Young transposable elements rewired gene regulatory networks in human and chimpanzee hippocampal intermediate progenitors

Sruti Patoori[1], Samantha Barnada[1], and Marco Trizzino[1,*]

[1] Department of Biochemistry and Molecular Biology, Sidney Kimmel Medical College, Thomas Jefferson University, Philadelphia, PA

* Corresponding author: Marco Trizzino, Department of Biochemistry and Molecular Biology, Sidney Kimmel Medical College, Thomas Jefferson University, 233 S 10th Street, BLSB 826, Philadelphia, PA, 19104. E-mail: marco.trizzino@jefferson.edu

Abstract

The hippocampus is associated with essential brain functions such as learning and memory. Human hippocampal volume is significantly greater than expected when compared to non-human apes, suggesting a recent expansion. Intermediate progenitors, which are able to undergo multiple rounds of proliferative division before a final neurogenic division, may have played a role in the evolutionary hippocampal expansion. To investigate the evolution of gene regulatory networks underpinning hippocampal neurogenesis in apes, we leveraged the differentiation of human and chimpanzee induced Pluripotent Stem Cells into TBR2-positive hippocampal intermediate progenitors (hpIPCs). We find that the gene networks active in hpIPCs are significantly different between humans and chimpanzees, with ~2,500 genes differentially expressed. We demonstrate that species-specific transposon-derived enhancers contribute to these transcriptomic differences. Young transposons, predominantly Endogenous Retroviruses (ERVs) and SINE-Vntr-Alus (SVAs), were co-opted as enhancers in a species-specific manner. Human-specific SVAs provided substrates for thousands of novel TBR2 binding sites, and CRISPR-mediated repression of these SVAs attenuates the expression of ~25% of the genes that are upregulated in human intermediate progenitors relative to the same cell population in the chimpanzee.

Summary statement: Evolution of human and chimpanzee hippocampal development was mediated by co-option of young retrotransposons into species-specific enhancers.

Key words: hippocampal development, transposable elements, iPSCs, intermediate progenitors, TBR2, SVAs

Introduction

The hippocampus is associated with many traits relevant in the context of human evolution. These include traits such as tool use and language which require social cognition and learning, as well as spatial memory, navigation, and episodic memory (Burgess et al., 2002; Eichenbaum, 2017a; Eichenbaum, 2017b; Squire, 1992; Tomasello and Herrmann, 2010). This region of the brain is also greatly affected by Alzheimer's Disease (AD), a neurodegenerative disorder characterized by cell death, plaques and tangles of misfolded proteins, and cognitive decline (Duyckaerts et al., 2009). It has been hypothesized that the cognitive AD phenotype is uniquely human and that non-human primates, including chimpanzees, do not exhibit AD-related dementia (Edler et al., 2017; Finch and Austad, 2015; Walker and Jucker, 2017). If humans are uniquely susceptible to AD, it is crucial to understand how the human hippocampus differs from that of our closest biological relatives, the chimpanzees.

Human hippocampal volume is 50% greater than expected when compared to the hippocampal volumes of non-human apes, possibly indicating a recent hippocampal expansion specific to the human lineage (Barger et al., 2014). However, the evolution of the human hippocampus and the developmental mechanisms driving the human-specific volume increase have not yet been thoroughly studied.

Recent studies have suggested that evolutionary changes to neuronal progenitors may have had an impact on cortical volume in primates by increasing proliferative potential (Martínez-Cerdeño et al., 2006; Rétaux et al., 2013; Florio and Huttner, 2014). A specific class of neuronal progenitors known as intermediate progenitor cells (IPCs) or "transit amplifying cells" are able to undergo multiple rounds of proliferative division before a final neurogenic division (Englund, 2005; Arnold et al., 2008; Pontious et al., 2008; Hevner, 2019). These cells express the neurodevelopmental transcription factor TBR2 (EOMES) and are found in the sub-ventricular zone of the developing neocortex and hippocampus (Bulfone et al., 1999; Kimura et al., 1999; Englund, 2005; Cipriani et al., 2016). Genetic ablation of TBR2 in these progenitors results in reduced cortical thickness in mice (Sessa et al., 2008), abnormal cortical cell differentiation (Mihalas et al., 2016), and impaired neurogenesis in the hippocampal formation (Hodge et al., 2012). As these IPCs are hypothesized to play a role in neocortical expansion, they may also be involved in the lineage-specific hippocampal expansion seen in humans.

Many of the differences between humans and chimpanzees are due to diverging gene regulatory sequences (Agoglia et al., 2021; Enard et al., 2002; Gokhman et al., 2021; King and Wilson, 1975; Wray, 2007). A recent study comparing gene expression between adult human, chimpanzee and macaque brain regions identified several genes specifically upregulated in the human hippocampus (Sousa et al., 2017). However, transcriptomic differences between primate species during specific timepoints of hippocampal development have not been investigated.

As samples of developing human and chimpanzee brain tissue are extremely limited, Induced Pluripotent Stem Cells (iPSCs) are an ideal system to conduct comparative studies of human and chimpanzee hippocampal development. Previous studies have employed iPSC-derived cortical organoids and iPSC-derived neuronal progenitor cells (NPCs) from human and chimpanzee for comparative and developmental genomic purposes (Marchetto et al., 2019; Mora-Bermúdez et al., 2016). Here, we leverage human and chimpanzee iPSC-derived hippocampal progenitors as models for comparative developmental and genomic studies with the goal of identifying species-specific differences in gene regulation during hippocampal development.

Several recent papers have recently demonstrated that transposable elements (TEs) can alter existing regulatory elements or generate entirely novel ones, as well as expand in a species- or lineage-specific manner (reviewed in Sundaram and Wysocka, 2020). Species-specific TE expansion and co-option into gene regulatory networks have been demonstrated as a mechanism for evolutionary change (Chuong et al., 2016; Fuentes et al., 2018; Jacques et al., 2013; Lynch et al., 2011; Lynch et al., 2015; Miao et al., 2020; Mika et al., 2021; Pontis et al., 2019; Trizzino et al., 2017). Endogenous Retroviruses (ERVs)

and SINE-Vntr-Alus (SVAs) are among the TE families more frequently associated with gene regulatory activity in the human genome (Chuong et al., 2016; Fuentes et al., 2018; Pontis et al., 2019; Trizzino et al., 2017; Trizzino et al., 2018). SVAs encompass six subfamilies, denoted as SVA-A through -F. Of the ~3,000 SVA copies in the human genome, nearly half are human specific, including all the SVA-E and SVA-F (Quinn and Bubb, 2014; Wang et al., 2005). The remaining half are also found in other great apes. SVAs are still replication competent and thus able to transpose in the human genome. ERVs are retrotransposons belonging to the Long Terminal Repeat (LTR) group. They are remnants of past retroviral infection events, and make up ~8% of the human genome (Tokuyama et al., 2018). Both SVAs and ERVs were recently found to be enriched within the sequences of active cis-regulatory elements (enhancers, promoters) in hippocampal tissue as compared to other human brain regions where they are predominantly repressed (Trizzino et al., 2018). Therefore, we hypothesize that species-specific ERV and SVA transposon activity may influence the gene regulatory networks necessary for human and chimpanzee hippocampal development.

Given the key function that intermediate progenitors played in the evolution of the primate brain (Florio and Huttner, 2014; Martínez-Cerdeño et al., 2006), we sought to identify molecular differences between iPSC-derived human and chimpanzee hippocampal intermediate progenitors (hpIPCs) in terms of gene expression and the regulatory activity of non-coding regions. We specifically examined gene expression (RNA-seq), gene regulation (ATAC-seq) and functional TE activity via CRISPR-interference.

After confirming that the hpIPC differentiated cells express the appropriate neurodevelopmental markers, we conducted a transcriptomic comparison between human and chimpanzee hpIPCs. This analysis revealed over 2,500 differentially expressed genes. We then used ATAC-seq to conduct extensive analyses of differential chromatin accessibility between human and chimpanzee hpIPCs. In both species, differentially accessible chromatin regions were more likely than expected to overlap a TE insertion. Further, these regions were found as both enriched and depleted for specific TE families. Notably, species-specific enrichment for ERV and SVA sequences within differentially accessible genomic sites correlated with species-specific changes in nearby gene expression. This is likely driven by transcription factors binding to the TE-derived regulatory sequences, as we demonstrate for TBR2 and SVA-derived enhancers. Finally, we used CRISPR-interference to repress all the accessible SVAs in progenitor-like cells and demonstrated that such repression results in global changes in gene expression and affects hundreds of important neurodevelopmental genes.

This work demonstrates that two young TE families have contributed significantly to the gene regulatory differences between human and chimpanzee hippocampal development, providing insight into how the human hippocampus evolved both its unique cognitive capacity and its susceptibility to neurodegenerative disease.

Results

An iPSC-derived model for human and chimpanzee hippocampal intermediate progenitors

We modeled hpIPCs in humans and chimpanzees using three human and three chimpanzee iPSC lines. All six iPSC lines were validated as pluripotent in previous studies (Gallego Romero et al., 2015; Pagliaroli et al., 2021; Pashos et al., 2017; Ward et al., 2018; Yang et al., 2015; Zhang et al., 2015). For both human and chimpanzee iPSCs, we used two female cell lines and one male line.

We used a previously published method to generate hpIPCs from iPSCs (Yu et al., 2014). In this protocol, the stem cells are treated with a media containing anticaudalizing factors and a Sonic Hedgehog antagonist (DKK1, Nogging, SB431542) to generate forebrain progenitor cell types (Yu et al., 2014). It should be noted that the hpIPCs are distinct from a more general neuronal progenitor cell (pan-NPC) as the pan-NPC media is supplemented only with FGF2 and B27. Moreover, and the hpIPCs can be further

145 induced to generate mature hippocampal CA3 pyramidal or dentate gyrus granule neurons (Sarkar et al.,
146 2018; Yu et al., 2014).
147 As we were specifically interested in TBR2-positive intermediate hippocampal progenitors, we
148 first differentiated one of the human lines for five days and performed a TBR2 time-course western blot
149 on cells collected at day 0, day 5 and day 11. The blot revealed day-5 as a time-point in which TBR2 protein
150 is highly expressed (Fig. 1A).
151 To confirm that day 5 was the optimal time-point in both species, we differentiated one human
152 and one chimpanzee line for five days and used immunocytochemistry to confirm the presence of OTX2,
153 which labels neuronal progenitors, as well as TBR2 and PAX6 which specifically mark the intermediate
154 progenitor cell type (Florio and Huttner, 2014; Hevner, 2019; Fig. 1B). A previous study of human fetal
155 hippocampal development demonstrated that 47% and 57% of TBR2-positive cells express PAX6 at
156 gestational weeks 11 and 13, respectively (Cipriani et al., 2016). We found that nearly all of the
157 cells treated with hpIPC media for five days were positive for all three neurodevelopmental markers in
158 both species (Fig. 1B). Together, these data indicate that our iPSC-derived model is suitable to study TBR2-
159 positive hippocampal intermediate progenitors in both human and chimpanzee.
160
161 **Important neurodevelopmental genes are differentially expressed between human and chimpanzee**
162 **hippocampal intermediate progenitors**
163 To investigate the differences between human and chimpanzee hippocampal intermediate
164 progenitors, we first aimed to characterize differential gene expression between the hpIPCs of the two
165 species. After five days of treatment with the hpIPC differentiation media, we collected cells for RNA
166 extraction (Fig. 2A). To ensure statistical power, we conducted bulk RNA-seq on two replicates each of all
167 six cell lines (i.e. three biological replicates and six technical replicates per species). The differentiation,
168 the harvesting, and the RNA processing were performed in mixed batches with samples from both species
169 to prevent batch effects. The libraries were sequenced using Illumina NextSeq500, generating 100 bp
170 Paired-End reads. Non-orthologous genes were omitted from the analysis and a total of 2,588 genes were
171 identified as differentially expressed (FDR <0.05 and log2foldChange > 1.5 or < -1.5). The genes with
172 log2foldChange > 1.5 were more highly expressed in the human hpIPCs ("Human UP") while the genes
173 with log2foldChange < -1.5 were more highly expressed in the chimpanzee hpIPCs ("Chimp UP"). In total,
174 1,686 (65.1%) of the differentially expressed genes (DEG) were "Human UP" and 901 (34.9%) of the DEG
175 were "Chimp UP" genes (Fig. 2B, Supplementary Table 1.1).
176 Hippocampal neurodevelopmental markers *PAX6*, *OTX2*, *NEUROD1* and *FOXG1* were found as
177 highly expressed in both species. This is consistent with the immunostaining in demonstrating that
178 the hpIPCs differentiation was successful and that the progenitors from both species are comparable.
179 The "Human UP" genes include *FOXP2*, *MTRNR2L8*, *DHX40*, *VPS13B*, *WDFY2*, and *PURB*, all of
180 which are associated with neurodevelopment or neurodegenerative disease (Abrajano et al., 2009;
181 Hickey et al., 2019; Kamboh et al., 2019; Kolehmainen et al., 2003; MacDermot et al., 2005; Mathys et al.,
182 2019; Sin et al., 2015; Taher et al., 2014). The "Chimp UP" genes include *HIST1H3A*, *BLC2L2*, and
183 *CLIC1* which have been reported as expressed in the hippocampus and associated with sleep, AD, and
184 neurite outgrowth (Averaimo et al., 2014; Datson et al., 2009; Wei, 2020).
185 To understand the transcriptional programs driving these differences in gene expression,
186 we conducted an Ingenuity Pathway Analysis. Three of the top five upstream regulators predicted by the
187 pathway analysis were the transcription factors CREB1, FOXA2 and TBR2 (EOMES) (Fig. 2C). CREB1 is
188 known to regulate genes involved in the nervous system and neurodevelopment (reviewed in Sakamoto
189 et al., 2011), FOXA2 controls dopaminergic neuronal development and disease (Kittappa et al., 2007), and
190 TBR2 plays a crucial role in cortical and hippocampal neurogenesis and is the signature marker of the
191 intermediate progenitor population (Cipriani et al., 2016; Englund, 2005). This pathway analysis was
192 consistent with the RNA-seq data as predicted targets of all three transcription factors were also found to

