Human-specific regulation of neural maturation identified by cross-primate transcriptomics

Sara Linker, Iñigo Narvaiza, Jonathan Hsu, Meiyan Wang, Fan Qiu, Ana P.D. Mendes, Ruth Oefner, Kalyani Kottilil, Amandeep Sharma, Lynne Randolph-Moore, et al.

Human-specific regulation of neural maturation identified by cross-primate transcriptomics

Sara B. Linker[1], Iñigo Narvaiza[1], Jonathan Y. Hsu[1], Meiyan Wang[1], Fan Qiu[1], Ana P. D. Mendes[1], Ruth Oefner[1], Kalyani Kottilil[1], Amandeep Sharma[1], Eunice Mejia[1], Renata Santos[1,2,3], Maria C. Marchetto[1,4,5*], Fred H. Gage[1,6*]

[1] Laboratory of Genetics, The Salk Institute for Biological Studies, 10010 North Pines Road, La Jolla, CA 92037, USA.

[2] Université Paris Cité, Institute of Psychiatry and Neuroscience of Paris (IPNP), INSERM U1266, Laboratory of Dynamics of Neuronal Structure in Health and Disease, 102 rue de la Santé, 75014 Paris, France.

[3] Institut des Sciences Biologiques, CNRS, 16 rue Pierre et Marie Curie, 75005 Paris, France.

[4] Department of Anthropology, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA.

[5] Center for Academic Research and Training in Anthropogeny (CARTA), University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA.

[6]Lead contact: Fred H. Gage

[*] These authors contributed equally

Correspondence should be addressed to Fred H. Gage (gage@salk.edu) and Maria C. Marchetto (mcmarchetto@ucsd.edu).

SUMMARY

Unique aspects of human behavior are often attributed to differences in the relative size and organization of the human brain; these structural aspects originate during early development. Recent studies indicate that human neurodevelopment is considerably slower than that in other nonhuman primates, a finding that is termed neoteny. One aspect of neoteny is the slow onset of action potentials. However, which molecular mechanisms play a role in this process remain unclear. To examine the evolutionary constraints on the rate of neuronal maturation, we have generated transcriptional data tracking five time points, from the neural progenitor state to eight-week-old neurons, in primates spanning the catarrhine lineage, including *Macaca mulatta*, *Gorilla gorilla*, *Pan paniscus*, *Pan troglodytes*, and *Homo sapiens*. Despite finding an overall similarity of many transcriptional signatures, species-specific and clade-specific distinctions were observed. Among the genes that exhibited human-specific regulation, we identified a key pioneer transcription factor, *GATA3*, that was uniquely up-regulated in humans during the neuronal maturation process. We further examined the regulatory nature of *GATA3* in human cells and observed that down-regulation quickened the speed of developing spontaneous action potentials, thereby modulating the human neotenic phenotype. These results provide evidence for the divergence of gene regulation as a key molecular mechanism underlying human neoteny.

INTRODUCTION

Since the split of *Homo sapiens* from the last common nonhuman primate (NHP) ancestor, the human brain has substantially altered its size, structure and connectivity. The human brain has a larger mass with respect to body weight, increased cortical neurons with respect to size, an expanded proliferative zone, and unique connectivity patterns [1-3]. The identification of human-specific genes that contribute to these phenotypes has been advanced by a better understanding of genomic copy number changes and neuronal modeling techniques [4,5]. For example, the human-specific genes *FOXP2* and *ARHGAP11B* increase dendritic arborization, cortical size and folding [6-8]. Changes to transcriptional regulation have also driven human-specific features such as human-accelerated regions (HARs), small regions of the genome with abnormally fast positive selection [9]. Human-specific regulatory regions in *FZD8*, *NPAS3*, and *GADD45G* control aspects of human neurodevelopment such as cell cycle progression, brain region specification, and cortical size [10-12]. Furthermore, organoid models revealed that human-specific expression of the conserved epithelial-to-mesenchymal transition regulator ZEB2 modulates neuroepithelial cell shape and the timing of neurodevelopment [13].

Human-specific neurodevelopment is not only marked by physical differences, but also by temporal changes. Human neurons, during both prenatal neurodevelopment and adult neurogenesis, exhibit an exceptionally delayed time course, a characteristic termed neoteny [14-17]. It is hypothesized that this longer developmental period plays a role in the aforementioned structural and connectivity differences. Signatures of human-specific neoteny have been observed and reproduced across systems including *in vivo*, induced pluripotent stem cell (iPSC), induced neuron (iN), and organoids models. Neural progenitor cells (NPCs) derived from human iPSCs proliferate for a longer period of time, and migrate more slowly, than NPCs derived from NHPs [18-21]. Neotenic phenotypes also persist during the maturation period, with human cells being capable of generating evoked action potentials at a slower rate than *Macaca mulatta* [22] and spontaneous action potentials more slowly than *Pan paniscus* and *Pan troglodytes* [19].

The genes supporting dendritic development are also delayed in expression in iNs [23]. Together, these studies indicate that iPSC-derived neurons can recapitulate aspects of neoteny in a dish. IPSC modeling is particularly useful when examining the molecular drivers of a cellular phenotype such as gene transcription. It has long been proposed that the phenotypic differences between closely related species may be driven, in part, by divergent transcriptional regulation rather than

by novel protein-coding sequence [24,25]; however, how these regulatory mechanisms play a role in the protracted maturation process in human neurons remains largely unknown.

Here, we modeled neurogenesis across five primate species - *Macaca mulatta* (rhesus), *Gorilla gorilla* (gorilla), *Pan paniscus* (bonobo), *Pan troglodytes* (chimpanzee), and *Homo sapiens* (human) - and assessed the differences in transcriptional dynamics. While the NPCs of humans, chimpanzees, and bonobos were highly similar, we found that transcriptional differences increased between all species throughout neuronal differentiation and maturation. We identified a pioneer transcription factor, *GATA3*, that exhibited elevated neuronal expression only in humans. Strikingly, down-regulation of *GATA3* increased the rate of physiological maturity in human neurons, indicating that the species-specific rate of physiological maturity is cell intrinsic and can be modulated by perturbing a single, conserved transcription factor. This finding provides evidence for the divergence of gene regulation as a contributor to human neoteny.

RESULTS

RNA-seq separates primate iPSC-derived neurons in association with phylogenetic distance

To study species-specific neuronal maturation, we generated iPSCs from human, bonobo, chimpanzee [26,27], gorilla, and rhesus samples [28]. iPSCs from each species were differentiated into NPCs using a pan-neuronal protocol [26] and were then matured into neurons for 8 weeks. As expected to all species, NPCs ($n_{, human}$ = 4 individuals; $n_{, chimpanzee}$ = 2 individuals; $n_{, bonobo}$ = 2 individuals; $n_{, gorilla}$ = 2 individuals; $n_{, rhesus}$= 1 individual, 2 clones) expressed similar levels of SOX2 (one-way ANOVA p = 0.52), and neurons similarly expressed synapsin-GFP (one-way ANOVA p = 0.67) and CTIP2, a marker of deep cortical neurons (one-way ANOVA p = 0.86) (**Figure 1A**). RNA was collected from neurons at two-week intervals (neurons = 2, 4, 6, and 8 weeks) to examine species-specific expression profiles (**Figure 1B**). RNA was also collected from NPCs to serve as a control time point (NPC = 0 weeks) to enable the separation of the transcriptional signatures that were important for neuronal maturation from the background of genes that were species-specific but maturation-independent.

Differences in the quality of a reference genome can impact expression estimates. Therefore, it was important to analyze expression within regions that exhibited high homology between the species. To minimize alignment artifacts and differences in the quality of reference genomes between species, all samples were first aligned to the available reference sequences for human, chimpanzee, gorilla, and rhesus macaque. The chimpanzee reference was used as a filter for both chimpanzee and the closely related bonobo. To restrict analysis to regions of high homology, we then assessed expression differences using only reads that aligned uniquely to all four species (**Figures S1A-S1D**). We first assessed the global expression patterns across all samples. Principal component analysis (PCA) of the temporal RNA-seq data revealed a clear separation of NPCs from neurons for all species along PC1 (21% of variance) (**Figure 1C**). PC2 (17% of variance) separated species in association with phylogenetic distance. Human samples were most similar to both bonobo and chimpanzee, which represented a separation of approximately 7 million years (my), followed by gorilla (10 my) [29], and rhesus (25 my) (**Figure 1C**). To determine which genes exhibited species-independent dynamics throughout differentiation, we performed mixed model regression with neuronal state and species as the fixed and random effects, respectively (n_{NPC}: human = 6, bonobo = 3, chimpanzee = 2, gorilla = 3, rhesus = 3; n_{Neuron}: human = 12, bonobo = 4, chimpanzee = 4, gorilla = 8, rhesus = 4), as well as a model that assessed transcriptional changes between NPCs and neurons as a function of phylogeny (*Homo*, Hominini, or Homininae). A total of 5,099 genes were detected as being differentially expressed between NPCs and neurons, with 87% independent of phylogeny and 13% exhibiting phylogeny-dependent temporal dynamics. As expected, genes that were consistently up-regulated in NPCs independent of species (2,276 genes) displayed roles in cell cycle progression (270 genes, padj < 1.83e-72), such as the cyclin-dependent kinases (nine cyclin C-terminal genes, padj < 2.2e-02) and DNA replication genes (94 genes padj < 6.04e-37) (**Figures 1D and 1E, and Data S1**). Genes up-regulated in neurons independent of species (2,182 genes) were related to neuron-specific structural components such as the synapse (106 synapse genes; padj < 4.76e-23) and ion transport, including multiple voltage-gated calcium, potassium, and sodium channels (139 genes, padj < 1.5e-10) (**Figures 1D and 1E, and Data S1**), indicating that NPCs from all species were differentiating into neurons.

To examine the neuronal maturation process, we identified genes that were differentially expressed as a function of time using only the neuronal samples that were 2- to 8-weeks post-differentiation from the NPC state in all species. We detected significant changes over time across all species in 60 genes, 28 of which were not previously identified as being different between the

NPC and neuronal states (**Data S1**). These 28 genes contained interesting regulators of neuronal development such as *AUTS2* and *DBN1*, but they were not enriched for a specific GO or KEGG pathway.

All species transition similarly from NPCs to neurons

To examine the consistent species-independent signature in more detail, we predicted the spatiotemporal expression patterns of all samples in comparison to the post-mortem reference using brainImageR [30-32]. Temporal analysis was performed first in comparison to the full BrainSpan dataset, wherein all samples displayed high similarity to the prenatal (< 40 weeks post-conception) time window (**Figure S1E**). To increase the resolution within the prenatal period, we repeated the analysis using only the samples from BrainSpan that were < 40 weeks post-conception. Temporally, all NPCs were predicted to model approximately human 13 weeks post-conception, a time point that increased with differentiation similarly for all species (**Figure S1E**). The similarity in temporal space indicated that NPCs from all species successfully transitioned from a neural progenitor to a neuronal state after induction of differentiation (see methods), thus facilitating an analysis of differences in species-specific dynamics during this transition. Since the majority of the transcriptional changes that were observed during differentiation occurred between the NPC and neuronal states and changed minimally throughout the maturation process, for the remainder of the analyses we grouped neurons from 2 to 8 weeks and analyzed differences between NPCs and neurons.

For spatial analysis, we also compared our dataset to the available prenatal microdissected time point of 15 weeks post-conception [30-32]. Genes up-regulated in NPCs were identified as enriched in the subventricular zone (SVZ) and genes up-regulated in neurons were enriched for regions containing mature neurons throughout the developing brain, such as the bed nucleus of stria terminalis in the telencephalon, the hypothalamus in the diencephalon, and the pontine nucleus in the hindbrain (**Figure 1F**). We further examined transcription factors associated with fore-, mid- and hindbrain patterning: SIX3, FOXG1, PAX6, OTX2, PAX2, EN1, IRX3, and GBX2. We noted a minor preference for forebrain expression patterns for rhesus macaque compared to human, chimpanzee and bonobo. However, there were not clear human-specific differences in expression of these patterning genes. Finally, to compare cell types, we predicted the cell type composition of each sample set using data from studies [33] or [34] as a reference set. All differentiating cells expressed genes consistent with astrocytes, glutamatergic neurons, and GABAergic neurons,

with consistent levels of cell type-specific expression profiles across species (**Figures S1F and S1G**). In combination with the staining results, this analysis indicated that all samples contained similar mixtures of neural cell types.

