genomics, PN-120233) to generate single-cell gel beads in emulsion. cDNA amplification was done as per the guidelines from 10x Genomics and sequencing libraries were generated with unique sample indices (SI) for each sample. Libraries for samples were multiplexed and sequenced on a Novaseq using a 150-cycle kit using the recommended read length. Raw base calls were demultiplexed to obtained sample specific FastQ files and reads were aligned to GRCh38 genome using the Cell Ranger pipeline from 10x Genomics (Cellranger count v5) with default parameters (–include-introns was used for nuclei mapping). The resulting matrix files were used for further downstream analysis. Seurat (version 3.1.1 and R version 3.4) was utilized for bioinformatics analysis. Low quality cells and genes were filtered out based on fraction of total number of genes detected in each cell (±3 nmads). All samples were merged and batch effects were removed using the Harmony algorithm (Korsunsky et al., 2019). Cells with at least 500 detected genes were retained and the data was normalized to transcript copies per 10,000, and log-normalized to reduce sequencing depth variability. For visualization and clustering, manifolds were calculated using UMAP methods (*RunUMAP,* Seurat) and 20 precomputed principal components and the shared nearest neighbor algorithm modularity optimization-based clustering algorithm (FindClusters, Seurat) with a resolution of 0.1. For each cell type differential expression between LacZ and

knockdown samples was carried out using Seurat function FindMarkers (Wilcox test, p adjusted < 0.01). For SPATA18 expression, only cells with greater than zero expression for SPATA18 were included for the analysis. Cell cycle scores were computed using Seurat function CellCycleScoring. Gene ontology overrepresentation analysis was performed using the enrichGO function in the clusterProfiler package using genes that are expressed in this dataset as background (padj < 0.05) (Yu et al., 2012).

Copy number estimation for the VNTR locus near ZNF558 in human and non-human great ape samples

To investigate the copy number variation for the large VNTR downstream of *ZNF558* in human and non-human great ape lineages, we applied a read-depth based copy number genotyper (Sudmant et al., 2010) to a collection of 1,112 high-coverage genomes from several publicly-available resources (Bergström et al., 2020; Mafessoni et al., 2020; Mallick et al., 2016; Meyer et al., 2012; Prado-Martinez et al., 2013; Prüfer et al., 2017). In short, sequencing reads were divided into multiples of 36-mer, which were then mapped to a repeat-masked human reference genome (GRCh38) using mrsFAST (Hach et al., 2010). Up to two mismatches per 36-mer were allowed in order to increase our mapping sensitivity and read depth of mappable sequences in our analysis was corrected for underlying GC content. Finally, copy number estimate for the locus of interest was computed by summarizing over all mappable bases for each sample.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical details of the experiments can be found in the figure legend of related figure panels. For western blot, qPCR. qRT-PCR and size measurements Student's test was used and $p < 0.05$ was considered significant. Quantification of qPCR and qRT-PCR was done using the delta delta Ct method. For bioinformatical statistical analysis see each individual method section and in relevant figure legends.