193 be among the 2,588 genes differentially expressed between human and chimpanzee hpIPCs (Fig. 2D,
194 Supplementary Table 1.2-1.4). The pathway analysis also determined that several of the differentially
195 expressed genes are involved in embryonic development (Fig. 2D, Supplementary Table 1.5). Overall,
196 these findings indicate that previously characterized neurodevelopmental gene regulatory networks are
197 utilized differently during human and chimpanzee hippocampal development.
198
199 **Human-specific chromatin accessibility patterns in hippocampal intermediate progenitors**
200 After identifying differentially expressed genes and the transcriptional networks that may be
201 involved, we sought to identify cis-regulatory differences between the human and chimpanzee hpIPCs. To
202 this end, we conducted ATAC-seq on the hpIPCs from both species. We used the same batches of
203 differentiated iPSCs for the ATAC-seq as we did for the RNA-seq (i.e. from the same batch of
204 differentiation), and generated 100 bp long Paired-End Illumina reads.
205 We first performed a "human-centric" analysis. We aligned the ATAC-seq reads from all six cell
206 lines to the respective reference genome assemblies (hg19 for the human cell lines, panTro5 for the
207 chimpanzee cell lines) and only retained uniquely mapped reads with high mapping quality (Samtools Q =
208 10 filtering). Next, we identified regions of accessible chromatin (peaks; FDR < 0.05) in all three human
209 cell lines. Only peaks replicated in all the three human lines were retained. To carry out a proper
210 comparison, we only retained replicated human ATAC-seq peaks with orthologs in the chimpanzee
211 genome (see methods; Fig. 3A). This filtering ultimately resulted in 82,235 human ATAC-seq peaks
212 replicated in all the human cell lines and with orthologs in the chimpanzee genome. These 82,235 regions
213 were used for downstream analysis. We found that the chromatin accessibility at these regions was highly
214 reproducible across all three human cell lines (Fig. 3B).
215 Next, we quantified the ATAC-seq read depth for each of the 82,235 regions and used DESeq2 to
216 identify sites exhibiting differential chromatin accessibility between the two species. In total, we identified
217 3,006 differentially accessible regions (FDR <0.05; log2foldChange > 1.5 or < -1.5; Supplementary Table
218 2.1-2.2). Of these regions, 92.3% were located at least 1 kb away from the closest transcription start site
219 (TSS), suggesting they could be putative enhancers, while the remaining were putative promoters (Fig.
220 3C). As expected, given that this analysis was performed with a human-centric approach, 90.1% of the
221 differentially accessible (DA) peaks were significantly more accessible in the human hpIPCs relative to
222 chimpanzee hpIPCs ("Human UP"; Fig. 3C).
223 As TE insertions can be a source of cis-regulatory evolution, we examined whether the DA regions
224 were more likely to overlap with a TE than those which were accessible to the same degree in both
225 species (non-DA, Fig. 3D and Supplementary Table 2.4-2.5). We observed that 1,335/3,006 (44.4%) DA
226 regions overlapped a TE (Supplementary Table 2.3). This is significantly higher than what was observed
227 for the non-DA peaks (33.2% overlapped a TE; Two-sided Fisher's Exact Test, p < 0.0001; Fig. 3D). This
228 indicates that chromatin regions with human-specific accessibility are significantly more likely to be TE-
229 derived than the regions with accessibility levels conserved between human and chimpanzee.
230 Next, we associated the nearest gene to each DA region and found that TE-derived DA regions
231 were significantly more likely to be near a differentially expressed gene relative to non-TE derived DA
232 regions (Two-sided Fisher's Exact Test, p = 0.016; Fig. 3E). Finally, we investigated whether specific TE
233 families were overrepresented among the TE-derived DA regions and found enrichment for LTRs. While
234 LTRs account for ~16% of all human annotated TEs, they represented 33.9% of the TEs overlapping DA
235 regions in our human-centric analysis (Two-Sided Fisher's Exact Test p < 0.0001; Fig. 3F and Supplementary
236 Table 2.6). Of these enriched LTRs, 97.1% were ERVs. Notably, 31.7% of those LTR-derived DA regions
237 were located near a differentially expressed gene (Two-Sided Fisher's Exact Test p < 0.0001).
238 Together, these data indicate that there are TE-derived cis-regulatory elements that have
239 significantly greater accessibility in humans than chimpanzees during hippocampal neurogenesis. These
240 TE insertions preceded the human-chimpanzee split, but the difference in accessibility is species-specific,

241 suggesting that the co-option into gene regulatory networks took place after the human-chimpanzee
242 divergence.
243
244 **Chimpanzee-specific chromatin accessibility patterns in hippocampal intermediate progenitors**
245 We repeated the ATAC-seq analysis as described above, but this time with a "chimpanzee-centric"
246 approach. We started from a set of 72,211 peaks found as replicated in all the three chimpanzee lines
247 and with orthologs in both species (Figs. 4A,B). With this approach, we identified 3,806 ATAC-seq peaks
248 as differentially accessible (DA) between human and chimpanzee (FDR <0.05; log2foldChange > 1.5 or < -
249 1.5; Supplementary Table 3.1-3.2), 82% of which displayed greater accessibility in chimpanzee as
250 compared to humans (i.e. "Chimp UP"; Fig. 4C). Similar to what we observed with the human-centric
251 analysis, 97.9% of the 3,806 DA regions were putative enhancers (distance > 1Kb from TSS; Fig. 4C).
252 As seen in the human-centric analysis, chimpanzee-specific DA peaks were more likely to be TE-
253 derived than those similarly accessible across species (Two-sided Fisher's Exact Test, $p < 0.0001$; Fig. 4D
254 and Supplementary Tables 3.3-3.5). Of the chimpanzee DA regions, 37.1% overlapped an annotated
255 chimpanzee TE, compared to only 28.1% of the non-DA regions (Two-sided Fisher's Exact Test, $p < 0.0001$;
256 Fig. 4D). However, the TE-derived enhancers in this chimpanzee-centric analysis were no more likely than
257 the non-TE derived ones to be located near differentially expressed genes (Fig. 4E).
258 We found enrichment for LTRs, which account for approximately 16% of the chimpanzee TEs but
259 represented 36.1% of the TE-derived DA regions (Two-sided Fisher's Exact Test, p-value < 0.0001; Fig. 4F
260 and Supplementary Table 3.6; 98.2% were ERVs), and SVAs, which account for just 0.25% of annotated
261 chimpanzee TEs but represented 1.1% of the TE-derived DA regions (Two-sided Fisher's Exact Test, $p <$
262 0.0001, Fig 4G and Supplementary Table 3.8). In particular, the SVA-B and C subfamilies were the most
263 enriched (Fig. 4H). Together, these data indicate that chimpanzee ERV and SVA transposons were co-
264 opted into regulatory elements important for the developing chimpanzee hippocampal intermediate
265 progenitors.
266
267 **Genomic features underlying species-specific LTR enrichment at hpIPC enhancers**
268 We aimed to further investigate genomic features potentially underlying the LTR enrichment
269 among the differentially accessible hippocampal progenitor regions. To this end, we conducted a motif
270 analysis using the MEME suite (Bailey et al., 2015). Binding motifs for CTCF and EGR2 were detected as
271 enriched in the LTR-derived differentially accessible regions identified from both the human-centric and
272 chimpanzee-centric analyses (Figs. 5A, 5C). EGR2 is an early response gene involved in learning and
273 memory, in the brain's response to stimuli, and in hippocampal synaptic plasticity (Cheval et al., 2012;
274 Mukherjee et al., 2021; Poirier et al., 2007). CTCF, a well-known regulator of chromatin structure, has
275 been implicated in various neurodevelopmental disorders (reviewed in Davis et al., 2018)
276 We identified 63 differentially expressed genes located near the human-enriched LTRs (Fig. 5B
277 and Supplementary Table 2.7). These included *GLUL*, a glutamine synthetase hypothesized to provide
278 neuroprotection in AD patients, (Kohane and Wood, 2021) and *DLK1*, a Notch ligand involved in SVZ
279 neurogenesis (Ferrón et al., 2011), both of which are upregulated in human hpIPCs compared to
280 chimpanzee hpIPCs.
281 We identified 34 differentially expressed genes located near the chimp-enriched LTRs (Fig. 5D and
282 Supplementary Table 3.7). For the chimp-centric analysis the differentially expressed genes located near
283 TE-derived enhancers with species-specific accessibility included *HIST1H3A* and *MTRNR2L8* (Fig. 5D).
284 These genes were highly upregulated in the chimpanzee and human hpIPCs, respectively (Fig. 5D).
285 Importantly, *HIST1H3A* has been associated with autism spectrum disorders and sleep deprivation
286 (Crawley et al., 2016; Wei, 2020) while *MTRN2L8* has been reported as upregulated in Alzheimer's patients
287 (Mathys et al., 2019). *BDNF*, which is involved in hippocampal neurogenesis and plays an important role

in the SVZ (Bath et al., 2012), was highly upregulated in the chimpanzee and also located near an LTR-derived enhancer with chimpanzee-specific accessibility.

Human-Specific SVAs play a major role in hippocampal neurogenesis

As mentioned earlier, the chimpanzee-centric analysis also identified SVA transposons as enriched within chimpanzee-specific enhancers (Fig. 6A and Supplementary Table 3.8). These SVAs were enriched for the identified binding motifs of the neurodevelopmental factors ASCL1, ZIC1, and KLF8 (Andersen et al., 2014; Aruga, 2004; Yi et al., 2014) as well as the JUN/FOS AP-1 dimer, which is a known enhancer activator (Raivich, 2008; Raivich and Behrens, 2006) (Fig 6A).

It is important to remark that all the analyses shown so far were exclusively based on genomic sites with characterized orthologs in both species, to ensure an "apple-to-apple" comparison. However, nearly 2,000 SVA copies, including the entire SVA-E and F subfamilies, are exclusive to the human genome. Given that previous studies found that SVAs are highly enriched in active enhancers and promoters of the human hippocampus (Trizzino et al., 2018), we sought to investigate this further. We focused on the SVA copies exclusively present in the human genome and not in any other primate genome (hereafter human-specific SVAs). We performed sequence-based motif analysis for all the human-specific SVAs and identified the binding motif for the hpIPCs signature factor TBR2 (EOMES) as the most enriched ($p = 10^{-2103}$; Fig. 6B). Motifs for other transcription factors associated to hippocampal neurogenesis and function were also recovered (SMAD4, VDR, PLAG1; Fig. 6B).

Next, we annotated all of the genes found within 50 kb from each human-specific SVA. A total of 2,216 genes were recovered using this approach. We performed Pathway Analysis on this set of genes and found that Melatonin Degradation and Nicotine Degradation were the two most significantly enriched pathways ($p = 7.6 \times 10^{-6}$ and $p = 2.5 \times 10^{-5}$ respectively; Fig. 6C) and PHF8 was recovered as the top upstream regulator for the gene network (Fig. 6C). Notably, both Melatonin and Nicotine Degradation pathways are strongly active in the human hippocampus. PHF8 is a histone demethylase that contributes to the regulation of mTOR. The mTOR pathway is hyperactive in the human hippocampus where it regulates the protein synthesis-dependent plastic changes underlying learning and memory (Bekinschtein et al., 2007; Fortress et al., 2013; Graber et al., 2013). Mutations in the *PHF8* gene cause cognitive impairment and intellectual disability (Chen et al., 2018).

Therefore, we sought to determine the contribution of human-specific SVAs to the TBR2-mediated gene regulatory network in human hpIPCs. Using our ATAC-seq data, we identified 1,816 human SVAs as accessible in human hpIPCs (Fig. 6D). Of these, nearly a quarter (434) displayed high accessibility, while 1,382 were moderately accessible (Fig. 6D). Next, we used chromatin immunoprecipitation followed by sequencing (ChIP-seq) to profile TBR2 binding in two human lines at day 5 of hpIPC differentiation. As with the previous sequencing experiments, we generated 100 bp long Paired-End reads and only retained uniquely mapping high quality reads (Samtools q=10 filtering) in order to maximize our chance to properly map reads on repetitive regions. This experiment revealed that 739 of the accessible SVAs showed TBR2 signal in the two human lines (Fig. 6D and Supplementary Table 2.8). Notably, 257 of the (48.3%) TBR2-bound SVAs were human-specific. (Fig. 6E and Supplementary Table 2.9). TBR2-bound human-specific SVAs were located near 37 genes that our RNA-seq analysis identified as differentially expressed between human and chimpanzee (Fig. 6F and Supplementary Table 2.10). These genes include *VPS13B* (up in humans), which is responsible for a rare developmental disease known as Cohen Syndrome (Kolehmainen et al., 2003), *NR4A2* (down in humans), which has been implicated in neurodevelopmental language impairment (Reuter et al., 2017), *DHX40* (down in humans), which is implicated in Alzheimer's (Taher et al., 2014), and *E2F1* (up in humans), which is a cell cycle regulator associated with several neurodegenerative diseases such as Alzheimer's (Zhang et al., 2010).

In summary, these data support a model in which human-specific SVAs provided a substrate for binding sites of TBR2 and other important hippocampal regulators. Therefore, it is likely they were co-

336 opted in the gene regulatory networks active during hippocampal neurogenesis, which led to human-
337 specific regulation of key genes.
338
339 **CRISPR-mediated SVA repression has massive repercussions on global gene expression**
340 To further assess the contribution of SVA transposons to the gene regulation of hippocampal
341 intermediate progenitors, we leveraged CRISPR-interference to simultaneously repress most of the active
342 SVAs. We utilized NCCIT cells treated with retinoic acid (RA) as the experimental system for this purpose.
343 The NCCIT cell line is derived from embryonal carcinoma and thus exhibits a gene expression signature
344 highly similar to human embryonic stem cells (Fuentes et al., 2018). Importantly, NCCITs treated for 7 days
345 with RA differentiate into intermediate neural progenitor-like cells expressing both PAX6 and TBR2
346 (Mandal et al., 2015; Fig 7A). We cloned a stable NCCIT line with a permanently incorporated doxycycline-
347 inducible, catalytically-dead Cas9 fused to a repressive KRAB domain (dCas9-KRAB). The KRAB domain
348 deposits repressive histone methylation (H3K9me3) to the regions targeted by the dCas9 via single guide-
349 RNAs (sgRNAs). Into this same line, we also permanently knocked-in two single guide-RNAs (sgRNAs) that
350 are able to target >80% of all SVAs (Pontis et al. 2019). Hereafter, we refer to the RA-treated CRISPR line
351 as RA-NCCITs. Remarkably, exposing the RA-NCCITs to doxycycline for 72 hours was sufficient to induce
352 dCas9 activation (Fig. 7B) and the deposition of repressive histone methylation (H3K9me3) on over 2,500
353 previously unmethylated SVAs (Fig. 7C). We next performed RNA-seq on the RA-NCCITs with or without
354 doxycycline treatment (three replicates per condition). First, we observed that the genome-wide
355 expression levels (TPMs) of the RA-NCCITs where highly correlated with those of the iPSC-derived human
356 hpIPCs (Pearson correlation = 0.93; $p < 2.2 \times 10^{-16}$). This indicates that RA-treated NCCITs are appropriate
357 to model hippocampal intermediate progenitors.
358 Then, we compared the expression levels of the RA-NCCITs with or without doxycycline treatment
359 and identified 5,795 differentially expressed genes (FDR < 0.05; Fig. 7D and Supplementary Table 4.1). Of
360 these genes, 677 were previously identified as differentially expressed when comparing human to
361 chimpanzee hpIPCs (Fig. 7E; Supplementary Table 4.2). In other words, the expression of over a quarter
362 (26.1%) of the genes that exhibited human-specific expression signature in hippocampal intermediate
363 progenitors seem to be under control of SVA transposons. Importantly, one of these genes was *FOXP2*,
364 which has been associated with the evolution of language and is implicated in several speech-disorders
365 (Enard, 2011; Liégeois et al., 2016; MacDermot et al., 2005).
366 Since the gRNAs used for this experiment were originally designed to target a DNA sequence
367 shared by the SVAs with the LTR5H family (Pontis et al. 2019), we restricted the analysis to the genes that
368 are associated to human SVAs (i.e. only considering the genes that represent the closest gene to an
369 annotated SVA: hereafter SVA-genes). By doing so, we found 611 SVA-genes as differentially expressed in
370 RA-NCCITs upon SVA repression (Supplementary Table 4.3). Of these, 90 were previously identified as
371 differentially expressed when comparing human to chimpanzee hpIPCs (Supplementary Table 4.4). Thus,
372 the expression of these 90 genes can be bona-fide considered as directly regulated by SVA-derived
373 enhancers in both primary hippocampal progenitors and in the NCCIT cell line. Remarkably, the large
374 majority of these genes (72.5%) had decreased expression upon SVA repression. These include *SOX2*,
375 *FGF2*, *PRDM1*, *NTRK2*, and *TFAP2B* (Supplementary Table 4.4).
376 Finally, we examined the genes previously identified as being near a TBR2-bound human-specific
377 SVA in iPSC-derived hpIPCs, and found that nearly a third of them lose expression in RA-NCCITs upon SVA
378 repression (Fig. 7F). In summary, our functional experiments indicate a widespread role for human-
379 specific SVA transposons as cis-regulatory elements during hippocampal neurogenesis.
380
381 **Discussion**
382 The hippocampus is susceptible to specific neurodegenerative disorders such as Alzheimer's
383 Disease but may have also played an important role in the evolution of human cognition. Spatial memory,