A human-specific transcriptional signature develops throughout neuronal differentiation, revealing GATA3 transcription factor as the top human-specific correlated gene

We next sought to understand the intrinsic species-specific transcriptional dynamics throughout neuronal differentiation. First, we explored the variance within NPCs or neurons alone using two separate PCAs. A PCA of NPCs alone clustered all Hominini (human, chimpanzee, and bonobo) along PC1 (**Figures 2A and S2A**). Gorilla (F-test $p < 9.07e-06$) and rhesus (F-test $p < 4.52e-11$) were significant outgroups. However, the clustering within Hominini shifted during differentiation. In a PCA of neurons alone, human samples clustered separately from all other species (F-test $p_{chimpanzee} < 3.95e-06$, $p_{bonobo} < 3.53e-08$, $p_{gorilla} < 5.30e-17$, $p_{rhesus} < 1.27e-18$) (**Figures 2B, S2A and S2B**). This clustering of Hominini in the NPC state and separation within the neuronal state was consistent after subsampling all species to equivalent sample size (Neuron F-test: $p_{chimpanzee} < 0.006$, $p_{bonobo} < 9.21e-05$, $p_{gorilla} < 8.38e-14$, $p_{rhesus} < 2.2e-16$) (**Figure S2C**), indicating that the transcriptional profile became increasingly species-specific during neurogenesis.

To identify human-specific expression dynamics, we performed Weighted Gene Co-expression Network Analysis (WGCNA) on genes that either increased or decreased in expression during human neural differentiation, as determined by differential expression between human NPCs and neurons. The rationale for expanding the starting gene list beyond those with detectable human-specific expression patterns was to include genes that might have human-specific dynamics but were not detected in differential expression analyses due to low power or variability in expression within the neuronal (2- to 8-week) state. Of 125 identified modules (83 up-regulated modules, 42 down-regulated modules), four exhibited a main effect of human-specific expression as well as human-specific dynamics over time (up-regulated module 1= 165 genes, up-regulated module 2= 366 genes, up-regulated module 3 = 91 genes, down-regulated module 1 = 358 genes) (**Figures 2C and S2D**). Only up-regulated module 1 contained genes that were not expressed in the NPC state and were highly expressed upon differentiation into neurons. Therefore, this module was further studied to identify genes with potential involvement in human-specific neural maturation. To identify which of these genes had the highest likelihood for impacting human transcription, a human-neuron correlation score was calculated. We hypothesized that regulatory factors relevant

for human specificity would be correlated to the expression of other human-specific genes in neurons and would have more connections to genes that exhibited human specificity in neurons than in NPCs. Therefore, we assessed the correlation of each human-specific up-regulated module-1 gene with all other genes expressed in neurons. We then counted the number of connections that had a correlation coefficient > 0.60 to genes that exhibited significant human-specific expression in NPCs or separately in neurons. The ratio between these two estimates, Neuron $_{\text{human-specific,k}}$ and NPC $_{\text{human-specific,k}}$, was defined as the human-neuron correlation (HNC) score. Nineteen up-regulated module-1 genes had an HNC score above 1, indicating that they had a strong association with human-specific genes in the neuronal state. *GATA3*, a pioneer transcription factor, and its antisense transcript *GATA3-AS1* were the top human-specific correlated genes (**Figures 2D and 2E**). *GATA3-AS1* was up-regulated at two weeks and then decreased over the 4- to 8-week period whereas the transcription factor *GATA3* increased at 2 weeks and then stayed constant over the remainder of the eight-week period. *GATA3-AS1* is known to regulate expression of the transcription factor *GATA3* in T cells [35], indicating that it may play a role in up-regulating the neighboring *GATA3*.

GATA3 is a pioneer transcription factor that is capable of reorganizing chromatin and is important for cell state transitions, including the epithelial to mesenchymal transition [36], T-cell differentiation [37], differentiation of neurons of the inner ear [38,39] and differentiation of neurons in the dorsal raphe [40]. Given the correlation of *GATA3* primarily with genes with human-specific expression in this study, and its known role as a transcription factor that drives cell fate in multiple systems, we felt that *GATA3* was a prime candidate to examine as a regulator of human-specific cortical neuronal maturation.

GATA3 expression was up-regulated specifically in human neurons during differentiation both at the RNA (two-way ANOVA padj < 5.7e-03; **Data S1; Figures 2D and 2E**) and protein (ANOVA $_{\text{GATA3 ~ Human + HSP90}}$ p $_{\text{human}}$ < 1.1e-03; **Figures 2F and S2F**) levels. To confirm that the *GATA3* expression differences observed *in vitro* were also present *in vivo*, we examined *GATA3* expression within post-mortem human gene expression data provided by the Allen Brain Atlas [30-32] and identified that *GATA3* was indeed expressed in the developing brain and was enriched in the intermediate zone (IZ) **(Figure S2E**). To determine the human specificity of *GATA3* expression *in vivo,* we examined transcription within the human and rhesus post-mortem IZ. Prenatal expression data were acquired from the NIH Blueprint NHP Atlas and the BrainSpan dataset for rhesus macaque and human, respectively. Both datasets were available from the Allen

Institute for Brain Science and were scaled across all available prenatal tissues prior to analysis, enabling us to examine the relative level of *GATA3* expression in the IZ within each species. We again observed that *GATA3* was significantly increased in the human compared to rhesus (ANOVA p < 9.85e-11; **Figure 2G**), further supporting that *GATA3* was elevated specifically in humans during early neurogenesis.

GATA3 is bound to neurodevelopmentally regulated genes

To determine which genes might be regulated by GATA3, we performed ChIP-seq in 2-week-old iPSC-derived neurons from three separate human individuals. A total of 3,199 genes were identified as being bound by GATA3 in all three samples (**Data S2**). As expected, binding sites were enriched for the canonical GATA3 motif (**Figures S3A and S3G**) and were highly over-represented in promoter regions and 5'-UTRs (**Figures 3A and 3B**). In line with previous work in T-cells [41,42], GATA3 binding was associated with both activation and repression of gene expression (**Figure S3C**). Functional enrichment of GATA3-bound genes identified an association with processes important for the NPC state, such as the cell cycle (133 genes, p-value < 5.7e-06), as well as processes important for the development of the mature neuron, such as protocadherin gamma genes (19 genes, p < 2.9e-11), cadherins (42 genes, p < 1.0e-06), and the synapse (74 genes, p < 4.9e-03) (**Data S2**). Approximately half (1,694 of 3,199) of GATA3-bound genes were found to be expressed in the SVZ/VZ in the Allen Brain Atlas prenatal dataset (**Figure 3C**).

Using the results of the multiple regression analysis, GATA3 was not found to be enriched at genes with conserved temporal expression patterns or at human-specific genes. This finding may have been due to the low power achieved with the differential expression test. We therefore determined in an unbiased manner the types of temporal dynamics that GATA3 binding was associated with through WGCNA analysis. Interestingly, even though GATA3 was only expressed in humans, the majority of these genes decreased expression throughout maturation in all species, indicating that GATA3 may play a compensatory role for factors that are present in other species. In addition to this putative compensatory role, GATA3 was also bound to genes that displayed human-specific dynamics throughout maturation (**Figure 3SD)**. For example, genes exhibited dynamics such as sustained expression from the NPC state or up-regulation in humans and lower expression in NHPs (ex: *CDC14A*, *SLC41A3*, *APBB2*, **Figure S3E**).

We compared GATA3-bound genes to genes associated with neurodevelopmental disorders using either the Simons Foundation Autism Research Initiative (SFARI) list (Hypergeometric p < 4.13e-05) or the Deciphering Developmental Disorders (DDD) gene list (Hypergeometric p < 1.16e-18). There was a significant enrichment for GATA3-bound genes within both of these lists (**Figure 3F**). We then calculated the enrichment score for a disorder that is not related to neurodevelopment, inflammatory bowel disease, and identified no association with GATA3-bound genes (Hypergeometric p < 0.74). Together, these analyses indicate that GATA3 binds to genes that are involved in early human cortical development.

GATA3-binding sites were under positive selective pressure in humans

Recent studies have focused on the impact of human-specific paralogs on neuronal neoteny [6,18,20,43,44]. We examined the set of 24 gene families that Suzuki et al. (2018) identified to contain human-specific paralogs and to be expressed in the developing brain [18]. Of these 24 gene families, 15 had members that were directly bound by GATA3 at the promoter in at least two of three samples, with 30 genes bound by GATA3 in total (**Table S1**). Interestingly, one such paralog was *NOTCH2NL* from the *NOTCH2* family, which was recently identified to regulate neotenic development in the NPC state [18,20,45] (**Figure 3B**). While this enrichment was not significant in comparison to all genes (Chi-squared test p = 0.16), it does suggest that GATA3 may play a role in regulating human-specific paralogs during the transition from NPC proliferation to neuron generation to neuronal differentiation.

Previously, Arbiza et al. described an algorithm, INSIGHT, that combined human polymorphism data with evolutionary data to calculate signatures of weak positive and negative selection on transcription factor binding sites [46,47]. They examined 78 transcription factors from the ENCODE consortium including GATA3 and transcription factors with similar sequences: GATA1 and GATA2. GATA2 and GATA3 exhibited the strongest signature of positive selection across all transcription factors, whereas GATA1, though similar in sequence, exhibited weak selection [46].

To test whether the sites bound by GATA3 in human neurons were similarly under positive selective pressure in humans, GATA3 motif locations were extracted from each ChIP-seq peak and were assessed for their divergence from NHP as well as the polymorphism status within humans using the INSIGHT algorithm [46,47]. By measuring the expected number of adaptive substitutions, we identified a significant increase in the probability that the GATA3-binding sites

had undergone strong positive selection in the human lineage compared to random chance (E[A] = 1.86 ± 0.46 p < 0.01, **Figure S3F**). Furthermore, by measuring the expected number of weakly deleterious polymorphisms (E[W] = 0.36 ± 0.35, **Figure S3G**), we identified no signature of weak negative selection. These results indicated that the binding sites of GATA3 identified in human neurons were largely unique to humans. Further work is required to determine if these mutations are necessary and sufficient for driving changes at the gene expression level.

GATA3 regulates the rate of physiological maturity in human neurons

To examine the direct impact of *GATA3* expression during human neuronal differentiation, we generated stable partial knockdowns using an shGATA3 (**Figures S4A-S4C**) in two NPC lines generated from two different individuals (WT33 and WT126) with two replicates per individual. We then differentiated them into neurons over four weeks. We performed RNA-seq on all samples at biweekly intervals (NPC = 0, neurons = 2 and 4 weeks). To determine genes directly regulated by GATA3 and to filter out genes influenced indirectly by the short hairpin construct, differential expression was calculated between human neurons treated with shGATA3 samples versus the combination of shControl and non-infected samples. A total of 155 genes were identified as being regulated by GATA3 with an FDR < 0.05. There was a significant enrichment of GATA3 binding within genes regulated by GATA3, with 106 of the 155 GATA3-regulated genes being bound by GATA3 in at least one human sample (Fisher exact test p < 3.37e-10) and 33 bound in all three human samples (Fisher exact test p < 9.79e-19) (**Figure 3D**). Of these 33 genes, 23 are expressed in the SVZ and VZ regions in the BrainSpan database of prenatal human development (**Table S2**). Most of the genes directly regulated by GATA3, as determined by ChIP-seq and modulated by GATA3 knockdown, exhibited decreases in expression from NPC to neuron in all species (93 of 106 genes**, Figure S4D**), supporting a role for GATA3 in regulating the repression of genes from the NPC to neuronal state. Although these genes were detected as down-regulated in a combined analysis across all species, Hominini samples had a larger fold change than gorilla or rhesus, indicating these genes decreased expression more rapidly from NPC to neurons in Hominini samples. A subset of genes bound by GATA3 and dysregulated by shGATA3 have known roles in cell cycle control, including *CDK6*, *CHEK1*, and *PCNA*.

To determine if there was an additional signature of human-specific expression modulated by knockdown, we performed PCA using all genes that were regulated by GATA3 with a raw p-value < 0.05. In contrast to the full transcriptome, GATA3-regulated genes displayed a species-specific

gene signature within neurons when examined by PCA (**Figures 3E and S3G**). PC1 (31% of variance) exhibited an expression profile that was shGATA3-specific and species-independent (**Figure 3E**) whereas PC2 (14% of variance) separated samples by species. As expected, samples expressing the shControl clustered with the non-infected human samples. Strikingly, the neuronal samples that expressed shGATA3 shifted down along PC2, away from humans, to a position that matched chimpanzee and bonobo neurons (**Figure 3E**). These results indicated that GATA3 further regulated a component of transcription that distinguished human neurons from *Pan*. Some of the genes with top loading values for PC2 were known early development-associated genes such as *OTX2, OTX1, and LEF1*. To identify genes at an individual level, we performed a differential expression analysis comparing human neurons to *Pan* and overlapped this gene list with direct GATA3 targets. Of the 33 high confidence direct GATA3 targets, 6 (18%) were significantly differentially expressed between human and *Pan* (p-adj < 0.05), with 3 shifted human expression to a more *Pan*-like state after knockdown (*DOCK8*,*STXBP4* and *TENM3*).