384 which is attributed to the hippocampus, may have contributed to the geographic expansion of ancient
385 humans. Characterizing the human-specific gene regulatory networks of hippocampal development
386 provides insight into its role in human evolution. Though there is no consensus on whether the cognitive
387 phenotypes seen in AD are uniquely human, understanding the unique properties of the human
388 hippocampus may lead the way to potential treatments. Thus, the work described here is relevant both
389 in terms of evolutionary developmental biology and evolutionary medicine.
390 To study the evolution of the human hippocampus from a developmental standpoint, we
391 investigated the extent to which TEs contributed to gene expression profiles of human and chimpanzee
392 hippocampal intermediate progenitors. TEs account for nearly ~50% of the human genome composition
393 and many elegant studies have established that at least a fraction of the TEs can regulate the host genes
394 in humans and other primates (Chuong et al., 2013; Chuong et al., 2016; Cosby et al., 2021; del Rosario et
395 al., 2014; Du et al., 2016; Fuentes et al., 2018; Jacques et al., 2013; Judd et al., 2021; Lynch et al., 2011;
396 Lynch et al., 2015; Mika et al., 2021; Modzelewski et al., 2021; Okhovat et al., 2020; Rayan et al., 2016;
397 Schmidt et al., 2012; Sundaram et al., 2014; Trizzino et al., 2017; Trizzino et al., 2018; Ward et al., 2018;
398 Xie et al., 2013).
399 The hpIPCs are a transient progenitor population during a critical developmental stage in the Sub-
400 ventricular zone (SVZ) of both the hippocampus and neocortex (Bulfone et al., 1999; Cipriani et al., 2016;
401 Englund, 2005; Kimura et al., 1999), and there is consensus that this progenitor population may have
402 played a role in the evolution of brain volume in mammals (Florio and Huttner, 2014; Martínez-Cerdeño
403 et al., 2006). To study the developmental evolution of human hpIPCs, we leveraged a comparative
404 approach centered on differentiating human and chimpanzee iPSCs into a neuronal population which
405 closely recapitulates hpIPCs, as demonstrated by the high expression of signature markers such as TBR2
406 and PAX6.
407 The transcriptomes of human and chimpanzee hpIPCs have not been previously compared, largely
408 due to the limited availability of primary tissue. Here, we carried out this comparison using our iPSC-
409 derived system and identified profound differences between the two species, with over 2,500 genes
410 differentially expressed at this stage. These genes include several which were previously associated with
411 cognitive function, language, neurodevelopment and neurodegeneration, many of which are upregulated
412 in humans relative to the chimpanzee (e.g. *MTRNR2L8*, *DHX40*, *VPS13B*, *WDFY2*, *PURB*).
413 We demonstrate that species-specific enhancers significantly contributed to the gene expression
414 differences we identified. This is consistent with recent studies that used primary brain tissues from
415 several species to profile species-specific cis-regulatory activity in mammals (Emera et al., 2016; Reilly et
416 al., 2015). Importantly, we find that these species-specific enhancers are enriched for young transposable
417 elements. Several studies have identified evolutionarily young L1 LINEs as active in the brain during
418 different developmental stages, suggesting that they could serve as alternative promoters for many genes
419 involved in neuronal functions (Coufal et al., 2009; Jönsson et al., 2019; Sur et al., 2017; Thomas et al.,
420 2012; Zhao et al., 2019). Here, we identified other young transposable elements, ERVs and SVAs, as
421 regulators of intermediate progenitor gene expression. The identification of ERVs as candidate enhancers
422 in human and chimpanzee intermediate progenitors is consistent with previous studies that
423 demonstrated that ERVs heavily impact the gene regulatory programs during immune response (Chuong
424 et al., 2013; Chuong et al., 2016), in pluripotency maintenance and development (Coluccio et al., 2018;
425 Fuentes et al., 2018; Miao et al., 2020), in the mammalian placenta (Lynch et al., 2011; Lynch et al., 2015;
426 Mika et al., 2021), in the primate liver (Trizzino et al., 2017), and in many cancer types (Ito et al.; Ivancevic
427 and Chuong, 2020; Shah et al., 2021). Similar to the L1s, the ERVs also possess a well-defined cis-regulatory
428 architecture (e.g. they have their own promoter), and this may have had a role in the co-option of these
429 TEs as functioning cis-regulatory elements.
430 The SVAs are particularly interesting from a human evolution standpoint, given that half of the
431 known copies are exclusively present in our species. Moreover, SVAs are among the few transposable

elements that still exhibit active transposition in the human genome. We and others have previously revealed that SVAs can be source of enhancers in primates (Playfoot et al., 2021; Pontis et al., 2019; Trizzino et al., 2017; Trizzino et al., 2018). The repression of some SVAs by specific zinc-finger proteins at specific stages of neuronal development is also a crucial mechanism for successful neurogenesis (Playfoot et al., 2021; Pontis et al., 2019; Turelli et al., 2020). Here, we demonstrate that SVAs are pervasive regulators of hippocampal neurogenesis and they act as enhancers in the hippocampal intermediate progenitor population. By using CRISPR-interference (CRISPR-i), we show that repressing hundreds of normally "de-repressed" SVAs alters the expression of thousands of genes. Intriguingly, our CRISPR-i experiments revealed that global SVA repression leads to the attenuation of the expression levels of over a quarter of the ~2,500 genes previously identified as with human-specific gene expression signal in hippocampal intermediate progenitors. These include critical neurodevelopmental regulators such as *FOXP2*, *HAND2*, *MEF2C*, *SOX2*, and *SOX4*. In normal conditions, these genes are more highly expressed in humans relative to the chimpanzee, but upon SVA repression these differences were diminished.

In conclusion, our findings indicate that the development of hippocampal neurons has been profoundly affected by the domestication of young transposable elements. These young TEs have been co-opted as functional enhancers and promoters and ultimately rewired the expression of hundreds of critically important neuro-regulators in the developing human hippocampus. The human-specific gene expression and the associated TE-derived enhancers we have identified here may play important roles both in human evolution and neurodegenerative disease.

Materials and Methods

Antibodies

Antibody	Catalog number	Use	Dilution
TBR2	Abcam ab23345	Immunofluorescence	1:250
TBR2	Abcam ab216870	ChIP	15 ug per ChIP
TBR2	Santa Cruz sc - 293481	Western Blot	1:500
PAX6	DHSB AB_528427	Immunofluorescence	1:20
OTX2	R&D Systems AF1979	Immunofluorescence	1:500
H3K9me3	Abcam ab8898	ChIP	3 ug per ChIP
GAPDH	CST #5174	Western Blot	1:1000

Human and Chimpanzee iPSC cultures

The human male iPSC line denoted as SV20 was obtained from the University of Pennsylvania, where it was generated, and validated by the expression of pluripotency markers and differentiation into various cell types in multiple studies (Pagliaroli et al., 2021; Pashos et al., 2017; Yang et al., 2015; Zhang et al., 2015). The human female iPSC line GM 23716 was obtained from the Coriell Institute for Medical Research (Camden, NJ) and validated by the expression of pluripotency markers and differentiation into cranial neural crest cells in a previous study (Pagliaroli et al., 2021). The human female iPSC line 21792 and all three of the chimpanzee iPSC lines were obtained from the laboratory of Yoav Gilad at the University of Chicago and validated in previous studies (Gallego Romero et al., 2015; Ward et al., 2018)

The iPSC lines were expanded in feeder-free, serum-free mTeSR™1 medium (STEMCELL Technologies). Cells were passaged ~1:10 at 80% confluency using ReLeSR (STEMCELL Technologies) and small cell clusters (50–200 cells) were subsequently plated on tissue culture dishes coated overnight with Geltrex™ LDEV-Free hESC-qualified Reduced Growth Factor Basement Membrane Matrix (Fisher-Scientific).

471
472 **hpIPC Differentiation**
473 The iPSC lines were differentiated into hpIPCs as previously described (Yu et al. 2014). 3 batches comprised
474 of one human and one chimpanzee iPSC line each were cultured until approximately 50-70% confluence
475 was reached and then treated with the hpIPC media for five days prior to collection for RNA-seq, ATAC-
476 seq, ChIP-seq or immunofluorescence. The hpIPC media consists of DMEM/F12, 0.5X N2, 0.5X B27, DKK-
477 1, cyclopamine, Noggin, and SB431542 as well as the antibiotic Penn-Strep.
478
479 **Western Blot**
480 For total lysate, cells were harvested and washed three times in 1X PBS and lysed in RIPA buffer (50mM
481 Tris-HCl pH7.5, 150mM NaCl, 1% Igepal, 0.5% sodium deoxycholate, 0.1% SDS, 500uM DTT) with proteases
482 inhibitors. Twenty µg of whole cell lysate were loaded in Novex WedgeWell 4-20% Tris-Glycine Gel
483 (Invitrogen) and separated through gel electrophoresis (SDS-PAGE) Tris-Glycine-SDS buffer (Invitrogen).
484 The proteins were then transferred to ImmunBlot PVDF membranes (ThermoFisher) for antibody probing.
485 Membranes were incubated with 10% BSA in TBST for 30 minutes at room temperature (RT), then
486 incubated for variable times with the suitable antibodies diluted in 5% BSA in 1X TBST, washed with TBST
487 and incubated with a dilution of 1:10000 of secondary antibody for one hour at RT. The antibody was then
488 visualized using Super Signal West Dura Extended Duration Substrat (ThermoFisher) and imaged with
489 Amersham Imager 680. Used antibodies are listed in the "Antibodies" section of the methods.
490
491 **Immunofluorescence**
492 Upon fixation (4% PFA for 10 minutes), cells were first treated with 10 mM sodium citrate (pH = 6) for 10
493 minutes at 95 C. The cells were then permeabilized in blocking solution (0.1% Triton X-100, re PBS, 5%
494 normal donkey serum) for one hour at room temperature and then incubated overnight at 4 C with the
495 primary antibody of interest. Subsequently, the cells were treated with blocking solution for 30 minutes
496 at room temperature and then fluorophore-conjugated secondary antibodies for 2 hours at room
497 temperature. The cell nuclei were stained with 4,6-diamidine-2-phenylindole dihydrochloride (DAPI;
498 Sigma-Aldrich; 50 mg/ml in PBS for 15 min at RT). Immunostained cells analyzed with confocal microscopy,
499 using a Nikon A1R+. Images were captured with a 40X objectives and a pinhole of 1.2 Airy unit. Analyses
500 were performed in sequential scanning mode to rule out cross-bleeding between channels. Used
501 antibodies are listed in the "Antibodies" section of the methods.
502
503 **Real-time quantitative polymerase chain reaction (RT-qPCR)**
504 Cells were lysed in Tri-reagent and RNA was extracted using the Direct-zol RNA MiniPrep kit (Zymo
505 research). 600ng of template RNA was retrotranscribed into cDNA using RevertAid first strand cDNA
506 synthesis kit (Thermo Scientific) according to manufacturer directions. 15ng of cDNA were used for each
507 real-time quantitative PCR reaction with 0.1 µM of each primer, 10 µL of PowerUp™ SYBR™ Green Master
508 Mix (Applied Biosystems) in a final volume of 20 µl, using QuantStudio 3 Real-Time PCR System (Applied
509 Biosystem). Thermal cycling parameters were set as following: 3 minutes at 95°C, followed by 40 cycles of
510 10 s at 95°C, 20 s at 63°C followed by 30 s at 72°C. Each sample was run in triplicate. 18S rRNA was used
511 as normalizer. Primer sequences are reported in Supplementary Data 1.
512
513 **ChIP-Seq and ChiP-qPCR**
514 Samples from different conditions were processed together to prevent batch effects. 15 million cells were
515 cross-linked with 1% formaldehyde for 5 min at room temperature, quenched with 125mM glycine,
516 harvested and washed twice with 1× PBS. The pellet was resuspended in ChIP lysis buffer (150 mM NaCl,
517 1% Triton X-100, 0,7% SDS, 500 µM DTT, 10 mM Tris-HCl, 5 mM EDTA) and chromatin was sheared to an
518 average length of 200–500 bp, using a Covaris S220 Ultrasonicator. The chromatin lysate was diluted with

519 SDS-free ChIP lysis buffer. For ChIP-seq, 10 µg of antibody (3 µg for H3K27ac) was added to 5 µg of
520 sonicated chromatin along with Dynabeads Protein A magnetic beads (Invitrogen) and incubated at 4 °C
521 overnight. On day 2, beads were washed twice with each of the following buffers: Mixed Micelle Buffer
522 (150 mM NaCl, 1% Triton X-100, 0.2% SDS, 20 mM Tris-HCl, 5 mM EDTA, 65% sucrose), Buffer 500 (500
523 mM NaCl, 1% Triton X-100, 0.1% Na deoxycholate, 25 mM HEPES, 10 mM Tris-HCl, 1 mM EDTA),
524 LiCl/detergent wash (250 mM LiCl, 0.5% Na deoxycholate, 0.5% NP-40, 10 mM Tris-HCl, 1 mM EDTA) and
525 a final wash was performed with 1× TE. Finally, beads were resuspended in 1× TE containing 1% SDS and
526 incubated at 65 °C for 10 min to elute immunocomplexes. Elution was repeated twice, and the samples
527 were further incubated overnight at 65 °C to reverse cross-linking, along with the untreated input (5% of
528 the starting material). On day 3, after treatment with 0.5 mg/ml Proteinase K for 1h at 65 °C, DNA was
529 purified with Zymo ChIP DNA Clear Concentrator kit and quantified with QUBIT.
530 For all ChIP-seq experiments, barcoded libraries were made with NEB ULTRA II DNA Library Prep Kit for
531 Illumina, and sequenced on Illumina NextSeq 500, producing 100 bp PE reads.
532 For ChIP-qPCR, on day 1 the sonicated lysate was aliquot into single immunoprecipitations of 2.5×10^6
533 cells each. A specific antibody or a total rabbit IgG control was added to the lysate along with Protein A
534 magnetic beads (Invitrogen) and incubated at 4 °C overnight. On day3, ChIP eluates and input were
535 assayed by real-time quantitative PCR in a 20 µl reaction with the following: 0.4 µM of each primer, 10 µl
536 of PowerUp SYBR Green (Applied Biosystems), and 5 µl of template DNA (corresponding to 1/40 of the
537 elution material) using the fast program on QuantStudio qPCR machine (Applied Biosystems). Thermal
538 cycling parameters were: 20sec at 95 °C, followed by 40 cycles of 1sec at 95°C, 20sec at 60°C. Used
539 antibodies are listed in the "Antibodies" section of the methods.
540
541 **ChIP-seq Analyses**
542 After removing the adapters with TrimGalore!, the sequences were aligned to the reference hg19, using
543 Burrows Wheeler Alignment tool (BWA), with the MEM algorithm[64] . Uniqueky mapping aligned reads
544 were filtered based on mapping quality (MAPQ > 10) to restrict our analysis to higher quality and likely
545 uniquely mapped reads, and PCR duplicates were removed. We called peaks for each individual using
546 MACS2[65] (H3K27ac) or Homer[66], at 5% FDR, with default parameters.
547
548 **RNA-Seq**
549 Cells were lysed in Tri-reagent (Zymo research) and total RNA was extracted using Quick-RNA Miniprep
550 kit (Zymo research) according to the manufacturer's instructions. RNA was further quantified using
551 DeNovix DS-11 Spectrophotometer while the RNA integrity was checked on Bioanalyzer 2100 (Agilent).
552 Only samples with RIN value above 8.0 were used for transcriptome analysis. RNA libraries were prepared
553 using 1 µg of total RNA input using NEBNext® Poly(A) mRNA Magnetic Isolation Module, NEBNext®
554 UltraTM II Directional RNA Library Prep Kit for Illumina® and NEBNext® UltraTM II DNA Library Prep Kit for
555 Illumina® according to the manufacturer's instructions (New England Biolabs).
556
557 **RNA-Seq Analyses**
558 After removing the adapters with TrimGalore!, Kallisto (Bray et al., 2016) was used to count reads mapping
559 to each gene. We analyzed differential gene expression levels with DESeq2 (Love et al., 2014) , with the
560 following model: design = ~condition, where condition indicates either Human or Chimpanzee.
561
562 **ATAC-Seq**
563 For ATAC-Seq experiments, 50,000 cells per condition were processed as described in the original ATAC-
564 seq protocol paper (Buenrostro et al. 2013). ATAC-seq data were processed with the same pipeline
565 described for ChIP-seq, with one modification: all mapped reads were offset by +4 bp for the forward-
566 strand and -5 bp for the reverse-strand. Peaks were called using MACS2 (Zhang et al., 2008).

Generation of the NCCIT-dCas9KRAB-SVAsgRNA Stable Cell Line

A dCas9-KRAB was cloned into a piggyBac transposon containing ampicillin and puromycin resistance (Addgene) which was obtained from the Wysocka Lab at Stanford University. The piggyBac dCas9-KRAB doxycycline-inducible plasmid, along with a piggyBac transposase (Cell Signaling), were transfected into NCCIT cells (ATCC) at ~70% confluency using a 6:1 ratio of Fugene HD (Promega) for 48 hours in ATCC-formulated RPMI media. Two days post-transfection, the media was changed and the transfected cells were selected for using 1µg puromycin per 1mL media. A piggyBac transposon plasmid containing two sgRNAs (SVAsgRNA1: 5'CTCCCTAATCTCAAGTACCC 3' and SVAsgRNA2: 5' TGTTTCAGAGAGCACGGGGT 3'; Integrated DNA Technologies) targeting ~80% of all annotated SVAs in humans (Pontis et al. 2019), along with a piggyBac transposase, were transfected into the NCCIT-dCas9KRAB cells using a 6:1 ratio of Fugene HD for 48 hours in ATCC-formulated RPMI media. Two days post-transfection, the media was changed and the transfected cells were selected for using 400µg geneticin per 1mL of media in addition to 1µg puromycin per 1mL media. The NCCIT-dCas9KRAB-SVAsgRNA cell line was maintained in ATCC-formulated RPMI media supplemented with 10% tet-free FBS, 1% L-glutamine, 1µg/mL puromycin, and 400µg/mL geneticin and incubated at 5% CO_2, 20% O_2 at 37°C.

Retinoic Acid-induced neuronal differentiation and CRISPR-interference of NCCIT-dCas9KRAB-SVAsgRNA cells

The NCCIT-dCas9KRAB-SVAsgRNA cells, at ~20% confluency, were treated with 10µM retinoic acid (RA) per 10mL media for 1 week to induce neuronal differentiation. At day 4, the media was refreshed and the cells were additionally treated with 2µg doxycycline per 1mL of media for 3 days. The cells were collected on day 7 of RA treatment (day 3 of doxycycline treatment) for qPCR and genomic experiments. Expression of the doxycycline-inducible dCas9 was verified via western blot.

Statistical and genomic analyses

All statistical analyses were performed using R v3.3.1 or Graphpad Prism version 9.2.0 for Mac OS X . BEDtools v2.27.1 (Quinlan and Hall, 2010) was used for genomic analyses. Pathway analysis was performed with Ingenuity Pathway Analysis Suite (QIAGEN Inc., https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis). Motif analyses were performed using the MEME-Suite (Bailey et al., 2015), and specifically with the Meme-ChIP application. Fasta files of the regions of interest were produced using BEDTools v2.27.1. Shuffled input sequences were used as background. E-values < 0.001 were used as threshold for significance. Orthologous ATAC-seq regions were identified using the University of California Santa Cruz Genome Browser tool LiftOver.

Acknowledgements

The authors are grateful for Dr. Yoav Gilad (University of Chicago) for providing all the chimpanzee lines and one of the human lines, and to Dr. Joanna Wysocka's lab, and in particular Dr. Raquel Fueyo (Stanford University), for providing CRISPR reagents and critical advice. The authors thank the Thomas Jefferson Stem Cell and Regenerative Neuroscience Center for the support in the process of optimization of the iPSC differentiation. The authors are grateful to Trizzino lab members Luca Pagliaroli, PhD, Chiara Scopa, PhD and Connor Ott for help with specific analyses and experiments. The authors thank the Genomic Facility at The Wistar Institute (Philadelphia, PA) for the Next Generation Illumina Sequencing. The PAX6 antibody developed by Kawakami, A. was obtained from the Developmental Hybridoma Bank, created by the NICHD of the NIH and maintained at The University of Iowa, Department of Biology, Iowa City, IA 52242.

Competing interests

The authors declare no competing interests.

Funding
For this work, M.T. was funded by the National Institute of Health (NIH-NIGMS R35GM138344) and by the G. Harold and Leila Y. Mathers Foundation.

Data availability
The original genome-wide data generated in this study have been deposited in the GEO database under accession code GSE189347.

Author contributions
MT and SP designed the project. SP performed most of the experiments. SB carried out the CRISPR experiments in NCCIT cells. MT and SP analyzed the data and wrote the manuscript. All the authors read and approved the manuscript.

References

Abrajano, J. J., Qureshi, I. A., Gokhan, S., Zheng, D., Bergman, A. and Mehler, M. F. (2009). REST and CoREST Modulate Neuronal Subtype Specification, Maturation and Maintenance. *PLOS ONE* 4, e7936.

Agoglia, R. M., Sun, D., Birey, F., Yoon, S.-J., Miura, Y., Sabatini, K., Pașca, S. P. and Fraser, H. B. (2021). Primate cell fusion disentangles gene regulatory divergence in neurodevelopment. *Nature* 592, 421–427.

Andersen, J., Urbán, N., Achimastou, A., Ito, A., Simic, M., Ullom, K., Martynoga, B., Lebel, M., Göritz, C., Frisén, J., et al. (2014). A Transcriptional Mechanism Integrating Inputs from Extracellular Signals to Activate Hippocampal Stem Cells. *Neuron* 83, 1085–1097.

Arnold, S. J., Huang, G.-J., Cheung, A. F. P., Era, T., Nishikawa, S.-I., Bikoff, E. K., Molnar, Z., Robertson, E. J. and Groszer, M. (2008). The T-box transcription factor Eomes/Tbr2 regulates neurogenesis in the cortical subventricular zone. *Genes & Development* 22, 2479–2484.

Aruga, J. (2004). The role of Zic genes in neural development. *Mol Cell Neurosci* 26, 205–221.

Averaimo, S., Gritti, M., Barini, E., Gasparini, L. and Mazzanti, M. (2014). CLIC1 functional expression is required for cAMP-induced neurite elongation in post-natal mouse retinal ganglion cells. *Journal of Neurochemistry* 131, 444–456.

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

Barger, N., Hanson, K. L., Teffer, K., Schenker-Ahmed, N. M. and Semendeferi, K. (2014). Evidence for evolutionary specialization in human limbic structures. *Front. Hum. Neurosci.* 8,.

Bath, K. G., Akins, M. R. and Lee, F. S. (2012). BDNF control of adult SVZ neurogenesis. *Developmental Psychobiology* 54, 578–589.

Bekinschtein, P., Katche, C., Slipczuk, L. N., Igaz, L. M., Cammarota, M., Izquierdo, I. and Medina, J. H. (2007). mTOR signaling in the hippocampus is necessary for memory formation. *Neurobiology of Learning and Memory* 87, 303–307.

654 Bray, N. L., Pimentel, H., Melsted, P. and Pachter, L. (2016). Near-optimal probabilistic RNA-seq
655 quantification. *Nat Biotechnol* 34, 525–527.

656 Bulfone, A., Martinez, S., Marigo, V., Campanella, M., Basile, A., Quaderi, N., Gattuso, C., Rubenstein, J.
657 L. R. and Ballabio, A. (1999). Expression pattern of the Tbr2 (Eomesodermin) gene during mouse
658 and chick brain development. *Mechanisms of Development* 84, 133–138.

659 Burgess, N., Maguire, E. A. and O'Keefe, J. (2002). The Human Hippocampus and Spatial and Episodic
660 Memory. *Neuron* 35, 625–641.

661 Chen, X., Wang, S., Zhou, Y., Han, Y., Li, S., Xu, Q., Xu, L., Zhu, Z., Deng, Y., Yu, L., et al. (2018). Phf8
662 histone demethylase deficiency causes cognitive impairments through the mTOR pathway. *Nat*
663 *Commun* 9, 114.

664 Cheval, H., Chagneau, C., Levasseur, G., Veyrac, A., Faucon-Biguet, N., Laroche, S. and Davis, S. (2012).
665 Distinctive features of Egr transcription factor regulation and DNA binding activity in CA1 of the
666 hippocampus in synaptic plasticity and consolidation and reconsolidation of fear memory.
667 *Hippocampus* 22, 631–642.

668 Chuong, E. B., Rumi, M. A. K., Soares, M. J. and Baker, J. C. (2013). Endogenous retroviruses function as
669 species-specific enhancer elements in the placenta. *Nat Genet* 45, 325–329.

670 Chuong, E. B., Elde, N. C. and Feschotte, C. (2016). Regulatory evolution of innate immunity through co-
671 option of endogenous retroviruses. *Science* 351, 1083–1087.

672 Cipriani, S., Nardelli, J., Verney, C., Delezoide, A.-L., Guimiot, F., Gressens, P. and Adle-Biassette, H.
673 (2016). Dynamic Expression Patterns of Progenitor and Pyramidal Neuron Layer Markers in the
674 Developing Human Hippocampus. *Cereb. Cortex* 26, 1255–1271.

675 Coluccio, A., Ecco, G., Duc, J., Offner, S., Turelli, P. and Trono, D. (2018). Individual retrotransposon
676 integrants are differentially controlled by KZFP/KAP1-dependent histone methylation, DNA
677 methylation and TET-mediated hydroxymethylation in naïve embryonic stem cells. *Epigenetics &*
678 *Chromatin* 11, 7.

679 Cosby, R. L., Judd, J., Zhang, R., Zhong, A., Garry, N., Pritham, E. J. and Feschotte, C. (2021). Recurrent
680 evolution of vertebrate transcription factors by transposase capture. *Science* 371, eabc6405.

681 Coufal, N. G., Garcia-Perez, J. L., Peng, G. E., Yeo, G. W., Mu, Y., Lovci, M. T., Morell, M., O'Shea, K. S.,
682 Moran, J. V. and Gage, F. H. (2009). L1 retrotransposition in human neural progenitor cells.
683 *Nature* 460, 1127–1131.

684 Crawley, J. N., Heyer, W.-D. and LaSalle, J. M. (2016). Autism and Cancer Share Risk Genes, Pathways,
685 and Drug Targets. *Trends in Genetics* 32, 139–146.

686 Datson, N. a., Morsink, M. c., Steenbergen, P. j., Aubert, Y., Schlumbohm, C., Fuchs, E. and de Kloet, E. r.
687 (2009). A molecular blueprint of gene expression in hippocampal subregions CA1, CA3, and DG is
688 conserved in the brain of the common marmoset. *Hippocampus* 19, 739–752.

689 Davis, L., Onn, I. and Elliott, E. (2018). The emerging roles for the chromatin structure regulators CTCF
690 and cohesin in neurodevelopment and behavior. *Cell. Mol. Life Sci.* 75, 1205–1214.

691 del Rosario, R. C. H., Rayan, N. A. and Prabhakar, S. (2014). Noncoding origins of anthropoid traits and a
692 new null model of transposon functionalization. *Genome Res* 24, 1469–1484.

693 Du, J., Leung, A., Trac, C., Lee, M., Parks, B. W., Lusis, A. J., Natarajan, R. and Schones, D. E. (2016).
694 Chromatin variation associated with liver metabolism is mediated by transposable elements.
695 *Epigenetics & Chromatin* 9, 28.

696 Duyckaerts, C., Delatour, B. and Potier, M.-C. (2009). Classification and basic pathology of Alzheimer
697 disease. *Acta Neuropathol* 118, 5–36.

698 Edler, M. K., Sherwood, C. C., Meindl, R. S., Hopkins, W. D., Ely, J. J., Erwin, J. M., Mufson, E. J., Hof, P. R.
699 and Raghanti, M. A. (2017). Aged chimpanzees exhibit pathologic hallmarks of Alzheimer's
700 disease. *Neurobiology of Aging* 59, 107–120.