In vitro-derived human neurons are delayed in their ability to generate spontaneous action potentials compared to iPSC-derived chimpanzee or bonobo neurons [19]. We therefore assessed the impact of *GATA3* knockdown on spontaneously generated action potentials using multi-electrode arrays (MEA). Neurons derived from two separate human individuals were plated onto a 96-well MEA plate. Non-infected, shControl, and shGATA3 samples were generated for each line and all samples were measured over a 20-day period following differentiation from NPCs. Strikingly, at the earliest time point examined (day 7), the human shGATA3-expressing neurons generated spontaneous action potentials whereas both the shControl and non-infected human cell lines exhibited minimal activity (mean firing rate: non-infected $_{day7}$ = 4.1e-03 $\pm$ 8.8e-03, shControl $_{day7}$ = 5.8e-05 $\pm$ 2.3e-04, shGATA3 $_{day7}$ = 0.11 $\pm$ 0.15; p-value $_{shControl,\ day7\ vs.\ non\text{-}infected,\ day7}$ = 0.86, p-value $_{shGATA3,\ day7\ vs.\ non\text{-}infected,\ day7}$ < 3.21e-05) (**Figure 4A**). The increased early activity in shGATA3 samples was consistent across multiple wells, indicating that the effect was highly reproducible (**Figure 4B**). Activity in shGATA3 samples was further associated with increased bursting, indicating that the shGATA3 neurons were not only physiologically mature but also highly connected within a given well (mean bursts: non-infected $_{day\ 7}$= 0.45 $\pm$1.1, shControl $_{day\ 7}$ = 0 $\pm$ 0, shGATA3 $_{day\ 7}$ = 2.92 $\pm$ 2.75; p-value $_{shControl,\ day\ 7\ vs.\ non\text{-}infected,\ day\ 7}$ = 0.35, p-value $_{shGATA3,\ day\ 7\ vs.\ non\text{-}infected,\ day\ 7}$ < 3.48e-06) (**Figure 4C**). Slower to mature, the non-infected and shControl neurons began to fire at rates similar to the day 7 shGATA3 samples at day 11 (mean firing rate: non-infected $_{day\ 11}$ 0.10 $\pm$ 0.12, shControl $_{day\ 11}$ = 0.05 $\pm$ 0.10 ; p-value $_{shControl\ ,day\ 11\ vs.\ non\text{-}infected,\ day\ 11}$ = 0.20, p-value $_{shGATA3,\ day\ 7\ vs.\ non\text{-}infected,\ day\ 11}$ = 0.85), but they still fired more slowly than shGATA3

neurons on the same day (mean firing rate: shGATA3 $_{day\ 11}$ = 0.45 $\pm$ 0.34, $_{shGATA3,\ day\ 11\ vs.\ non\text{-}infected,\ day\ 11}$ = 3.74e-07) (**Figure 4A**). We continued to analyze activity for 20 days following differentiation and tested the difference between sh and noninfected control using a multiple regression model (activity measure ~ sample + date + sample X date). The "firing rate" activity measure was not significantly different between shControl and controls (main effect p = 0.51, interaction with time p = 0.82), whereas the main effect was significantly increased in shGATA3 (main effect p < 0.004, interaction with time p = 0.68). With respect to the "bursts" activity measure, the shControl neurons displayed somewhat lower bursting; however, this effect was not significantly different from control (main effect p = 0.06, interaction with time p = 0.79). Conversely, the shGATA3 wells displayed a consistently higher level of network bursting and a shift in the slope of bursting over time (main effect p < 0.002, interaction with time < 0.045). (**Figure 4C**). While further work must be done to determine if the presence of GATA3 drives a delay in neuronal maturation, these results indicated that shGATA3 shifted the neotenous, delayed activity state of human neurons forward in time. Together these results show that GATA3 is uniquely up-regulated in human neurons in comparison to other NHPs and acts to delay the rate of maturity of spontaneous action potentials, a key aspect of human neoteny.

DISCUSSION

One key aspect feature separating humans from other apes is the protracted developmental period of neurogenesis that can be distinguished by the delayed generation of both spontaneous and evoked action potentials. While the molecular components required to build a functional synapse are numerous, we identified that inhibition of a single pioneer transcription factor, GATA3, accelerated the developmental timing of this human maturation process. *GATA3* was expressed at low levels in NPCs and expression increased during human neuronal maturation. The neurons in this study expressed neuronal markers (synapsin, MAP2, and CTIP2) as well as markers of immaturity such as the migration-associated gene *DCX,* indicating that they represent neurons in an immature state. Given that our findings are consistent *in vivo*, we speculate from our results that, as human neurons migrate out of the SVZ and through the IZ to their final destination in the cortex, *GATA3* is elevated, thereby facilitating a human-specific delay in physiological maturation.

What drives *GATA3* up-regulation in human neurons remains unclear. However, the upstream, and non-overlapping, non-coding RNA, *GATA3-AS1*, was also uniquely up-regulated in humans. *GATA3-AS1* was previously shown to enhance expression of *GATA3* in T-helper 2 cells [35], moving

us to speculate that up-regulation of *GATA3-AS1* may also enhance *GATA3* expression in cortical neurons. Interestingly, directly downstream of *GATA3-AS1*, there is a human-accelerated region (HAR; HAR-388), which is a region of DNA that has undergone rapid changes in the human lineage [48-50]. A study of cortical tissue from post-mortem human, rhesus, and mouse found that this HAR was one of 48 HARs genome-wide to act as a human-specific enhancer in the developing cortex [51]. Further work is required to determine if this HAR is directly causal in promoting human-specific expression patterns of *GATA3-AS1* and *GATA3* in human cortical neurons.

The coding sequence of GATA3 is highly conserved in primates; however, unique binding sites of GATA3 have been driven to rapid fixation in humans, as identified both in this study and previously in non-neuronal cell lines [46]. As polymorphism data becomes available for other nonhuman primates, it will be important to repeat this analysis to determine how selection is different between the human and NHP lineages on these transcription factor binding sites. The unique expression pattern and subsequent downstream species-specific regulation by an otherwise conserved transcription factor illuminate the shift in regulatory domains that have shaped human-specific neoteny. While we have identified many interesting genes bound by GATA3 that are known to be important for neuronal function (ex: *GABRB1*, *GRIN3A*, *APP*, *NLGN1*, *SYN2*), future work is required to identify which direct targets of GATA3 are responsible for the delayed maturation phenotype. Interestingly, in light of recent findings showing the importance of human-specific paralogs in neoteny in the neural progenitor state, we observed that many of these paralogs were directly bound by GATA3, indicating a potential role for GATA3 in their regulation. Together, our findings underscore the importance of identifying how human-specific gene regulation, including these newly identified paralogs, is brought into an already functional transcriptional network of an organism, potentially through the coevolution of transcription factor binding sites.

ACKNOWLEDGMENTS

This work was supported by the Brinson Foundation, the JPB Foundation, the Milky Way Research Foundation and NIH (R01 AG057706-05, 5 R01 AG056511-04, 5P30 AG062429-03). M. C. Marchetto and F.H. Gage are funded by the Larry L. Hillblom Foundation. The authors would like to thank M.L. Gage for editorial comments and T. Toda for insightful discussions that helped to shape the progress of this project. We would also like to thank Alan H. Nagahara and

Mark H. Tuszynski for providing rhesus macaque tissue. This work was supported by the NGS Core Facility of the Salk Institute with funding from NIH-NCI CCSG: P30 014195, the Chapman Foundation and the Helmsley Charitable Trust. RNA sequencing of neural progenitor cells (NPCs) was conducted at the IGM Genomics Center, University of California, San Diego, La Jolla, CA.

AUTHOR CONTRIBUTIONS

Conceptualization, MCM, SBL, IN, FHG; Methodology, SBL, MCM; Formal Analysis, SBL, JYH, KK, FQ, AS; Investigation, MW, RS, AM, RO, AS, EM, MCM; Data Curation, SBL; Writing- Original Draft, SBL, MCM; Writing- Reviewing and Editing, FHG, RS; Visualization, SBL, MCM; Supervision, MCM, FHG; Funding Acquisition, FHG, MCM.

DELARATIONS OF INTERESTS

The authors declare no competing interests.

FIGURE LEGENDS

Figure 1. All species transition similarly from neural progenitor cells (NPCs) to neurons

(A) Representative images of NPCs (top panels) and 4-week-old neurons (middle and bottom panels) for human, chimpanzee, bonobo, gorilla, and rhesus samples and quantification of the percentage of cells expressing the analyzed markers (right; n = independent experiments). Nestin-positive NPCs from all species similarly expressed the progenitor marker SOX2 (green). MAP2-positive neurons similarly expressed GFP from a synapsin promoter and exhibited equivalent endogenous levels of the deep layer neuronal marker CTIP2. Scale bar: 50 μm.

(B) Schematics showing the time points used for RNA extraction and next generation sequencing from human, chimpanzee, bonobo, gorilla and rhesus (timeline: 0, 2, 4, 6 and 8 weeks of neuronal differentiation).

(C) Principal component analysis (PCA) of all samples indicated a separation of cell state along PC1 and phylogeny along PC2; week 0 = NPC (orange highlight) and week > 0 = neuron (blue highlight).

(D) Heatmap of all cell state-dependent and species-independent genes (conserved expression).

(E) Representative genes with conserved expression dynamics, cyclin-dependent kinase 1 (*CDK1*) and synapsin (*SYN1*).

(F) Spatial analysis of genes with conserved expression patterns indicates an enrichment (dark red) in the subventricular zone (SVZ) within NPCs (top) and an enrichment in mature neuronal regions across multiple brain regions such as the pontine nucleus (PN) and the hypothalamus (HTH) (bottom); FC = fold-change. Boxplot whiskers for all plots represent 1.5 x interquartile range.

See also Figure S1.

Figure 2. A human-specific transcriptional signature develops throughout neuronal differentiation

(A) PC1 from a principal component analysis (PCA) of neural progenitor cells (NPCs) indicated a deviation of the human transcriptome from other primates during differentiation. ns = not significant, * = $p < 0.05$. Boxplot whiskers represent 1.5 * interquartile range.

(B) PC1 from a principal component analysis (PCA) of neurons indicated a deviation of the human transcriptome from other primates during differentiation. ns = not significant, * = $p < 0.05$. Boxplot whiskers represent 1.5 * interquartile range.

(C) The top dynamic, human-specific WGCNA module (up-regulated module-1) is shown with eigenvalues plotted for each species. H = human, B = bonobo, C = chimpanzee, G = gorilla, R = rhesus.
(D) Human-neuron correlation (HNC) score for the top correlated up-regulated module-1 genes. Size of each dot is associated with the human-specific effect size of each up-regulated module-1 genes.
(E) *GATA3* and *GATA3-AS1* RNA expression.
(F) Western blot for GATA3 in two-week-old neurons from three human individuals, as well as chimpanzee, bonobo, gorilla, and rhesus samples. Jurkat and 293T cells were included as positive controls for GATA3 expression.
(G) Scaled *GATA3* expression in the IZ from rhesus and human using data acquired from the Allen Brain Institute; * = $p < 0.05$. Boxplot whiskers represent 1.5 * interquartile range.
See also Figure S2.

Figure 3. GATA3 impacts gene expression in human neurons
(A) Presence of GATA3 (green) or input (gray) peaks normalized against input. * = GATA3-enriched regions $p < 0.05$.
(B) Example gene, *NOTCH2NL*, with enrichment of GATA3 signal at the 5'-UTR.
(C) Spatial enrichment of GATA3-bound genes with respect to the 15 post-conception-week-old postmortem human brain. Enrichment was largely in the SVZ with additional enrichment in the ventromedial hypothalamic nucleus (VMH), caudate nucleus (CA), and optic tract (OT).
(D) Overlap of genes bound by GATA3 in at least one human sample (Bound) and GATA3-regulated genes as indicated by shGATA3 expression (Regulated).
(E) PCA of all species using genes identified as GATA3-regulated within humans. Eigenvalues along PC1 (top) and PC2 (bottom) are combined into a single boxplot for each species. shGATA3: shG (black), shControl: shC (gray), Human no sh: green, Bonobo no sh: pink, Chimpanzee no sh: red, Gorilla no sh: purple, Rhesus no sh: blue. Boxplot whiskers represent 1.5 * interquartile range.
(F) Odds Ratio of the overlap between genes bound by GATA3 and Irritable Bowel Syndrome (IBD), Simons Foundation Autism Research Initiative (SFARI), and Deciphering Developmental Disorders (DDD)-associated genes.
See also Figures S3 and S4.

Figure 4. GATA3 regulates the rate of physiological maturity in human neurons

(A) Mean firing rate for non-infected human cells (Human), shControl (shC), and shGATA3 (shG) for the first week of recording. ns = not significant, * = $p < 0.05$.

(B) Raster plots for mean firing rate (top, red) and bursts (bottom, black) for the respective days examined in A.

(C) Mean firing rate and bursts recorded for one month following differentiation. Error bars = standard deviation.

See also Figure S4.

STAR★Methods

Resource availability

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Fred H. Gage (gage@salk.edu).

Materials availability

All unique/stable reagents generated in this study are available from the lead contact with a completed Materials Transfer Agreement.