701 Eichenbaum, H. (2017a). The role of the hippocampus in navigation is memory. *Journal of*
702 *Neurophysiology* 117, 1785–1796.

703 Eichenbaum, H. (2017b). Prefrontal–hippocampal interactions in episodic memory. *Nat Rev Neurosci* 18,
704 547–558.

705 Emera, D., Yin, J., Reilly, S. K., Gockley, J. and Noonan, J. P. (2016). Origin and evolution of
706 developmental enhancers in the mammalian neocortex. *PNAS* 113, E2617–E2626.

707 Enard, W. (2011). FOXP2 and the role of cortico-basal ganglia circuits in speech and language evolution.
708 *Current Opinion in Neurobiology* 21, 415–424.

709 Enard, W., Khaitovich, P., Klose, J., Zöllner, S., Heissig, F., Giavalisco, P., Nieselt-Struwe, K., Muchmore,
710 E., Varki, A., Ravid, R., et al. (2002). Intra- and Interspecific Variation in Primate Gene Expression
711 Patterns. *Science* 296, 340–343.

712 Englund, C. (2005). Pax6, Tbr2, and Tbr1 Are Expressed Sequentially by Radial Glia, Intermediate
713 Progenitor Cells, and Postmitotic Neurons in Developing Neocortex. *Journal of Neuroscience* 25,
714 247–251.

715 Ferrón, S. R., Charalambous, M., Radford, E., McEwen, K., Wildner, H., Hind, E., Morante-Redolat, J. M.,
716 Laborda, J., Guillemot, F., Bauer, S. R., et al. (2011). Postnatal loss of Dlk1 imprinting in stem
717 cells and niche astrocytes regulates neurogenesis. *Nature* 475, 381–385.

718 Finch, C. E. and Austad, S. N. (2015). Commentary: is Alzheimer's disease uniquely human? *Neurobiology*
719 *of Aging* 36, 553–555.

720 Florio, M. and Huttner, W. B. (2014). Neural progenitors, neurogenesis and the evolution of the
721 neocortex. *Development* 141, 2182–2194.

722 Fortress, A. M., Fan, L., Orr, P. T., Zhao, Z. and Frick, K. M. (2013). Estradiol-induced object recognition
723 memory consolidation is dependent on activation of mTOR signaling in the dorsal hippocampus.
724 *Learn. Mem.* 20, 147–155.

725 Fuentes, D. R., Swigut, T. and Wysocka, J. (2018). Systematic perturbation of retroviral LTRs reveals
726 widespread long-range effects on human gene regulation. *eLife* 7, e35989.

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

730 Gokhman, D., Agoglia, R. M., Kinnebrew, M., Gordon, W., Sun, D., Bajpai, V. K., Naqvi, S., Chen, C., Chan,
731 A., Chen, C., et al. (2021). Human–chimpanzee fused cells reveal cis -regulatory divergence
732 underlying skeletal evolution. *Nature Genetics* 53, 467–476.

733 Graber, T. E., McCamphill, P. K. and Sossin, W. S. (2013). A recollection of mTOR signaling in learning and
734 memory. *Learn. Mem.* 20, 518–530.

735 Hevner, R. F. (2019). Intermediate progenitors and Tbr2 in cortical development. *J. Anat.* 235, 616–625.

736 Hickey, S. L., Berto, S. and Konopka, G. (2019). Chromatin Decondensation by FOXP2 Promotes Human
737 Neuron Maturation and Expression of Neurodevelopmental Disease Genes. *Cell Reports* 27,
738 1699-1711.e9.

739 Hodge, R. D., Nelson, B. R., Kahoud, R. J., Yang, R., Mussar, K. E., Reiner, S. L. and Hevner, R. F. (2012).
740 Tbr2 Is Essential for Hippocampal Lineage Progression from Neural Stem Cells to Intermediate
741 Progenitors and Neurons. *Journal of Neuroscience* 32, 6275–6287.

742 Ito, J., Kimura, I., Soper, A., Coudray, A., Koyanagi, Y., Nakaoka, H., Inoue, I., Turelli, P., Trono, D. and
743 Sato, K. Endogenous retroviruses drive KRAB zinc-finger protein family expression for tumor
744 suppression. *Science Advances* 6, eabc3020.

745 Ivancevic, A. and Chuong, E. B. (2020). Transposable elements teach T cells new tricks. *PNAS* 117, 9145–
746 9147.

747 Jacques, P.-É., Jeyakani, J. and Bourque, G. (2013). The Majority of Primate-Specific Regulatory
748 Sequences Are Derived from Transposable Elements. *PLOS Genetics* 9, e1003504.

749 Jönsson, M. E., Ludvik Brattås, P., Gustafsson, C., Petri, R., Yudovich, D., Pircs, K., Verschuere, S.,
750 Madsen, S., Hansson, J., Larsson, J., et al. (2019). Activation of neuronal genes via LINE-1
751 elements upon global DNA demethylation in human neural progenitors. *Nat Commun* 10, 3182.

752 Judd, J., Sanderson, H. and Feschotte, C. (2021). Evolution of mouse circadian enhancers from
753 transposable elements. *Genome Biol* 22, 193.

754 Kamboh, M. I., Fan, K.-H., Yan, Q., Beer, J. C., Snitz, B. E., Wang, X., Chang, C.-C. H., Demirci, F. Y.,
755 Feingold, E. and Ganguli, M. (2019). Population-based genome-wide association study of
756 cognitive decline in older adults free of dementia: identification of a novel locus for the
757 attention domain. *Neurobiol Aging* 84, 239.e15-239.e24.

758 Kimura, N., Nakashima, K., Ueno, M., Kiyama, H. and Taga, T. (1999). A novel mammalian T-box-
759 containing gene, Tbr2, expressed in mouse developing brain. *Developmental Brain Research*
760 115, 183–193.

761 King, M. and Wilson, A. (1975). Evolution at two levels in humans and chimpanzees. *Science* 188, 107–
762 116.

763 Kittappa, R., Chang, W. W., Awatramani, R. B. and McKay, R. D. G. (2007). The foxa2 Gene Controls the
764 Birth and Spontaneous Degeneration of Dopamine Neurons in Old Age. *PLOS Biology* 5, e325.

765 Kohane, A. A. and Wood, T. R. (2021). *Neurodevelopmental clustering of gene expression identifies lipid*
766 *metabolism genes associated with neuroprotection and neurodegeneration*.

767 Kolehmainen, J., Black, G. C. M., Saarinen, A., Chandler, K., Clayton-Smith, J., Träskelin, A.-L., Perveen, R.,
768 Kivitie-Kallio, S., Norio, R., Warburg, M., et al. (2003). Cohen syndrome is caused by mutations in
769 a novel gene, COH1, encoding a transmembrane protein with a presumed role in vesicle-
770 mediated sorting and intracellular protein transport. *Am J Hum Genet* 72, 1359–1369.

771 Liégeois, F. J., Hildebrand, M. S., Bonthrone, A., Turner, S. J., Scheffer, I. E., Bahlo, M., Connelly, A. and
772 Morgan, A. T. (2016). Early neuroimaging markers of FOXP2 intragenic deletion. *Sci Rep* 6,
773 35192.

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

776 Lynch, V. J., Leclerc, R. D., May, G. and Wagner, G. P. (2011). Transposon-mediated rewiring of gene
777 regulatory networks contributed to the evolution of pregnancy in mammals. *Nat Genet* 43,
778 1154–1159.

779 Lynch, V. J., Nnamani, M. C., Kapusta, A., Brayer, K., Plaza, S. L., Mazur, E. C., Emera, D., Sheikh, S. Z.,
780 Grützner, F., Bauersachs, S., et al. (2015). Ancient Transposable Elements Transformed the
781 Uterine Regulatory Landscape and Transcriptome during the Evolution of Mammalian
782 Pregnancy. *Cell Rep* 10, 551–561.

783 MacDermot, K. D., Bonora, E., Sykes, N., Coupe, A.-M., Lai, C. S. L., Vernes, S. C., Vargha-Khadem, F.,
784 McKenzie, F., Smith, R. L., Monaco, A. P., et al. (2005). Identification of FOXP2 Truncation as a
785 Novel Cause of Developmental Speech and Language Deficits. *The American Journal of Human*
786 *Genetics* 76, 1074–1080.

787 Mandal, C., Jung, K. H., Kang, S. C., Choi, M. R., Park, K. S., Chung, I. Y. and Chai, Y. G. (2015). Knocking
788 down of UTX in NCCIT cells enhance cell attachment and promote early neuronal cell
789 differentiation. *BioChip J* 9, 182–193.

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

793 Martínez-Cerdeño, V., Noctor, S. C. and Kriegstein, A. R. (2006). The role of intermediate progenitor cells
794 in the evolutionary expansion of the cerebral cortex. *Cereb Cortex* 16 Suppl 1, i152-161.

795 Mathys, H., Davila-Velderrain, J., Peng, Z., Gao, F., Mohammadi, S., Young, J. Z., Menon, M., He, L.,
796 Abdurrob, F., Jiang, X., et al. (2019). Single-cell transcriptomic analysis of Alzheimer's disease.
797 *Nature* 570, 332–337.

798 Miao, B., Fu, S., Lyu, C., Gontarz, P., Wang, T. and Zhang, B. (2020). Tissue-specific usage of transposable
799 element-derived promoters in mouse development. *Genome Biol* 21, 255.

800 Mihalas, A. B., Elsen, G. E., Bedogni, F., Daza, R. A. M., Ramos-Laguna, K. A., Arnold, S. J. and Hevner, R.
801 F. (2016). Intermediate Progenitor Cohorts Differentially Generate Cortical Layers and Require
802 Tbr2 for Timely Acquisition of Neuronal Subtype Identity. *Cell Reports* 16, 92–105.

803 Mika, K., Marinić, M., Singh, M., Muter, J., Brosens, J. J. and Lynch, V. J. (2021). Evolutionary
804 transcriptomics implicates new genes and pathways in human pregnancy and adverse
805 pregnancy outcomes. *eLife* 10, e69584.

806 Modzelewski, A. J., Shao, W., Chen, J., Lee, A., Qi, X., Noon, M., Tjokro, K., Sales, G., Biton, A., Anand, A.,
807 et al. (2021). A mouse-specific retrotransposon drives a conserved Cdk2ap1 isoform essential for
808 development. *Cell* 184, 5541-5558.e22.

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

812 Mukherjee, D., Gonzales, B. J., Ashwal-Fluss, R., Turm, H., Groysman, M. and Citri, A. (2021). Egr2
813 induction in spiny projection neurons of the ventrolateral striatum contributes to cocaine place
814 preference in mice. *eLife* 10, e65228.

815 Okhovat, M., Nevonen, K. A., Davis, B. A., Michener, P., Ward, S., Milhaven, M., Harshman, L., Sohota, A.,
816 Fernandes, J. D., Salama, S. R., et al. (2020). Co-option of the lineage-specific LAVA
817 retrotransposon in the gibbon genome. *PNAS* 117, 19328–19338.

818 Pagliaroli, L., Porazzi, P., Curtis, A. T., Scopa, C., Mikkers, H. M. M., Freund, C., Daxinger, L., Deliard, S.,
819 Welsh, S. A., Offley, S., et al. (2021). Inability to switch from ARID1A-BAF to ARID1B-BAF impairs
820 exit from pluripotency and commitment towards neural crest formation in ARID1B-related
821 neurodevelopmental disorders. *Nat Commun* 12, 6469.

822 Pashos, E. E., Park, Y., Wang, X., Raghavan, A., Yang, W., Abbey, D., Peters, D. T., Arbelaez, J., Hernandez,
823 M., Kuperwasser, N., et al. (2017). Large, Diverse Population Cohorts of hiPSCs and Derived
824 Hepatocyte-like Cells Reveal Functional Genetic Variation at Blood Lipid-Associated Loci. *Cell*
825 *Stem Cell* 20, 558-570.e10.

826 Playfoot, C. J., Duc, J., Sheppard, S., Dind, S., Coudray, A., Planet, E. and Trono, D. (2021). Transposable
827 elements and their KZFP controllers are drivers of transcriptional innovation in the developing
828 human brain. *Genome Res.* 31, 1531–1545.

829 Poirier, R., Cheval, H., Mailhes, C., Charnay, P., Davis, S. and Laroche, S. (2007). Paradoxical Role of an
830 Egr Transcription Factor Family Member, Egr2/Krox20, in Learning and Memory. *Front Behav*
831 *Neurosci* 1, 6.

832 Pontious, A., Kowalczyk, T., Englund, C. and Hevner, R. F. (2008). Role of intermediate progenitor cells in
833 cerebral cortex development. *Dev Neurosci* 30, 24–32.

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

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

840 Quinn, J. P. and Bubb, V. J. (2014). SVA retrotransposons as modulators of gene expression. *Mobile*
841 *Genetic Elements* 4, e32102.

842 Raivich, G. (2008). c-Jun expression, activation and function in neural cell death, inflammation and
843 repair. *J Neurochem* 107, 898–906.

844 Raivich, G. and Behrens, A. (2006). Role of the AP-1 transcription factor c-Jun in developing, adult and
845 injured brain. *Prog Neurobiol* 78, 347–363.

846 Rayan, N. A., Del Rosario, R. C. H. and Prabhakar, S. (2016). Massive contribution of transposable
847 elements to mammalian regulatory sequences. *Semin Cell Dev Biol* 57, 51–56.

848 Reilly, S. K., Yin, J., Ayoub, A. E., Emera, D., Leng, J., Cotney, J., Sarro, R., Rakic, P. and Noonan, J. P.
849 (2015). Evolutionary genomics. Evolutionary changes in promoter and enhancer activity during
850 human corticogenesis. *Science* 347, 1155–1159.

851 Rétaux, S., Bourrat, F., Joly, J.-S. and Hinaux, H. (2013). Perspectives in Evo-Devo of the Vertebrate Brain.
852 In *Advances in Evolutionary Developmental Biology*, pp. 151–172. John Wiley & Sons, Ltd.

853 Reuter, M. S., Krumbiegel, M., Schlüter, G., Ekici, A. B., Reis, A. and Zweier, C. (2017). Haploinsufficiency
854 of NR4A2 is associated with a neurodevelopmental phenotype with prominent language
855 impairment. *Am J Med Genet A* 173, 2231–2234.

856 Sakamoto, K., Karelina, K. and Obrietan, K. (2011). CREB: a multifaceted regulator of neuronal plasticity
857 and protection. *Journal of Neurochemistry* 116, 1–9.

858 Sarkar, A., Mei, A., Paquola, A. C. M., Stern, S., Bardy, C., Klug, J. R., Kim, S., Neshat, N., Kim, H. J., Ku, M.,
859 et al. (2018). Efficient Generation of CA3 Neurons from Human Pluripotent Stem Cells Enables
860 Modeling of Hippocampal Connectivity In Vitro. *Cell Stem Cell* 22, 684-697.e9.

861 Schmidt, D., Schwalie, P. C., Wilson, M. D., Ballester, B., Gonçalves, A., Kutter, C., Brown, G. D., Marshall,
862 A., Flicek, P. and Odom, D. T. (2012). Waves of retrotransposon expansion remodel genome
863 organization and CTCF binding in multiple mammalian lineages. *Cell* 148, 335–348.

864 Sessa, A., Mao, C.-A., Hadjantonakis, A.-K., Klein, W. H. and Broccoli, V. (2008). Tbr2 directs conversion
865 of radial glia into basal precursors and guides neuronal amplification by indirect neurogenesis in
866 the developing neocortex. *Neuron* 60, 56–69.

867 Shah, A. H., Gilbert, M., Ivan, M. E., Komotar, R. J., Heiss, J. and Nath, A. (2021). The role of human
868 endogenous retroviruses in gliomas: from etiological perspectives and therapeutic implications.
869 *Neuro-Oncology* 23, 1647–1655.

870 Sin, C., Li, H. and Crawford, D. A. (2015). Transcriptional Regulation by FOXP1, FOXP2, and FOXP4
871 Dimerization. *Journal of Molecular Neuroscience* 55, 437–448.

872 Sousa, A. M. M., Zhu, Y., Raghanti, M. A., Kitchen, R. R., Onorati, M., Tebbenkamp, A. T. N., Stutz, B.,
873 Meyer, K. A., Li, M., Kawasawa, Y. I., et al. (2017). Molecular and cellular reorganization of neural
874 circuits in the human lineage. *Science* 358, 1027–1032.

875 Squire, L. R. (1992). Memory and the Hippocampus: A Synthesis From Findings With Rats, Monkeys, and
876 Humans.

877 Sundaram, V. and Wysocka, J. (2020). Transposable elements as a potent source of diverse cis-
878 regulatory sequences in mammalian genomes. *Philosophical Transactions of the Royal Society B:*
879 *Biological Sciences* 375, 20190347.

880 Sundaram, V., Cheng, Y., Ma, Z., Li, D., Xing, X., Edge, P., Snyder, M. P. and Wang, T. (2014). Widespread
881 contribution of transposable elements to the innovation of gene regulatory networks. *Genome*
882 *Res* 24, 1963–1976.

883 Sur, D., Kustwar, R. K., Budania, S., Mahadevan, A., Hancks, D. C., Yadav, V., Shankar, S. K. and Mandal, P.
884 K. (2017). Detection of the LINE-1 retrotransposon RNA-binding protein ORF1p in different
885 anatomical regions of the human brain. *Mobile DNA* 8, 17.

886 Taher, N., McKenzie, C., Garrett, R., Baker, M., Fox, N. and Isaacs, G. D. (2014). Amyloid-β Alters the DNA
887 Methylation Status of Cell-fate Genes in an Alzheimer's Disease Model. *Journal of Alzheimer's*
888 *Disease* 38, 831–844.

889 Thomas, C. A., Paquola, A. C. M. and Muotri, A. R. (2012). LINE-1 Retrotransposition in the Nervous
890 System. *Annual Review of Cell and Developmental Biology* 28, 555–573.

891 Tokuyama, M., Kong, Y., Song, E., Jayewickreme, T., Kang, I. and Iwasaki, A. (2018). ERVmap analysis
892 reveals genome-wide transcription of human endogenous retroviruses. *PNAS* 115, 12565–
893 12572.

894 Tomasello, M. and Herrmann, E. (2010). Ape and Human Cognition: What's the Difference? *Current*
895 *Directions in Psychological Science* 19, 3–8.

896 Trizzino, M., Park, Y., Holsbach-Beltrame, M., Aracena, K., Mika, K., Caliskan, M., Perry, G. H., Lynch, V. J.
897 and Brown, C. D. (2017). Transposable elements are the primary source of novelty in primate
898 gene regulation. *Genome Res.*

899 Trizzino, M., Kapusta, A. and Brown, C. D. (2018). Transposable elements generate regulatory novelty in
900 a tissue-specific fashion. *BMC Genomics* 19, 468.

901 Turelli, P., Playfoot, C., Grun, D., Raclot, C., Pontis, J., Coudray, A., Thorball, C., Duc, J., Pankevich, E. V.,
902 Deplancke, B., et al. (2020). Primate-restricted KRAB zinc finger proteins and target
903 retrotransposons control gene expression in human neurons. *Science Advances* 6, eaba3200.

904 Walker, L. C. and Jucker, M. (2017). The Exceptional Vulnerability of Humans to Alzheimer's Disease.
905 *Trends in Molecular Medicine* 23, 534–545.

906 Wang, H., Xing, J., Grover, D., Hedges, D. J., Han, K., Walker, J. A. and Batzer, M. A. (2005). SVA Elements:
907 A Hominid-specific Retroposon Family. *Journal of Molecular Biology* 354, 994–1007.

908 Ward, M. C., Zhao, S., Luo, K., Pavlovic, B. J., Karimi, M. M., Stephens, M. and Gilad, Y. (2018). Silencing
909 of transposable elements may not be a major driver of regulatory evolution in primate iPSCs.
910 *eLife* 7, e33084.

911 Wei, Y. (2020). Comparative transcriptome analysis of the hippocampus from sleep-deprived and
912 Alzheimer's disease mice. *Genet Mol Biol* 43, e20190052.

913 Wray, G. A. (2007). The evolutionary significance of cis-regulatory mutations. *Nat Rev Genet* 8, 206–216.

914 Xie, M., Hong, C., Zhang, B., Lowdon, R. F., Xing, X., Li, D., Zhou, X., Lee, H. J., Maire, C. L., Ligon, K. L., et
915 al. (2013). DNA hypomethylation within specific transposable element families associates with
916 tissue-specific enhancer landscape. *Nat Genet* 45, 836–841.

917 Yang, W., Liu, Y., Slovik, K. J., Wu, J. C., Duncan, S. A., Rader, D. J. and Morrisey, E. E. (2015). Generation
918 of iPSCs as a Pooled Culture Using Magnetic Activated Cell Sorting of Newly Reprogrammed
919 Cells. *PLoS One* 10, e0134995.

920 Yi, R., Chen, B., Zhao, J., Zhan, X., Zhang, L., Liu, X. and Dong, Q. (2014). Krüppel-like Factor 8 Ameliorates
921 Alzheimer's Disease by Activating β-Catenin. *J Mol Neurosci* 52, 231–241.

922 Yu, D. X., Di Giorgio, F. P., Yao, J., Marchetto, M. C., Brennand, K., Wright, R., Mei, A., Mchenry, L., Lisuk,
923 D., Grasmick, J. M., et al. (2014). Modeling Hippocampal Neurogenesis Using Human Pluripotent
924 Stem Cells. *Stem Cell Reports* 2, 295–310.

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

928 Zhang, J., Li, H., Yabut, O., Fitzpatrick, H., D'Arcangelo, G. and Herrup, K. (2010). Cdk5 Suppresses the
929 Neuronal Cell Cycle by Disrupting the E2F1–DP1 Complex. *J. Neurosci.* 30, 5219–5228.

930 Zhang, H., Xue, C., Shah, R., Bermingham, K., Hinkle, C. C., Li, W., Rodrigues, A., Tabita-Martinez, J.,
931 Millar, J. S., Cuchel, M., et al. (2015). Functional Analysis and Transcriptomic Profiling of iPSC-
932 Derived Macrophages and Their Application in Modeling Mendelian Disease. *Circulation*
933 *Research* 117, 17–28.

934 Zhao, B., Wu, Q., Ye, A. Y., Guo, J., Zheng, X., Yang, X., Yan, L., Liu, Q.-R., Hyde, T. M., Wei, L., et al.
935 (2019). Somatic LINE-1 retrotransposition in cortical neurons and non-brain tissues of Rett
936 patients and healthy individuals. *PLOS Genetics* 15, e1008043.

937
938
939 **Figure legends**
940
941 **Figure 1** – Differentiation of human and chimpanzee iPSCs into hippocampal intermediate progenitors (A)
942 Western Blot of human differentiated iPSCs against TBR2 (73kDa) at day 0 (iPSCs), day 5, and day 11 of
943 treatment with hpIPC differentiation media. GAPDH is shown as the loading control. (B)
944 Immunofluorescent staining of human and chimpanzee hpIPCs after five days against
945 neurodevelopmental markers PAX6, TBR2 and OTX2. Nuclei are counterstained with DAPI.
946
947 **Figure 2** – Differential Gene Expression In Human and Chimpanzee hpIPCs. (A) Schematic of iPSC
948 differentiation followed by RNA-seq library generation and analysis. (B) Volcano plot depicting "Human
949 UP" genes (right) and "Chimp UP" genes (left) with a log2FoldChange threshold of 1.5 and -1.5,
950 respectively, and a p-value threshold of 0.05. (C) Top upstream regulators of the differentially expressed
951 genes predicted by Ingenuity Pathway Analysis, ranked by -log10(p-value). (D) Heatmaps depicting the
952 expression of genes predicted to be under the control of the transcription factors CREB1, FOXA2 and TBR2
953 or predicted to play a role in embryonic development. The rows of the heatmaps indicate the transcript
954 names, columns indicate the species, sample number and replicate. "HS1 Rep 1" indicates Homo sapiens
955 sample 1 replicate 1 and "PT2 Rep 2" indicates Pan troglodytes sample 2 replicate 2
956
957 **Figure 3** – Human Centric Chromatin Accessibility Analysis. (A) Schematic of iPSC differentiation followed
958 by ATAC-seq library generation and analysis. (B) UCSC Genome Browser visualization of human ATAC-seq
959 libraries (HS Rep 1, Rep 2, Rep 3) aligned to hg19 genome assembly. (C) Distribution of the differentially
960 accessible (DA) peaks into enhancers or promoters, and into Human UP (greater accessibility in humans)
961 or Chimp UP (greater accessibility in chimpanzees). (D) Differentially accessible chromatin regions ($p <$
962 0.05, $n = 3006$) are more likely to overlap with transposable elements (TE) than non-differentially
963 accessible regions ($p > 0.9$, $n = 3{,}006$). (E) The 1335 TE-derived DA regions are more likely to be near a
964 differentially expressed gene, as compared to DA regions that do not overlap a TE. (F) Distribution of the
965 five major TE classes in the human genome, compared with their distribution among the TE-derived DA
966 chromatin regions, and among the TE-derived DA chromatin regions proximal to a differentially expressed
967 gene
968
969 **Figure 4** – Chimpanzee Centric Chromatin Accessibility Analysis. (A) Schematic of iPSC differentiation
970 followed by ATAC-seq library generation and analysis. (B) UCSC Genome Browser visualization of human
971 ATAC-seq libraries (PT Rep 1, Rep 2, Rep 3) aligned to hg19 genome assembly. (C) Distribution of the
972 differentially accessible (DA) peaks into enhancers or promoters, and into Human UP (greater accessibility
973 in humans) or Chimp UP (greater accessibility in chimpanzees). (D) Differentially accessible chromatin
974 regions ($p < 0.05$, $n = 3{,}806$) are more likely to overlap with transposable elements than other accessible
975 regions ($p > 0.9$, $n = 3806$. (E) Distribution of DA regions that overlap a TE and proximity to genes
976 differentially expressed between human and chimpanzee hpIPCs. (F) Distribution of the five major TE
977 classes in the chimpanzee genome, compared with their distribution among the 1410 TE-derived DA
978 chromatin regions (G) Breakdown of the 20 Miscellanous TEs represented in the 1,410 TE-derived DA
979 chromatin regions. (H) Distribution of SVA family of TEs in the human genome as compared to their
980 distribution in the TE-derived DA chromatin regions ($n = 16$).
981
982 **Figure 5** – LTRs Are Enriched Among Human and Chimpanzee Differentially Accessible Transposons. (A)
983 Distribution of LTRs compared to non-LTRs in the human genome and in the 1,335 human TE-derived DA
984 regions, with predicted binding motifs. (B) Differentially expressed genes close to the human-enriched DA

LTRs. (C) Distribution of LTRs compared to non-LTRs in the chimpanzee genome and in the 1,410 chimpanzee TE-derived DA regions, with predicted binding motifs. (D) Differentially expressed genes close to the chimpanzee-enriched DA LTRs.

Figure 6 – Human-specific SVAs Bind Tbr2 and Influence Neurodevelopment (A) The SVAs enriched within the differentially accessible chimpanzee peaks are enriched for binding motifs of neurodevelopmental transcription factors KLF8, ZIC1, and ASCL1 as well as the AP-1 dimer. (B) Human SVAs are enriched for the binding motif of TBR2. (C) Genes proximal to human-specific SVAs are predicted to be regulated by neurodevelopmental TFs and function in important neuronal pathways. (D) 1,816 human SVAs exhibit ATAC-seq signal in hpIPCs (red = signal, white = noise) and of these, 739 exhibit TBR2 ChIP signal (black = signal, yellow = background). (E) Breakdown of 739 accessible, TBR2 bound SVAs into "human-specific" and "conserved in chimpanzees". (F) Differentially expressed genes close to the human-specific TBR2-bound SVAs.

Figure 7 – CRISPR-interference on human SVAs. (A) Treating NCCITs with Retinoic Acid for 7 days leads to induction of TBR2 and PAX6, as shown in qRT-PCR. (B) Treating the RA-NCCIT stable CRISPR line for 72 hours with doxycycline leads to dCas9 activation, as shown in the Cas9 immunoblot. (C) dCas9 activation results in deposition of H3K9me3 at most human SVAs as shown in the H3K9me3 ChIP-seq heatmap. (D) Volcano plot depicting genes differentially expressed upon doxycycline treatment. (E) Venn diagram showing overlap between genes differentially expressed between human and chimpanzee's hippocampal intermediate progenitors and between RA-NCCITs with and without doxycycline treatment (i.e. with and without SVA repression). (F) Pie chart illustrating the fraction of TBR2-controlled genes that are differentially expressed upon SVA repression in RA-NCCITs.