Data and code availability

- RNA-seq data and ChIP-seq data generated in the current study have been deposited at GEO and are publicly available as of the date of publication (GEO repository with the accession number GSE120271).
- All code generated for this analysis is available at: https:/github.com/saralinker/multispecies_alignment
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

Experimental model and subject details

All experiments were performed with cells derived from three human individuals (WT126, WT33, Adrc40), one chimpanzee (PR00818), two bonobos (PR01086 and AG05253), two gorillas (PR00053 and PR00075), and one rhesus (A1) [26-28]. The use of fibroblasts from chimpanzee, bonobo and gorilla was approved by the US Fish and Wildlife Service, under permit MA206206. Protocols describing the use of iPSCs and human embryonic stem cells (hESCs) were previously approved by the University of California, San Diego (UCSD), the Salk Institute Institutional Review Board and the Embryonic Stem Cell Research Oversight Committee [27].

Methods details

Cell culture and neuronal differentiation

Fibroblasts from three neurotypical human subjects (WT126, WT33, Adrc40), one chimpanzee (PR00818), two bonobos (PR01086 and AG05253), two gorillas (PR00053 and PR00075), and one rhesus were reprogrammed to iPSCs and differentiated into NPCs as previously described

[26-28]. Established iPSC colonies were kept in feeder-free conditions on matrigel matrix, mTesr1 and passed using mechanical dissociation. Briefly, to obtain NPCs from iPSCs, embryoid bodies (EBs) were formed by mechanical dissociation of iPSC clusters and plated into low-adherence dishes in DMEM/F12 plus N2 and B27 medium in the presence of 500 ng/ml recombinant human noggin for forebrain induction for approximately 7 days. Then, floating EBs were plated onto poly-L-ornithine/laminin-coated dishes in DMEM/F12 plus N2 and B27 with the addition of 500 ng/ml noggin. Rosettes were collected after seven days and then dissociated with accutase and plated again onto coated dishes in DMEM/F12 plus N2 and B27 and 10 ng/ml of FGF2. Homogeneous populations of NPCs were obtained after one to two passages with accutase under the same conditions. To obtain neurons, NPCs were cultured in DMEM/F12 plus N2 and B27 with the addition of laminin (1 µg/ml), BDNF (20 ng/ml), GDNF (20 ng/ml) and cyclic AMP (500 µg/ml) for up to 8 weeks.

Immunofluorescence in vitro

Cells were fixed in 4% paraformaldehyde for 10 min and then permeabilized with 0.5% Triton-X100 in PBS buffer. Cells were then blocked in 0.5% Triton-X100 with 5% donkey serum for 1h before incubation with primary antibody overnight at 4°C. After three washes in PBS, cells were incubated with secondary antibodies (Jackson ImmunoResearch) for 1h at room temperature. Fluorescent signals were detected using a Zeiss inverted microscope and images were processed with Photoshop (Adobe Systems). Primary antibodies used in this study were human anti-Nestin (1:250, Millipore cat# ABD69); anti-SOX2 (1:200, Santa Cruz cat# sc-17320); anti-MAP2 (1:500, Abcam, cat# ab5392); anti-GFP (1:500, Invitrogen, cat# A6455); and anti-CTIP2 (1:200, Abcam, cat# ab18465). For neuron quantification, neurons were transduced with a lentivirus expressing EGFP under the control of the synapsin promoter. The Syn::EGFP reporter vector was obtained by cloning the Synapsin-1 promoter (a gift from Dr G. Thiel, Hamburg, Germany) in a lentivirus backbone.

Multi-electrode array analysis

Using the 96-well multi-electrode array (MEA) plate from Axion Biosystems, we plated cells derived from three human samples. Each well was seeded with 10,000 NPCs that were induced into neuronal differentiation. Each well was coated with poly-L-ornithine and laminin prior to cell seeding. Cells were fed three times a week and measurements were taken before media changes. Recordings were performed in a Maestro MEA system and AxIS software (Axion Biosystems) using a bandwidth with a filter for 10Hz to 2.5kHz cutoff frequencies. Spike detection

was performed using an adaptive threshold set to 5.5 times the standard deviation of the estimated noise on each electrode. Each plate rested for 10 min for acclimatization in the Maestro instrument and then was recorded for an additional 10 min to calculate spike and burst rates. Multi-electrode data analysis was performed using the Axion Biosystems Neural Metrics Tool, and an active electrode was considered once five spikes occurred over the length of 1 min (5 spikes/min). Bursts were identified in the data recorded from each individual electrode using an adaptive Poisson surprise algorithm. Network bursts were identified for each well using a non-adaptive algorithm requiring a minimum of 10 spikes with a maximum inter-spike interval of 100 ms.

RNA sequencing

Total cellular RNA was extracted from ~5x10^6 cells using the RNeasy Mini kit (Qiagen), according to the manufacturer's instructions. PolyA+ RNA was fragmented and prepared into sequencing libraries using the Illumina TruSeq RNA sample preparation kit. NPCs were analyzed on an Illumina HiSeq 2000 sequencer at the UCSD Biomedical Genomics Laboratory (BIOGEM). NPC libraries were sequenced using single-end 100 bp reads at a depth of 15-30 million reads per library (100 ± 25 bp fragments). Neurons were analyzed on an Illumina HiSeq 2500 sequencer at the Salk Institute's Next-Generation Sequencing Core. Neuron libraries were sequenced using paired-end 100 bp reads at a depth of 15-30 million reads per library (100 ± 25 bp fragments). Only read one of the paired-end reads was used for equivalency with the NPC libraries. Raw FASTQ files were trimmed for low quality reads using SolexaQA [52]. RNA-sequencing FASTQ files were aligned to the Human Hg19, *Pan troglodytes* PanTro4, *Gorilla gorilla* GorGor3, and *Macaca mulatta* RheMac3 reference using the STAR aligner [53]. All reads were filtered to retain only those that uniquely mapped to all four reference sequences, with a maximum of three mismatches to the species of origin and a maximum of five mismatches to the alternate species with no clipping. The filtered set of reads was annotated using the human genome coordinates and the respective Ensembl gene annotations. Counts per gene were calculated using the HTseq-count algorithm [54].

ChIP-sequencing

Approximately 0.5-1x10^7 cells were crosslinked with 1% formaldehyde for 10 min, followed by addition of 2.5 M glycine to the final concentration of 125 mM for 5 min at room temperature. Five µg GATA3 antibody (abcam: ab199428) was used for each IP. ChIP was performed as previously described with minor modifications [55]. Libraries were constructed with the NEBNext® Ultra™ II

DNA Library Prep Kit for Illumina (New England Biolabs). All ChIP-sequencing experiments were performed using an Illumina HiSeq 2500 Sequencer with 50 bp single-end reads and a minimum depth of 45 million reads per sample.

Western blot

Total protein extracts were prepared using RIPA buffer with protease and phosphatase inhibitors according to the procedure from Thermo Scientific. Proteins were separated by SDS-PAGE in NuPAGE™ precasted gels (4-12% Bis-Tris), with NuPAGE™ MOPS running buffer and transferred onto nitrocellulose membranes (Amersham Hybond). The primary antibodies used for immunoblotting were recombinant (rabbit) anti-GATA3 (1:1,000; abcam: ab199428) and monoclonal mouse anti-GAPDH (1:20,000; Santa Cruz: sc32233). Primary antibody binding was detected with horseradish peroxidase-conjugated anti-rabbit or anti-mouse and ECL luminescence solution (Millipore) as per the manufacturer's instructions.

Knockdown experiments

Three human NPC lines were infected with GATA3 lentiviral construct cloned from the SHCLNG-NM_002051 short hairpin GATA3 construct (Sigma Aldrich). Stable NPC lines were generated by selection with 1 µ/ml of puromycin over a one-week period. NPCs were then either collected for RNA sequencing or passaged for differentiation into neurons at a 1:6 dilution. Neurons were then fed three times a week for a total of either two or four weeks and the cells were collected for RNA sequencing.

Quantification and statistical analysis

Formatting of transcription counts

Counts were converted to Official Human Gene Symbol annotation using biomaRt (host = www.enembl.org, path = "ENSEMBL_MART_ENSEMBL", dataset = "hsapiens_gene_ensembl") in R [56]. The filter was set as "ensembl_gene_id" and "hgnc_symbol" as attributes. Genes with no conversion between ensembl and hgnc symbol were removed from downstream analysis. Expression counts were averaged across all ensembl annotations for genes with multiple ensembl annotations corresponding to a single hgnc symbol. Transcripts per million (TPM) were calculated similarly through biomaRt with the filter set to "hgnc_symbol" and attributes to "hgnc_symbol" and "transcript_length". All individual gene count values were first normalized by transcript length, returning a vector of transcript length-normalized counts. For each sample, the resulting vector was normalized by the total transcript length-normalized counts for all genes within the sample.

Values were then multiplied by 1e-06 to return a vector of TPM. TPM values were $\log_2$ + 1 transformed for downstream analyses.

Differential expression

All plots of TPM values were $\log_2$ + 1 transformed. Differential expression analyses were performed using either (1) a mixed-model regression analysis using lmer from the lme4 package with species as the random effect [57] or (2) as an exact test through edgeR [58]. TPM values were used as input for all mixed regression tests and raw counts were used for the exact test in edgeR. P-values for mixed regression analysis were calculated from t-values using the pt function in R. Raw p-values were adjusted for multiple-testing using the p.adjust function in R. The following models were used for mixed regression.

- Species-independent cell-state dependent: TPM ~ NPC + (1|species)
- Clade-dependent cell-state independent: TPM ~ clade + (1|species)
- Clade-dependent cell-state dependent: TPM ~ NPC + clade + NPC x clade + (1|species)

Principal components analysis

PCA was performed using pcaMethods [59] in R using $\log_2$ +1 transformed TPM values with the default method of svd. For PCA analyses examining all time points, NPCs only, or neurons only, all genes were used as input. For the PCA examining how shGATA3-regulated genes were expressed across NHP, genes identified as significantly different between shGATA3 and the combination of shControl and non-treated human samples (edgeR, padj < 0.05) was used for analysis.

Spatiotemporal analysis

Developmental timing and regional similarity estimates were calculated in comparison to the microarray dataset available from the Allen Brain Atlas (ABA) [31] using brainImageR as described previously [30]. In brief, for spatial analysis, the presence-absence file associated with each reference file was used to determine the reference set of genes expressed in each microdissected region. For each microdissected region, a hypergeometric distribution was used to test the significance of overlap between the user-defined gene set and the reference set. To plot gene-set enrichment, the values for each microdissected region were combined into one value representing the larger brain region shown in the brain plot. Fold change was calculated in comparison to random chance on 20 iterations. For temporal analysis, the full gene expression matrix was first scaled. Only postmortem samples less than 40 post-conception weeks of age

(prenatal) were used to train a random forest model predicting developmental time point of the sample using the randomForest library in R [60]. To refine the analysis, the random forest model was run twice. The first time returned a list of genes from the whole transcriptome with high likelihood of association with developmental time. The second round of random forest was refined by filtering out genes that were below the top 10% of informative strength in the initial model (MeanDecreaseGini). The final classifier was then used to predict the prenatal developmental time point of each sample.

Weighted gene co-expression network analysis

Genes increasing during differentiation in human samples were identified using the exact test from edgeR by comparing expression in NPCs to expression in neurons (log fold-change > 0). These genes were used as input to the weighted gene co-expression network analysis (WGCNA). The network analysis was performed with blockwiseModules with a maxBlockSize = 10000, power = 7, TOMType = “signed”, minModuleSize = 15, and reassignThreshold = 0. Module eigengenes were calculated with the moduleEigengenes function. The association between eigengene and phenotype was calculated with mixed model regression analysis (module eigengene ~ species + week + species x week). Modules with human-specific dynamics were defined as those with a p-value < 0.05 for the interaction terms between chimpanzee, bonobo, gorilla, or rhesus and neuronal state (NPC or neuron). Modules were named based on the analysis type and the rank order of significance (ex: first module built from genes up-regulated over time = up-regulated module 1).

ChIP-sequencing analysis

FATSQ files were aligned to the human reference genome (Hg19). Alignments were then converted to tag directories using the makeTagDirectory script in HOMER [61], and peaks were counted using the findPeaks function with -style factor. GATA3 peaks were called input correction from the corresponding individual. Peak annotations were identified using the annotatePeaks function.

SUPPLEMENTAL DATA AND TABLES

Data S1. Differentially expressed genes between human and nonhuman primate neurons, related to Figure 1.

Data S2. GATA3-bound genes, related to Figure 2.

Table S1. Human-specific paralogs bound by GATA3, related to Figure 3.

Table S2. GATA3-bound genes expressed in the subventricular and ventricular zones, related to Figure 3.