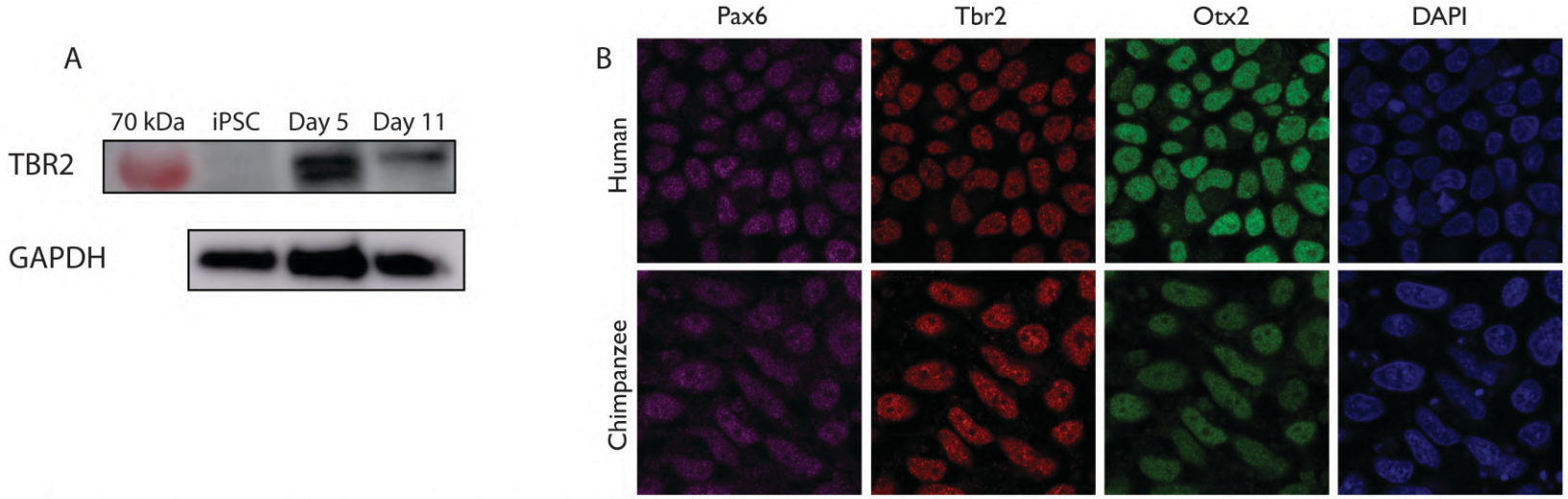

Figure 1 – Differentiation of human and chimpanzee iPSCs into hippocampal intermediate progenitors (A) Western Blot of human differentiated iPSCs against TBR2 (73kDa) at day 0 (iPSCs), day 5, and day 11 of treatment with hpIPC differentiation media. GAPDH is shown as the loading control. (B) Immunofluorescent staining of human and chimpanzee hpIPCs after five days against neurodevelopmental markers PAX6, TBR2 and OTX2. Nuclei are counterstained with DAPI.

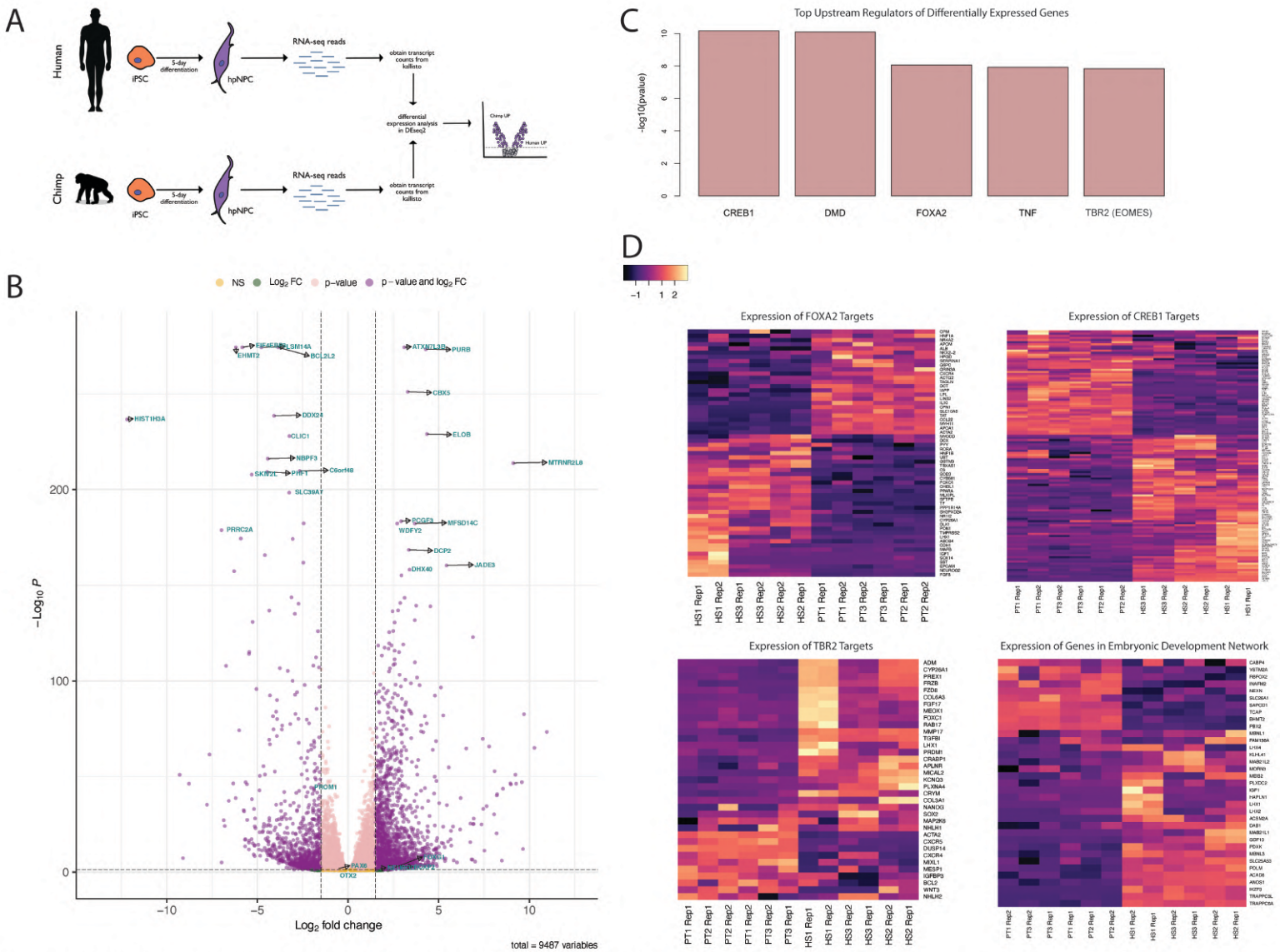

Figure 2 – Differential Gene Expression In Human and Chimpanzee hpIPCs. (A) Schematic of iPSC differentiation followed by RNA-seq library generation and analysis. (B) Volcano plot depicting "Human UP" genes (right) and "Chimp UP" genes (left) with a log2FoldChange threshold of 1.5 and -1.5, respectively, and a p-value threshold of 0.05. (C) Top upstream regulators of the differentially expressed genes predicted by Ingenuity Pathway Analysis, ranked by -log10(p-value). (D) Heatmaps depicting the expression of genes predicted to be under the control of the transcription factors CREB1, FOXA2 and TBR2 or predicted to play a role in embryonic development.

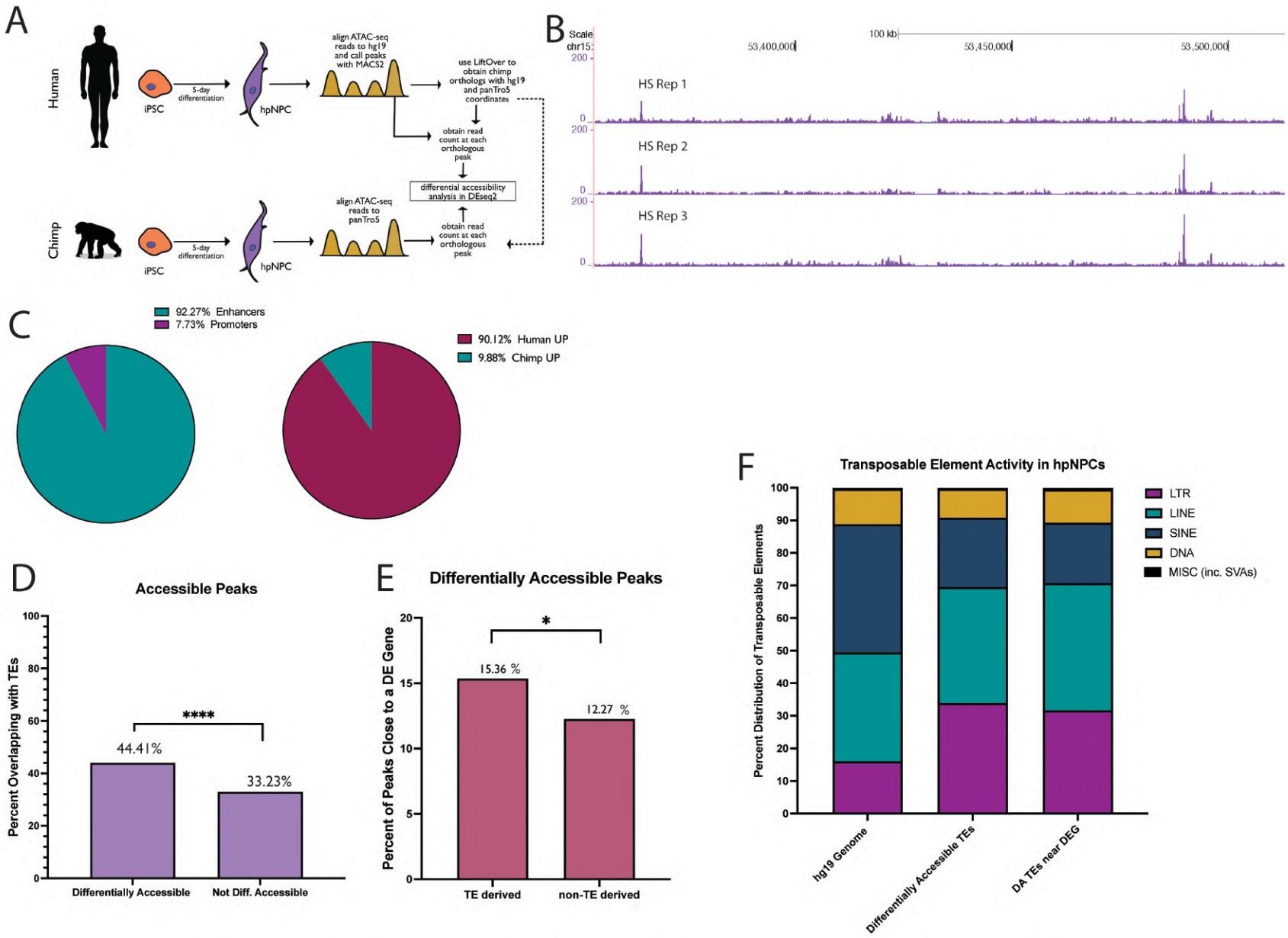

Figure 3 – Human Centric Chromatin Accessibility Analysis. (A) Schematic of iPSC differentiation followed by ATAC-seq library generation and analysis. (B) UCSC Genome Browser visualization of human ATAC-seq libraries (HS Rep 1, Rep 2, Rep 3) aligned to hg19 genome assembly. (C) Distribution of the differentially accessible (DA) peaks into enhancers or promoters, and into Human UP (greater accessibility in humans) or Chimp UP (greater accessibility in chimpanzees). (D) Differentially accessible chromatin regions ($p < 0.05$, $n = 3006$) are more likely to overlap with transposable elements (TE) than non-differentially accessible regions ($p > 0.9$, $n = 3{,}006$). (E) The 1335 TE-derived DA regions are more likely to be near a differentially expressed gene, as compared to DA regions that do not overlap a TE. (F) Distribution of the five major TE classes in the human genome, compared with their distribution among the TE-derived DA chromatin regions, and among the TE-derived DA chromatin regions proximal to a differentially expressed gene

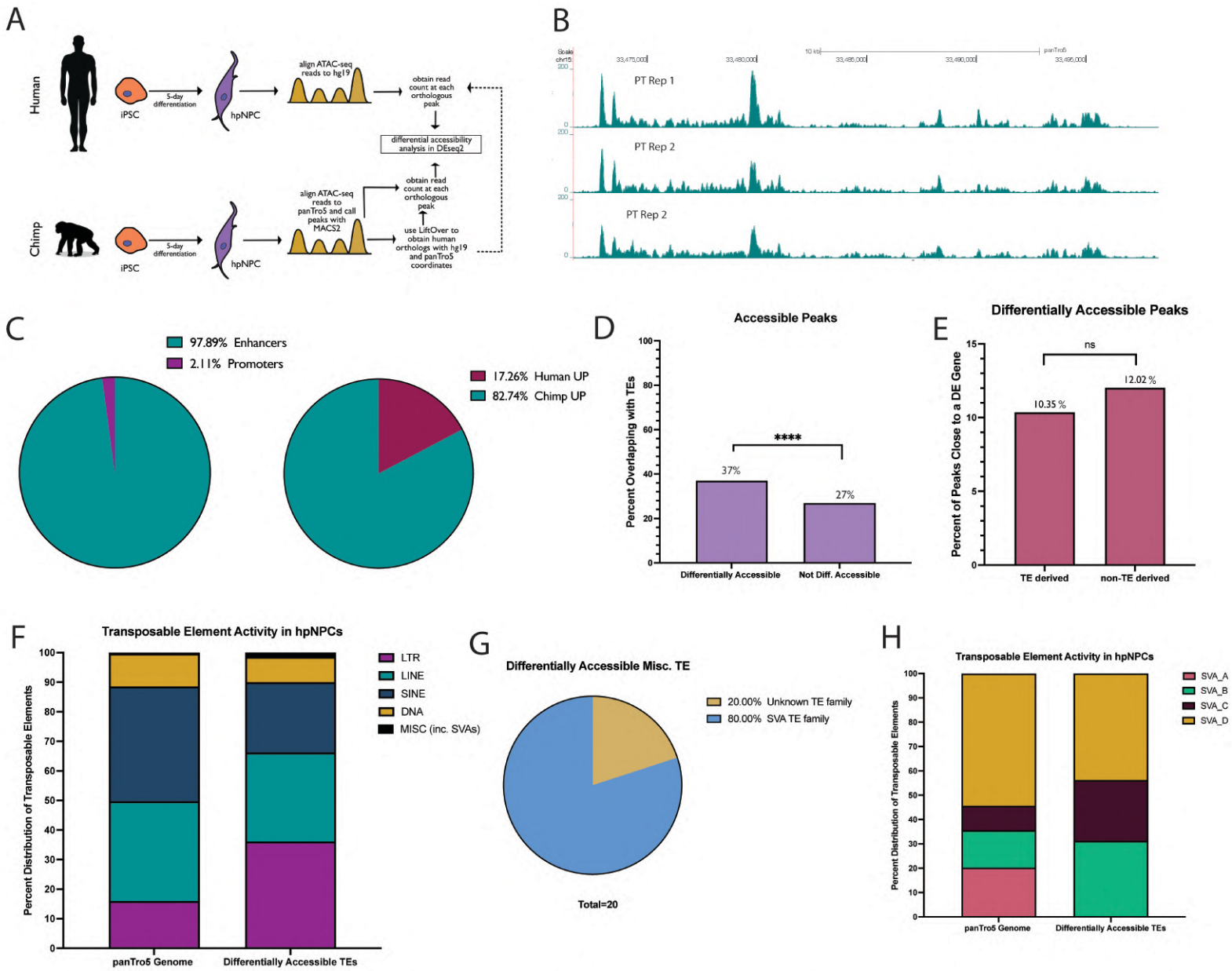

Figure 4 - Chimpanzee Centric Chromatin Accessibility Analysis. (A) Schematic of iPSC differentiation followed by ATAC-seq library generation and analysis. B) UCSC Genome Browser visualization of human ATAC-seq libraries (PT Rep 1, Rep 2, Rep 3) aligned to hg19 genome assembly. (C) Distribution of the differentially accessible (DA) peaks into enhancers or promoters, and into Human UP (greater accessibility in humans) or Chimp UP (greater accessibility in chimpanzees). (D) Differentially accessible chromatin regions ($p < 0.05$, n = 3,806) are more likely to overlap with transposable elements than other accessible regions ($p > 0.9$, n = 3806. (E) Distribution of DA regions that overlap a TE and proximity to genes differentially expressed between human and chimpanzee hpIPCs. (F) Distribution of the five major TE classes in the chimpanzee genome, compared with their distribution among the 1410 TE-derived DA chromatin regions (G) Breakdown of the 20 Miscellanous TEs represented in the 1,410 TE-derived DA chromatin regions. (H) Distribution of SVA family of TEs in the human genome as compared to their distribution in the TE-derived DA chromatin regions (n = 16).