REFERENCES

1. Lui, J.H., Hansen, D.V., and Kriegstein, A.R. (2011). Development and evolution of the human neocortex. Cell *146*, 18-36.
2. Smart, I.H., Dehay, C., Giroud, P., Berland, M., and Kennedy, H. (2002). Unique morphological features of the proliferative zones and postmitotic compartments of the neural epithelium giving rise to striate and extrastriate cortex in the monkey. Cereb. Cortex *12*, 37-53.
3. Sousa, A.M.M., Meyer, K.A., Santpere, G., Gulden, F.O., and Sestan, N. (2017). Evolution of the human nervous system function, structure, and development. Cell *170*, 226-247.
4. Dennis, M.Y., Harshman, L., Nelson, B.J., Penn, O., Cantsilieris, S., Huddleston, J., Antonacci, F., Penewit, K., Denman, L., Raja, A., et al. (2017). The evolution and population diversity of human-specific segmental duplications. Nat. Ecol. Evol. *1*, 69.
5. Sudmant, P.H., Kitzman, J.O., Antonacci, F., Alkan, C., Malig, M., Tsalenko, A., Sampas, N., Bruhn, L., Shendure, J., Project, G., et al. (2010). Diversity of human copy number variation and multicopy genes. Science *330*, 641-646.
6. Florio, M., Albert, M., Taverna, E., Namba, T., Brandl, H., Lewitus, E., Haffner, C., Sykes, A., Wong, F.K., Peters, J., et al. (2015). Human-specific gene *ARHGAP11B* promotes basal progenitor amplification and neocortex expansion. Science *347*, 1465-1470.
7. Enard, W., Gehre, S., Hammerschmidt, K., Hölter, S.M., Blass, T., Somel, M., Brückner, M.K., Schreiweis, C., Winter, C., Sohr, R., et al. (2009). A humanized version of Foxp2 affects cortico-basal ganglia circuits in mice. Cell *137*, 961-971.
8. Heide, M., Haffner, C., Murayama, A., Kurotaki, Y., Shinohara, H., Okano, H., Sasaki, E., and Huttner, W.B. (2020). Human-specific *ARHGAP11B* increases size and folding of primate neocortex in the fetal marmoset. Science *369*, 546-550.
9. Pollard, K.S., Salama, S.R., King, B., Kern, A.D., Dreszer, T., Katzman, S., Siepel, A., Pedersen, J.S., Bejerano, G., Baertsch, R., et al. (2006). Forces shaping the fastest evolving regions in the human genome. PLoS Genet. *2*, e168.
10. Boyd, J.L., Skove, S.L., Rouanet, J.P., Pilaz, L.J., Bepler, T., Gordân, R., Wray, G.A., and Silver, D.L. (2015). Human-chimpanzee differences in a *FZD8* enhancer alter cell-cycle dynamics in the developing neocortex. Curr. Biol *25*, 772-779.

11. Kamm, G.B., López-Leal, R., Lorenzo, J.R., and Franchini, L.F. (2013). A fast-evolving human *NPAS3* enhancer gained reporter expression in the developing forebrain of transgenic mice. Philos. Trans. R. Soc. Lond. B-Biol. Sci. *368*, 20130019.
12. McLean, C.Y., Reno, P.L., Pollen, A.A., Bassan, A.I., Capellini, T.D., Guenther, C., Indjeian, V.B., Lim, X., Menke, D.B., Schaar, B.T., et al. (2011). Human-specific loss of regulatory DNA and the evolution of human-specific traits. Nature *471*, 216-219.
13. Benito-Kwiecinski, S., Giandomenico, S.L., Sutcliffe, M., Riis, E.S., Freire-Pritchett, P., Kelava, I., Wunderlich, S., Martin, U., Wray, G.A., McDole, K., et al. (2021). An early cell shape transition drives evolutionary expansion of the human forebrain. Cell *184*, 2084-2102.e2019.
14. Amlien, I.K., Fjell, A.M., Tamnes, C.K., Grydeland, H., Krogsrud, S.K., Chaplin, T.A., Rosa, M.G.P., and Walhovd, K.B. (2016). Organizing principles of human cortical development-thickness and area from 4 to 30 years: insights from comparative primate neuroanatomy. Cereb. Cortex *26*, 257-267.
15. Miller, D.J., Duka, T., Stimpson, C.D., Schapiro, S.J., Baze, W.B., McArthur, M.J., Fobbs, A.J., Sousa, A.M., Sestan, N., Wildman, D.E., et al. (2012). Prolonged myelination in human neocortical evolution. Proc. Natl. Acad. Sci. U. S. A. *109*, 16480-16485.
16. Somel, M., Franz, H., Yan, Z., Lorenc, A., Guo, S., Giger, T., Kelso, J., Nickel, B., Dannemann, M., Bahn, S., et al. (2009). Transcriptional neoteny in the human brain. Proc. Natl. Acad. Sci. U. S. A. *106*, 5743-5748.
17. Zhou, Y., Su, Y., Li, S., Kennedy, B.C., Zhang, D.Y., Bond, A.M., Sun, Y., Jacob, F., Lu, L., Hu, P., et al. (2022). Molecular landscapes of human hippocampal immature neurons across lifespan. Nature *607*, 527-533.
18. Suzuki, I.K., Gacquer, D., Van Heurck, R., Kumar, D., Wojno, M., Bilheu, A., Herpoel, A., Lambert, N., Cheron, J., Polleux, F., et al. (2018). Human-specific *NOTCH2NL* genes expand cortical neurogenesis through Delta/Notch regulation. Cell *173*, 1370-1384.e1316.
19. Marchetto, M.C., Hrvoj-Mihic, B., Kerman, B.E., Yu, D.X., Vadodaria, K.C., Linker, S.B., Narvaiza, I., Santos, R., Denli, A.M., Mendes, A.P., et al. (2019). Species-specific maturation profiles of human, chimpanzee and bonobo neural cells. Elife *8*, e37527.
20. Fiddes, I.T., Lodewijk, G.A., Mooring, M., Bosworth, C.M., Ewing, A.D., Mantalas, G.L., Novak, A.M., van den Bout, A., Bishara, A., Rosenkrantz, J.L., et al. (2018). Human-specific *NOTCH2NL* genes affect Notch signaling and cortical neurogenesis. Cell *173*, 1356-1369.e1322.

21. Kanton, S., Boyle, M.J., He, Z., Santel, M., Weigert, A., Sanchís-Calleja, F., Guijarro, P., Sidow, L., Fleck, J.S., Han, D., et al. (2019). Organoid single-cell genomic atlas uncovers human-specific features of brain development. Nature *574*, 418-422.
22. Otani, T., Marchetto, M.C., Gage, F.H., Simons, B.D., and Livesey, F.J. (2016). 2D and 3D stem cell models of primate cortical development identify species-specific differences in progenitor behavior contributing to brain size. Cell Stem Cell *18*, 467-480.
23. Schörnig, M., Ju, X., Fast, L., Ebert, S., Weigert, A., Kanton, S., Schaffer, T., Nadif Kasri, N., Treutlein, B., Peter, B.M., et al. (2021). Comparison of induced neurons reveals slower structural and functional maturation in humans than in apes. Elife *10*, e59323.
24. Britten, R.J., and Davidson, E.H. (1969). Gene regulation for higher cells: a theory. Science *165*, 349-357.
25. King, M.C., and Wilson, A.C. (1975). Evolution at two levels in humans and chimpanzees. Science *188*, 107-116.
26. Marchetto, M.C., Carromeu, C., Acab, A., Yu, D., Yeo, G.W., Mu, Y., Chen, G., Gage, F.H., and Muotri, A.R. (2010). A model for neural development and treatment of Rett syndrome using human induced pluripotent stem cells. Cell *143*, 527-539.
27. Marchetto, M.C.N., Narvaiza, I., Denli, A.M., Benner, C., Lazzarini, T.A., Nathanson, J.L., Paquola, A.C.M., Desai, K.N., Herai, R.H., Weitzman, M.D., et al. (2013). Differential L1 regulation in pluripotent stem cells of humans and apes. Nature *503*, 525-529.
28. Ramsay, L., Marchetto, M.C., Caron, M., Chen, S.H., Busche, S., Kwan, T., Pastinen, T., Gage, F.H., and Bourque, G. (2017). Conserved expression of transposon-derived non-coding transcripts in primate stem cells. BMC Genomics *18*, 214.
29. Langergraber, K.E., Prüfer, K., Rowney, C., Boesch, C., Crockford, C., Fawcett, K., Inoue, E., Inoue-Muruyama, M., Mitani, J.C., Muller, M.N., et al. (2012). Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. Proc. Natl. Acad. Sci. U. S. A. *109*, 15716-15721.
30. Linker, S.B., Hsu, J.Y., Pfaff, A., Amatya, D., Ko, S.M., Voter, S., Wong, Q., and Gage, F.H. (2018). BrainImageR: spatiotemporal gene set analysis referencing the human brain. Bioinformatics *35*, 343-345.
31. Hawrylycz, M.J., Lein, E.S., Guillozet-Bongaarts, A.L., Shen, E.H., Ng, L., Miller, J.A., van de Lagemaat, L.N., Smith, K.A., Ebbert, A., Riley, Z.L., et al. (2012). An anatomically comprehensive atlas of the adult human brain transcriptome. Nature *489*, 391-399.

32. Miller, J.A., Ding, S.L., Sunkin, S.M., Smith, K.A., Ng, L., Szafer, A., Ebbert, A., Riley, Z.L., Royall, J.J., Aiona, K., et al. (2014). Transcriptional landscape of the prenatal human brain. Nature *508*, 199-206.
33. Lake, B.B., Ai, R., Kaeser, G.E., Salathia, N.S., Yung, Y.C., Liu, R., Wildberg, A., Gao, D., Fung, H.L., Chen, S., et al. (2016). Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. Science *352*, 1586-1590.
34. Fan, X., Fu, Y., Zhou, X., Sun, L., Yang, M., Wang, M., Chen, R., Wu, Q., Yong, J., Dong, J., et al. (2020). Single-cell transcriptome analysis reveals cell lineage specification in temporal-spatial patterns in human cortical development. Sci. Adv. *6*, eaaz2978.
35. Gibbons, H.R., Shaginurova, G., Kim, L.C., Chapman, N., Spurlock, C.F., 3rd, and Aune, T.M. (2018). Divergent lncRNA *GATA3-AS1* regulates *GATA3* transcription in T-helper 2 cells. Front. Immunol. *9*, 2512.
36. Yan, W., Cao, Q.J., Arenas, R.B., Bentley, B., and Shao, R. (2010). GATA3 inhibits breast cancer metastasis through the reversal of epithelial-mesenchymal transition. J. Biol. Chem. *285*, 14042-14051.
37. Ho, I.C., Tai, T.S., and Pai, S.Y. (2009). GATA3 and the T-cell lineage: essential functions before and after T-helper-2-cell differentiation. Nat. Rev. Immunol. *9*, 125-135.
38. Lawoko-Kerali, G., Rivolta, M.N., and Holley, M. (2002). Expression of the transcription factors GATA3 and Pax2 during development of the mammalian inner ear. J. Comp. Neurol. *442*, 378-391.
39. Lawoko-Kerali, G., Rivolta, M.N., Lawlor, P., Cacciabue-Rivolta, D.I., Langton-Hewer, C., van Doorninck, J.H., and Holley, M.C. (2004). GATA3 and NeuroD distinguish auditory and vestibular neurons during development of the mammalian inner ear. Mech. Dev. *121*, 287-299.
40. Haugas, M., Tikker, L., Achim, K., Salminen, M., and Partanen, J. (2016). Gata2 and Gata3 regulate the differentiation of serotonergic and glutamatergic neuron subtypes of the dorsal raphe. Development *143*, 4495-4508.
41. Wei, G., Abraham, B.J., Yagi, R., Jothi, R., Cui, K., Sharma, S., Narlikar, L., Northrup, D.L., Tang, Q., Paul, W.E., et al. (2011). Genome-wide analyses of transcription factor GATA3-mediated gene regulation in distinct T cell types. Immunity *35*, 299-311.
42. Wan, Y.Y. (2014). GATA3: a master of many trades in immune regulation. Trends Immunol. *35*, 233-242.

43. Florio, M., Namba, T., Paabo, S., Hiller, M., and Huttner, W.B. (2016). A single splice site mutation in human-specific *ARHGAP11B* causes basal progenitor amplification. Sci. Adv. *2*, e1601941.
44. Charrier, C., Joshi, K., Coutinho-Budd, J., Kim, J.E., Lambert, N., de Marchena, J., Jin, W.L., Vanderhaeghen, P., Ghosh, A., Sassa, T., et al. (2012). Inhibition of SRGAP2 function by its human-specific paralogs induces neoteny during spine maturation. Cell *149*, 923-935.
45. Florio, M., Heide, M., Pinson, A., Brandl, H., Albert, M., Winkler, S., Wimberger, P., Huttner, W.B., and Hiller, M. (2018). Evolution and cell-type specificity of human-specific genes preferentially expressed in progenitors of fetal neocortex. Elife *7*, e32332.
46. Arbiza, L., Gronau, I., Aksoy, B.A., Hubisz, M.J., Gulko, B., Keinan, A., and Siepel, A. (2013). Genome-wide inference of natural selection on human transcription factor binding sites. Nat. Genet. 45, 723-729.
47. Gronau, I., Arbiza, L., Mohammed, J., and Siepel, A. (2013). Inference of natural selection from interspersed genomic elements based on polymorphism and divergence. Mol. Biol. Evol. *30*, 1159-1171.
48. Pollard, K.S., Salama, S.R., Lambert, N., Lambot, M.A., Coppens, S., Pedersen, J.S., Katzman, S., King, B., Onodera, C., Siepel, A., et al. (2006). An RNA gene expressed during cortical development evolved rapidly in humans. Nature *443*, 167-172.
49. Doan, R.N., Bae, B.I., Cubelos, B., Chang, C., Hossain, A.A., Al-Saad, S., Mukaddes, N.M., Oner, O., Al-Saffar, M., Balkhy, S., et al. (2016). Mutations in human accelerated regions disrupt cognition and social behavior. Cell *167*, 341-354.e312.
50. Lindblad-Toh, K., Garber, M., Zuk, O., Lin, M.F., Parker, B.J., Washietl, S., Kheradpour, P., Ernst, J., Jordan, G., Mauceli, E., et al. (2011). A high-resolution map of human evolutionary constraint using 29 mammals. Nature *478*, 476-482.
51. Reilly, S.K., Yin, J., Ayoub, A.E., Emera, D., Leng, J., Cotney, J., Sarro, R., Rakic, P., and Noonan, J.P. (2015). Evolutionary changes in promoter and enhancer activity during human corticogenesis. Science *347*, 1155-1159.
52. Cox, M.P., Peterson, D.A., and Biggs, P.J. (2010). SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. BMC Bioinformatics *11*, 485.
53. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15-21.
54. Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq-a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166-169.

55. Hawkins, R.D., Hon, G.C., Lee, L.K., Ngo, Q., Lister, R., Pelizzola, M., Edsall, L.E., Kuan, S., Luu, Y., Klugman, S., et al. (2010). Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. Cell Stem Cell *6*, 479-491.
56. R Development Core Team (2010). R: A Language and environment for statistical computing. R Foundation for Statistical Computing.
57. Bates, D., Machler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. J. Stat. Softw. *67*, 1-48.
58. Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139-140.
59. Stacklies, W., Redestig, H., Scholz, M., Walther, D., and Selbig, J. (2007). pcaMethods - a bioconductor package providing PCA methods for incomplete data. Bioinformatics *23*, 1164-1167.
60. Liaw A, W.M. (2002). Classification and regression by randomForest. R News *2*, 18-22.
61. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576-589.

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rat monoclonal anti-Ctip2	abcam	RRID:AB_2064130
Mouse monoclonal anti-GAPDH	Santa Cruz Biotechnology	RRID:AB_627679
Recombinant anti-GATA3 (rabbit)	abcam	RRID:AB_2819013
Rabbit polyclonal anti-GFP	ThermoFisher	RRID:AB_221570
Chicken polyclonal anti-MAP2	abcam	RRID:AB_2138153
Rabbit polyclonal anti-nestin	Millipore	RRID:AB_2744681
Goat polyclonal anti-SOX2	Santa Cruz Biotechnology	RRID:AB_2286684
Secondary antibodies	Jackson Immunoresearch	N/A
Secondary antibodies horseradish peroxidase-conjugated	Millipore	N/A
Chemicals, peptides, and recombinant proteins		
Accutase	STEMCELL	AT-104
B-27 supplement (50x), minus vitamin A	ThermoFisher Scientific	12587010
BDNF (Brain-derived neurotrophic factor)	Peprotech	450-02-1mg
Cyclic AMP (Dibutyryl cyclic-AMP sodium salt)	TOCRIS	1141
DMEM/F-12 GlutaMAX supplement	ThermoFisher Scientific (Gibco)	10565018
DMSO	Sigma-Aldrich/Merck	D2650
ECL luminescence solution	Millipore	
FGF2 (human)	Joint Protein Central	Human FGF2
GDNF (glial cell derived neurotrophic factor)	Peprotech	450-10-1mg
Laminin	Invitrogen	23017-015
Matrigel	Corning	354234
mTeSR1	STEMCELL	85850
N-2 supplement (100x)	ThermoFisher Scientific	17502048
NEBNext Ultra II DNA Library Prep Kit for Illumina	New England Biolabs	E7645
Noggin (human)	Proteintech	HZ-1118
NuPAGE MOPS running buffer	ThermoFisher Scientific	NP0001
Penicillin/Streptomycin	ScienCell	0503
Poly-L-ornithine	Sigma-Aldrich	P3655
Puromycin	Gibco	A11138-03
RIPA buffer with protease and phosphatase inhibitors	ThermoFisher Scientific	89900 78442
RNeasy Mini kit	Qiagen	74106
Rock Inhibitor, Y-27632	AdooQ	A21448
Deposited data		
RNA- and ChIP-sequencing	GEO repository	GSE120271
Experimental models: Cell lines		
Human WT33 (female) iPSC	Laboratory of genetics Salk Institute	N/A

Human WT126 (male) iPSC	Laboratory of genetics Salk Institute	N/A
Human ADRC40 (male) iPSC	Laboratory of genetics Salk Institute	RRID:CVCL_HA43
Chimpanzee PR00818 (female) iPSC	Laboratory of genetics Salk Institute	RRID:CVCL_2Z83
Bonobo PR01086 (male) iPSC	Laboratory of genetics Salk Institute	RRID :CVCL_2Z81
Bonobo AG05253 (male) iPSC	Laboratory of genetics Salk Institute	RRID:CVCL_2Z84
Gorilla PR00053 (male) iPSC	Laboratory of genetics Salk Institute	N/A
Gorilla PR00075 (female) iPSC	Laboratory of genetics Salk Institute	N/A
Rhesus macaque (male) iPSC	Laboratory of genetics Salk Institute	N/A
Software and algorithms		
biomaRt	R package, open-source	RRID:SCR_019214
brainImageR	R package, open-source	https://github.com/saralinker/brainImageR
edgeR	R package, open-source	RRID:SCR_012802
HOMER	software	RRID:SCR_010881
HTSeq-count	algorithm, open-source	RRID:SCR_011867
R	R Core team	https://www.r-project.org
randomForest	R package, open-source	RRID:SCR_015718
SolexaQA	R package, open-source	RRID:SCR_005421
STAR aligner	R package, open-source	RRID:SCR_004463
Other		
Lentiviral particles synapsin::EGFP	Available from Addgene	https://www.addgene.org/search/catalog/plasmids/?q=hSyn1+GFP
Lentiviral particles human shGATA3	Merck	SHCLNG-NM_002051
96-well multi-eletrode array plates	Axion Biosystems	M768-tMEA-96W
NuPAGE precasted gels (4-12% Bis-Tris)	Fisher Scientific	102470002

Figure 1

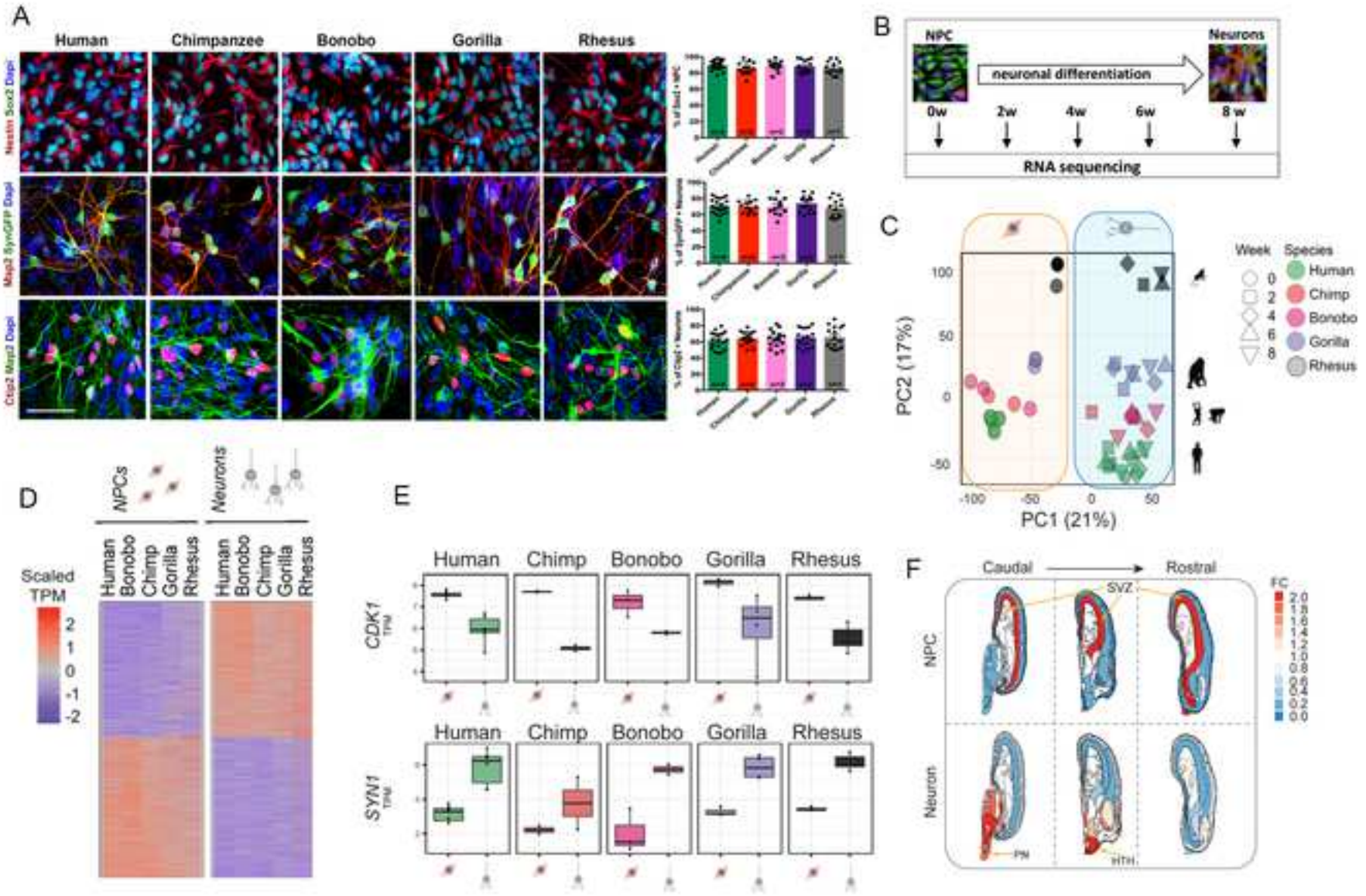

A
Human
Chimpanzee
Bonobo
Gorilla
Rhesus
B
NPC
Neurons
neuronal differentiation
0w
2w
4w
6w
8 w
RNA sequencing
C
PC2 (17%)
PC1 (21%)
Week
Species
Human
Chimp
Bonobo
Gorilla
Rhesus
D
NPCs
Neurons
Scaled TPM
E
CDK1
SYN1
F
Caudal
Rostral
SVZ
NPC
Neuron
FC