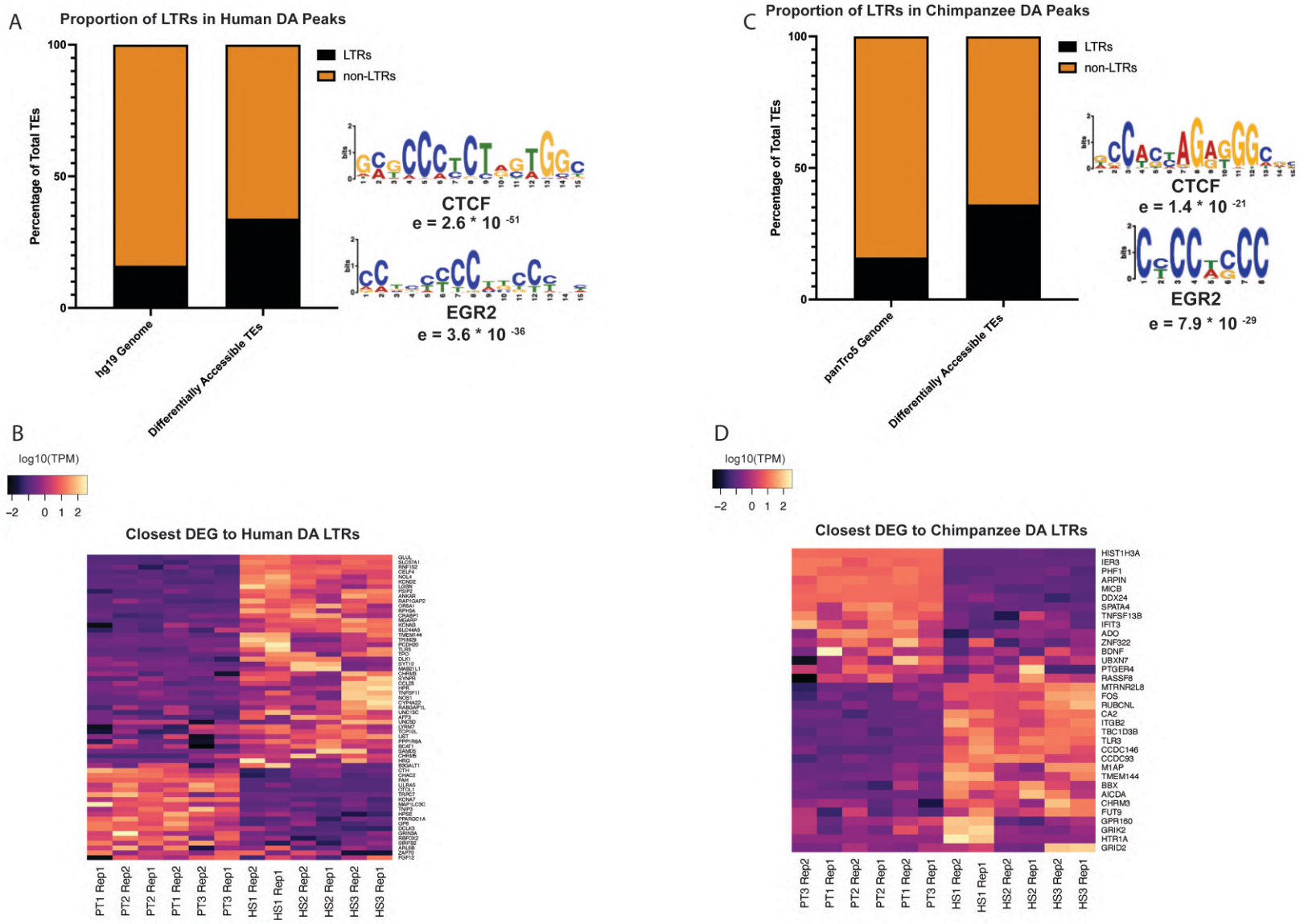

Figure 5 – LTRs Are Enriched Among Human and Chimpanzee Differentially Accessible Transposons. (A) Distribution of LTRs compared to non-LTRs in the human genome and in the 1,335 human TE-derived DA regions, with predicted binding motifs. (B) Differentially expressed genes close to the human-enriched DA LTRs. (C) Distribution of LTRs compared to non-LTRs in the chimpanzee genome and in the 1,410 chimpanzee TE-derived DA regions, with predicted binding motifs. (D) Differentially expressed genes close to the chimpanzee-enriched DA LTRs.

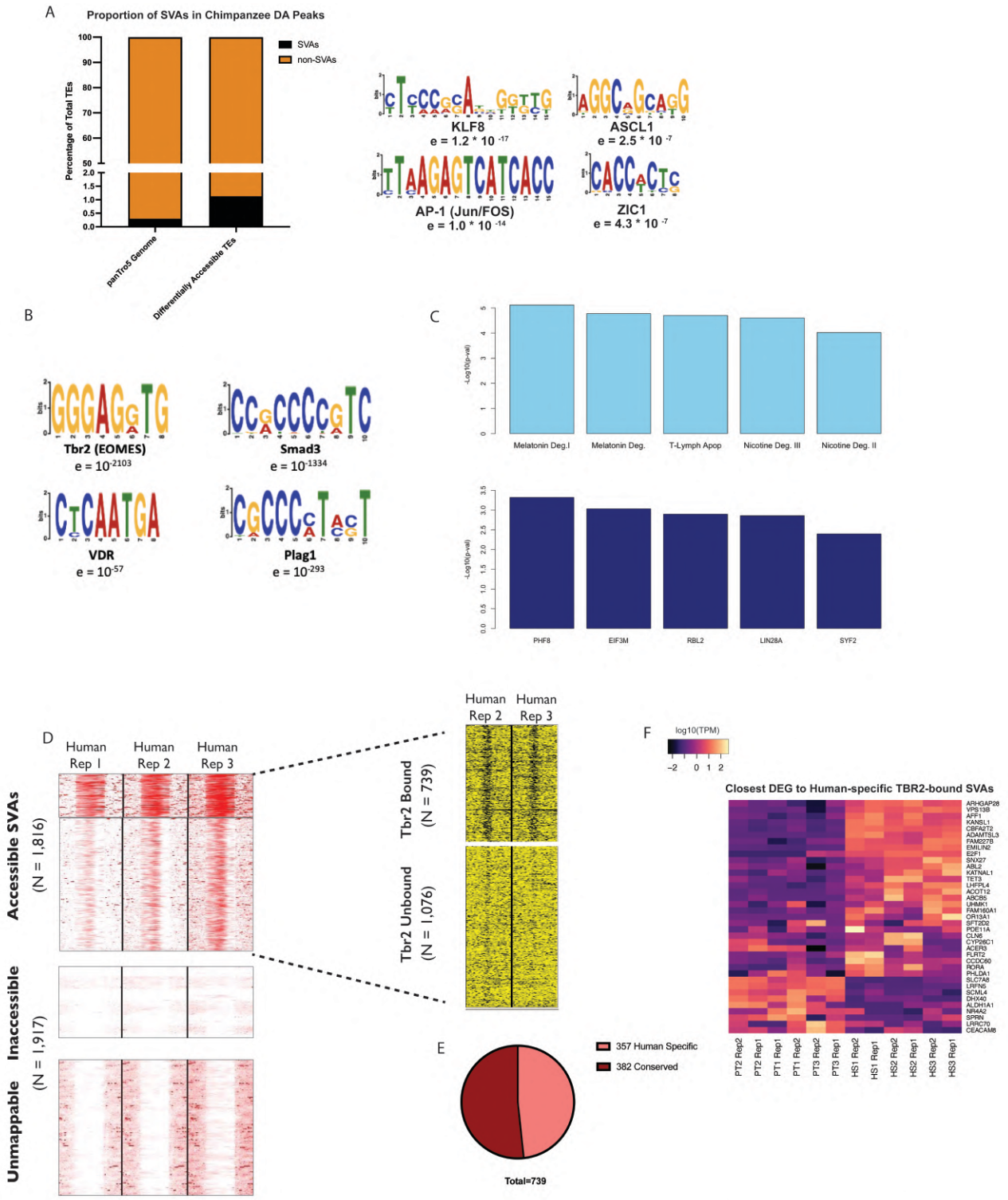

Figure 6 – Human-specific SVAs Bind Tbr2 and Influence Neurodevelopment (A) The SVAs enriched within the differentially accessible chimpanzee peaks are enriched for binding motifs of neurodevelopmental transcription factors KLF8, ZIC1, and ASCL1 as well as the AP-1 dimer. (B) Human SVAs are enriched for the binding motif of TBR2 (C) Genes proximal to human-specific SVAs are predicted to be regulated by neurodevelopmental TFs and function in important neuronal pathways (D) 1,816 human SVAs exhibit ATAC-seq signal in hpIPCs (red = signal, white = noise) and of these, 739 exhibit TBR2 ChIP signal (black = signal, yellow = noise). (E) Breakdown of 739 accessible, TBR2 bound SVAs into "human-specific" and "conserved in chimpanzees". (F) Differentially expressed genes close to the human-specific TBR2-bound SVAs.

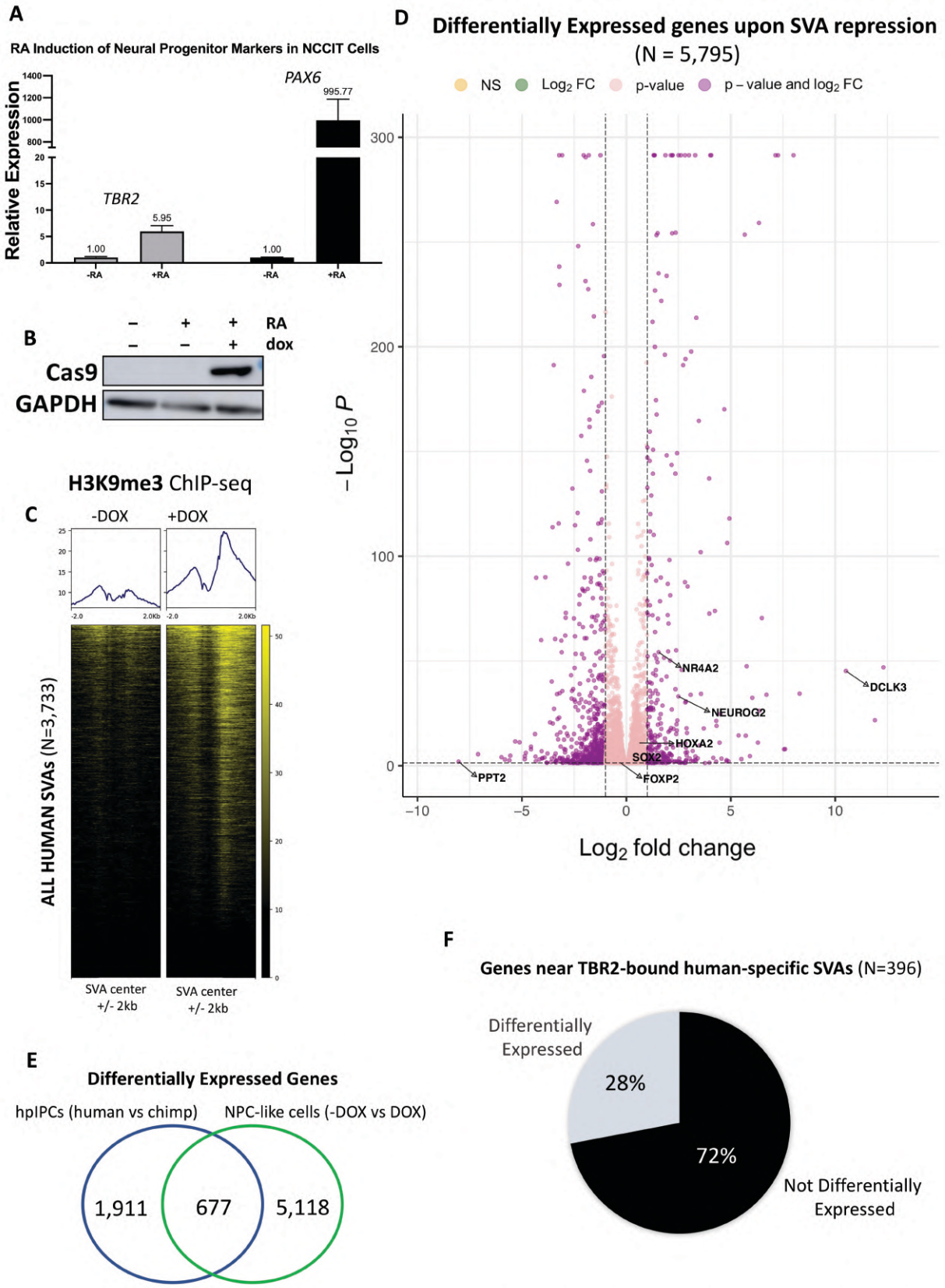

Figure 7 – CRISPR-interference on human SVAs. (A) Treating NCCITs with Retinoic Acid for 7 days leads to induction of TBR2 and PAX6, as shown in qRT-PCR. (B) Treating the RA-NCCIT stable CRISPR line for 72 hours with doxycycline leads to dCas9 activation, as shown in the Cas9 immunoblot. (C) dCas9 activation results in deposition of H3K9me3 at most human SVAs as shown in the H3K9me3 ChIP-seq heatmap (yellow = signal). (D) Volcano plot depicting genes differentially expressed upon doxycycline treatment. (E) Venn diagram showing overlap between genes differentially expressed between human and chimpanzee's hippocampal intermediate progenitors and between RA-NCCITs with and without doxycycline treatment (i.e. with and without SVA repression). (F) Pie chart illustrating the fraction of TBR2-controlled genes that are differentially expressed upon SVA repression in RA-NCCITs.