Figure 2

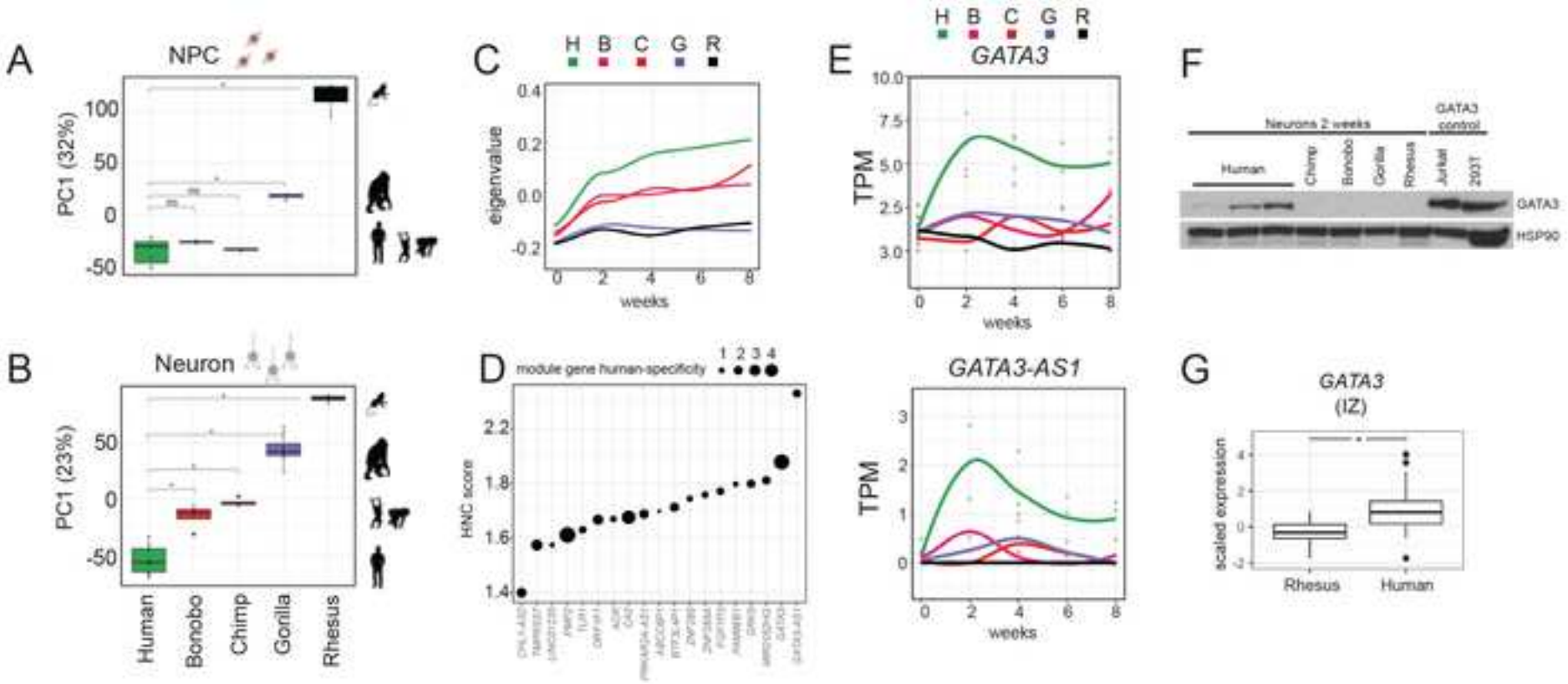

A
NPC
PC1 (32%)
100
50
0
-50
B
Neuron
PC1 (23%)
50
0
-50
Human
Bonobo
Chimp
Gorilla
Rhesus
C
H B C G R
eigenvalue
0.4
0.2
0.0
-0.2
0 2 4 6 8
weeks
D
module gene human-specificity
1 2 3 4
HNC score
2.2
1.8
1.6
1.4
E
H B C G R
GATA3
TPM
10.0
7.5
5.0
2.5
3.0
0 2 4 6 8
weeks
GATA3-AS1
TPM
3
2
1
0
0 2 4 6 8
weeks
F
Neurons 2 weeks
GATA3 control
Human
Chimp
Bonobo
Gorilla
Rhesus
Jurkat
293T
GATA3
HSP90
G
GATA3
(IZ)
scaled expression
Rhesus
Human

Figure 3

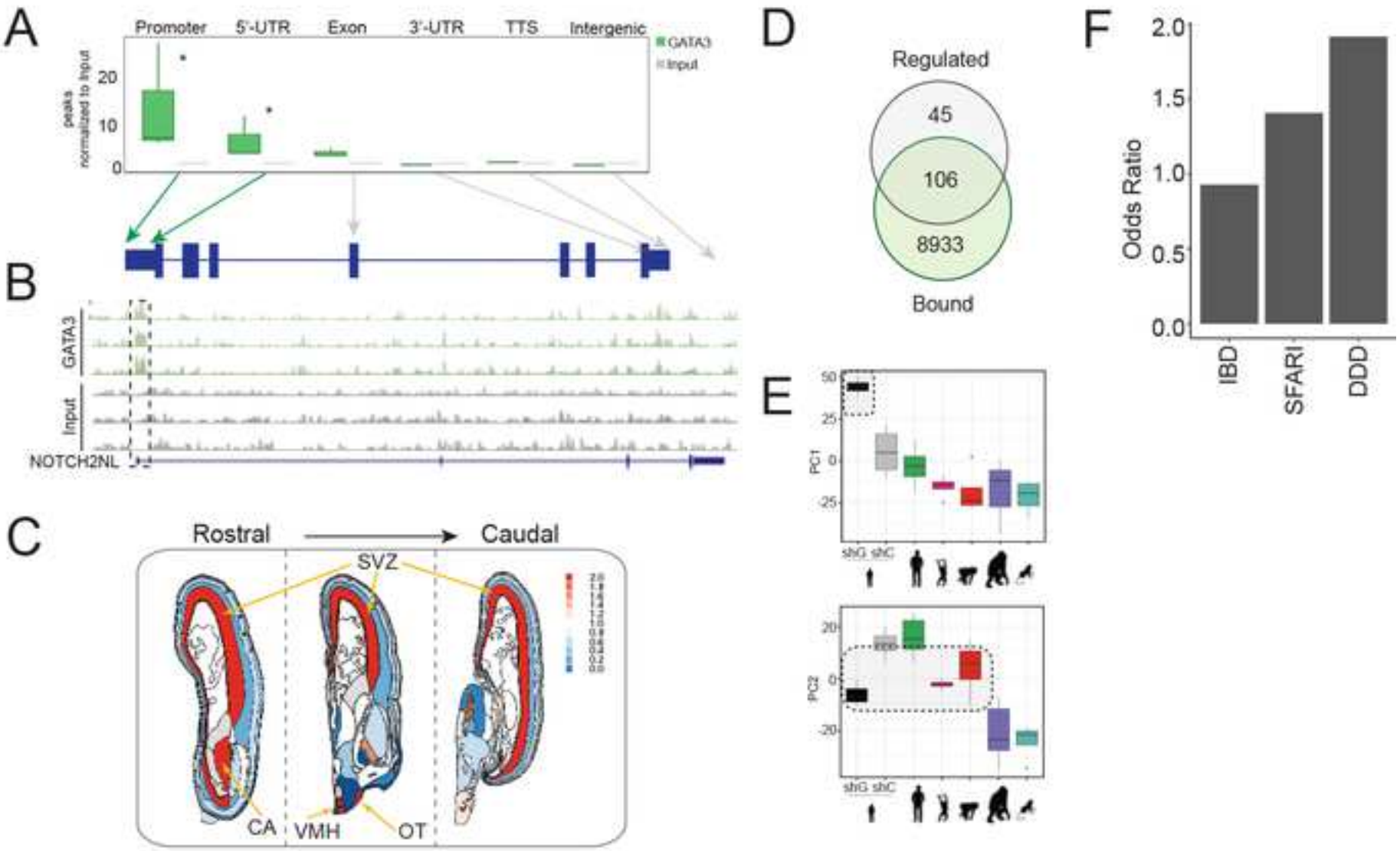
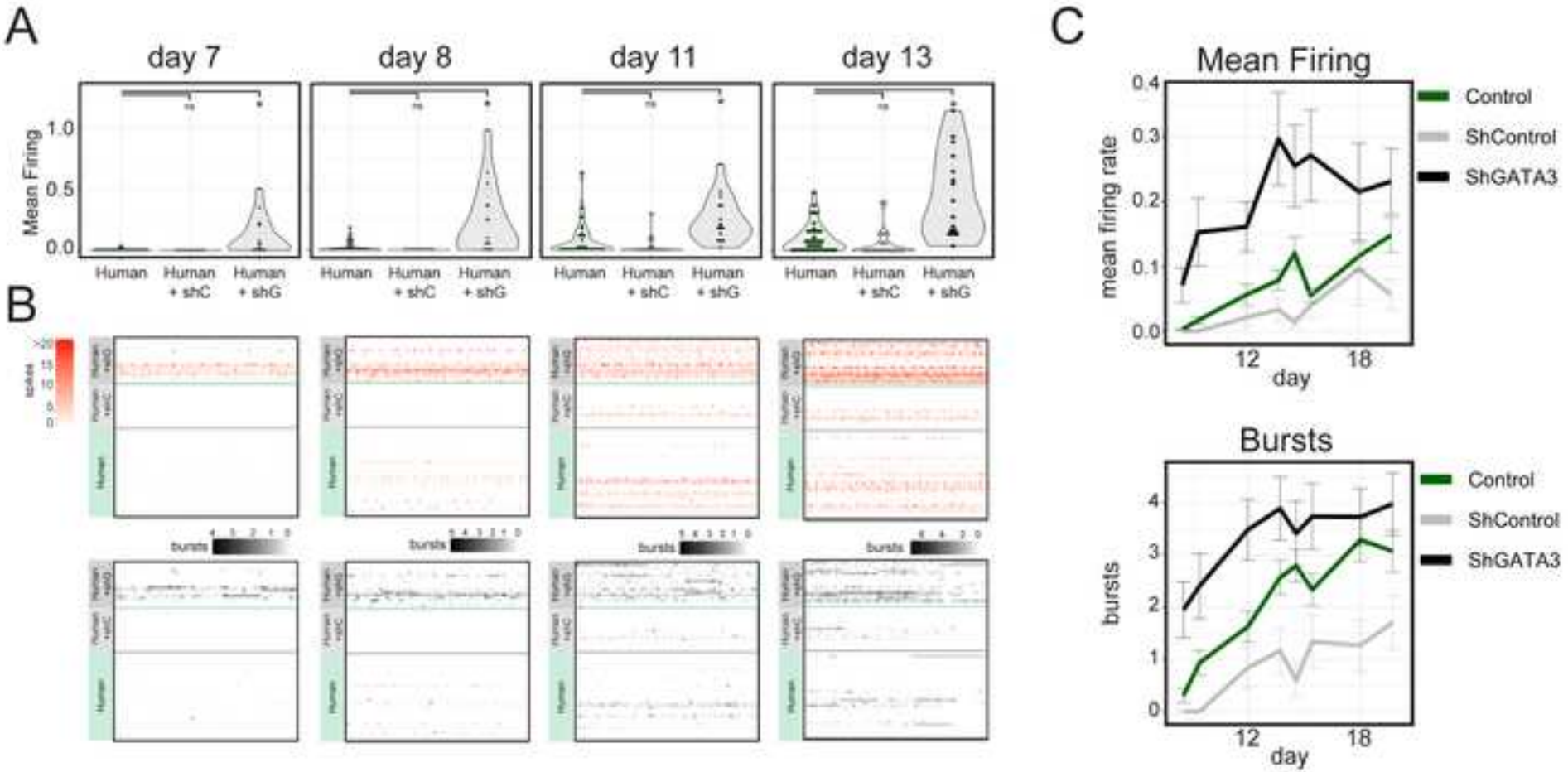

A
Promoter
5'-UTR
Exon
3'-UTR
TTS
Intergenic
GATA3
Input
peaks normalized to Input
20
10
0
B
GATA3
Input
NOTCH2NL
C
Rostral
Caudal
SVZ
CA
VMH
OT
D
Regulated
45
106
8933
Bound
E
PC1
PC2
shG
shC
F
Odds Ratio
2.0
1.5
1.0
0.5
0.0
IBD
SFARI
DDD

Figure 4


A
day 7
day 8
day 11
day 13
Mean Firing
1.0
0.5
0.0
Human
Human + shC
Human + shG
B
spikes
bursts
C
Mean Firing
mean firing rate
0.4
0.3
0.2
0.1
0.0
12
18
day
Control
ShControl
ShGATA3
Bursts
bursts
4
3
2
1
0
12
18
day
Control
ShControl
ShGATA3

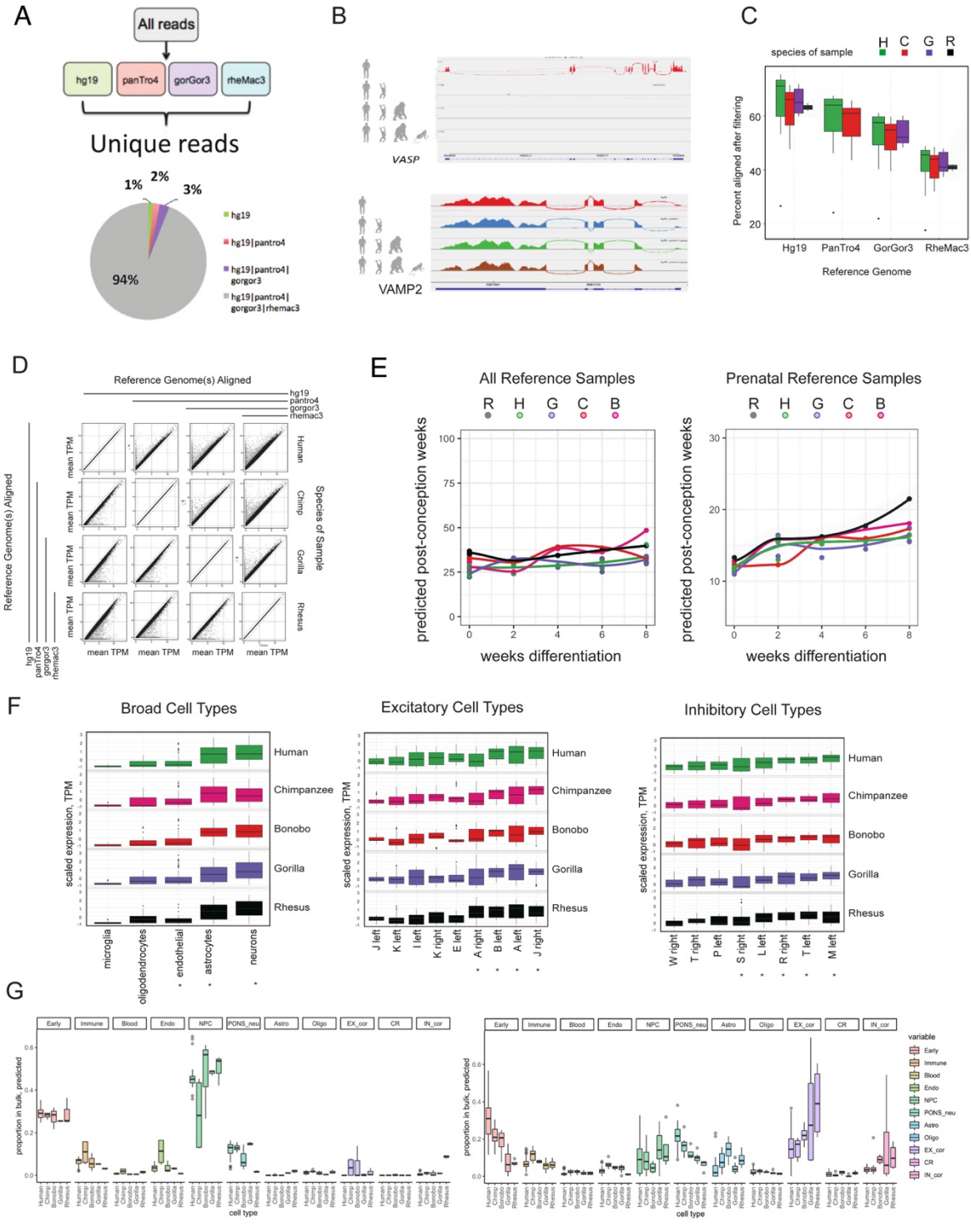


Figure S1. Species alignment strategy, related to Figure 1

(A) All reads for each sample were aligned to the human (hg19), chimpanzee (pantro4), gorilla (gorgor3), and rhesus (rhemac3) reference genomes. Reads that aligned uniquely in all 4 genomes were used for counting. All genes counts were annotated with respect to the human reference genome.

(B) Example genes from a single sample with reads aligned only to one reference (top, *VASP*) or equivalently to all 4 reference genomes (bottom, *VAMP2*).

(C) Percent reads aligned uniquely after each filtering step.

(D) Correlation of mean TPM for each species after filtering for each genome alignment.

(E) Predicted developmental time for all samples using the full transcriptome and the BrainSpan reference dataset of either 8 post-conception weeks to 40 years of age (left) or only prenatal reference samples (right). H = human, B = bonobo, C = chimpanzee, G = gorilla, R = rhesus.

(F) Total scaled expression for top marker genes of broad cell types (left), excitatory cell types (middle), or inhibitory cell types as defined in Lake et al. Science (2016) (right).

(G) Comparison of bulk RNA-seq to single-cell data from Fan et al. Science Advances (2020), which captures neurons from early gestation using CIBERSORT, Chen et al. Methods Mol Bio (2018). NPCs from all species (G-left) were most highly similar to early gestation NPCs followed by a signature for early neurons. Neurons from all species (G-right) have the highest overlap with early gestation neurons followed by excitatory cortical neurons then pons neurons and astrocytes.

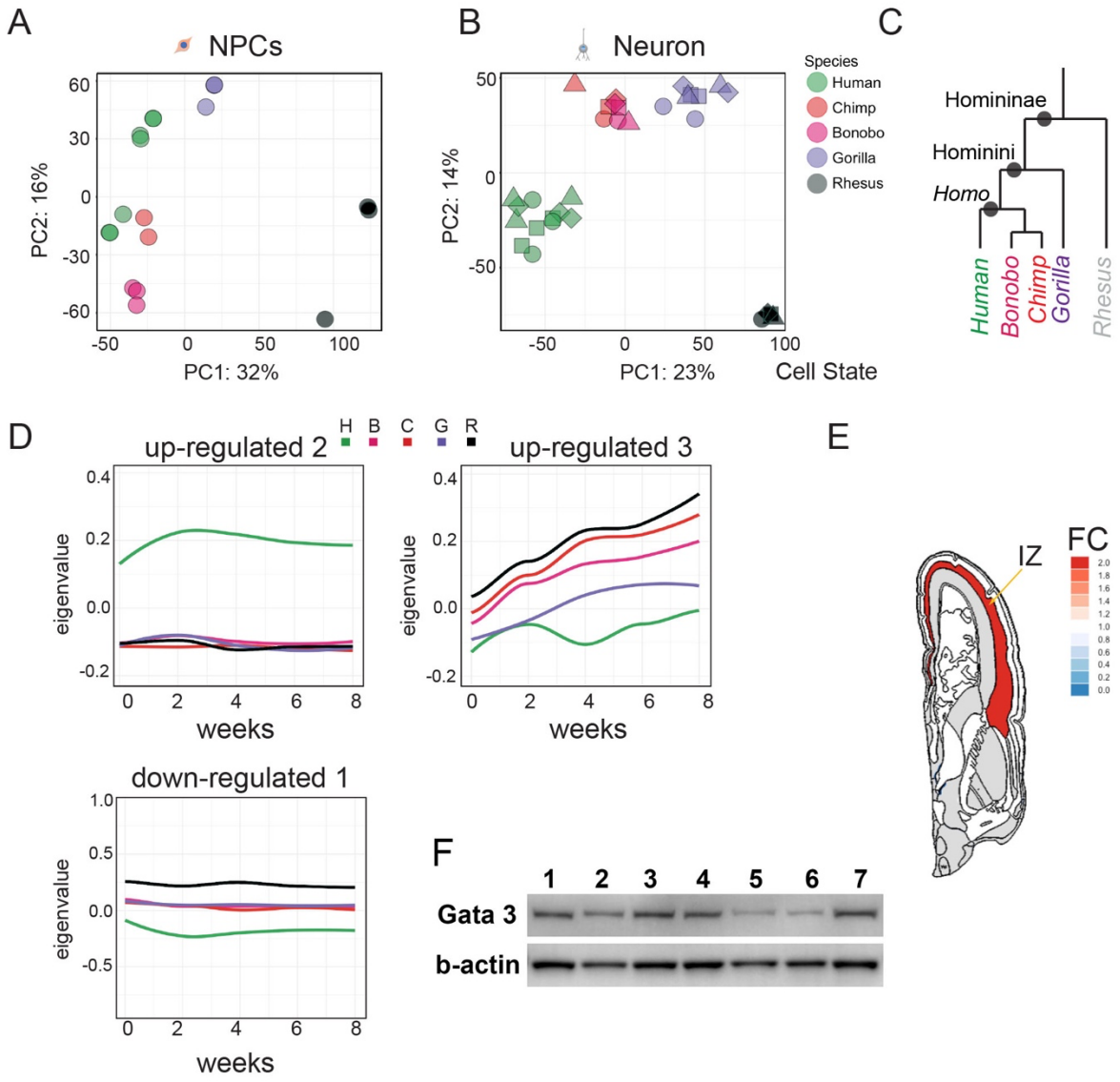


Figure S2. Developmental species-dependent transcriptional signature, related to Figure 2

(A) Principal component analysis (PCA) of NPCs from all species.

(B) Principal component analysis (PCA) of neurons from all species.

(C) Clades examined for differential expression analysis.

(D) Eigenvalues for human specific modules. H = human, B = bonobo, C = chimpanzee, G = gorilla, R = rhesus.

(E) Spatial enrichment of GATA3 expression.

(F) Western blot for GATA3 in 2-week-old human neurons. While results show variable expression levels of GATA3 protein, all human neurons expressed GATA3. 1-7 represent different lines of

human iPSC-derived neurons. 1-*WT33 clone1*; 2-*WT33 clone7*; 3-*WT126 clone5*; 4- *WT126 clone8*; 5- *ADRC40*; 6- *Cove*; 7- *Cent*.

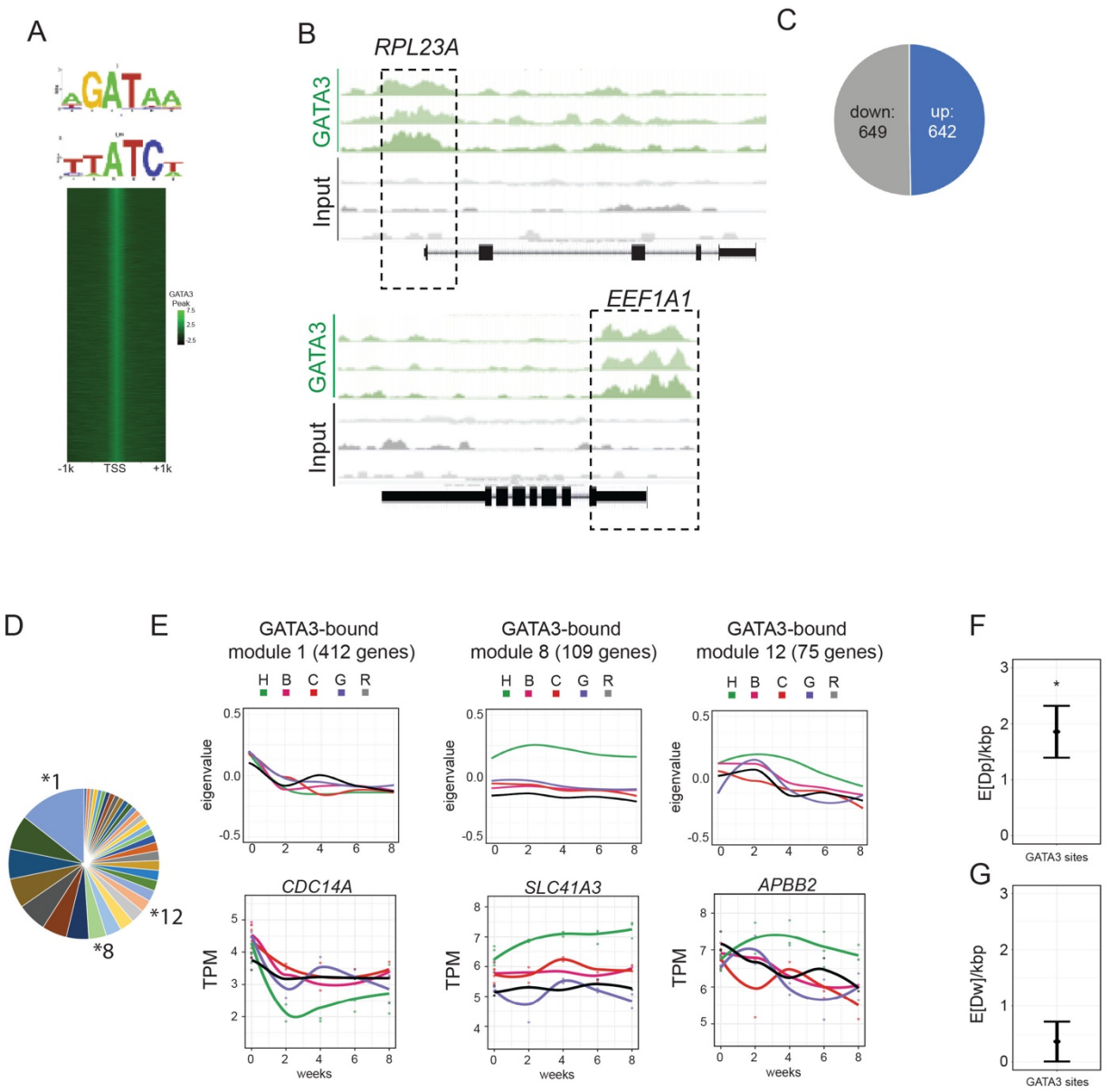


Figure 3S. GATA3 binding in human neurons, related to Figure 3

(A) GATA3 motif enriched within GATA3 ChIP-seq peaks and heatmap of GATA3 peak heights centered around transcription start sites (TSS).

(B) Example UCSC browser views of genes from families that contain human-specific paralogs bound at the promoter by GATA3 in human neurons. Green = GATA3 binding in lines generated from 3 separate human individuals; gray = input from the same 3 lines.

(C) Pie chart of GATA3-bound genes found to be significantly up- (up = 642 genes) or down-regulated (down= 649 genes) between human NPCs and neurons (p-adjusted < 0.05).

(D) Pie chart of the 37 modules identified through WGCNA analysis on only GATA3-bound genes. *# refers to the specific modules plotted in E.

(E) Three modules and representative genes from GATA3-bound WGCNA analysis showing the most common expression profile (module 1), and two human-specific profiles (modules 8 and 12).

(F) Estimation of divergence under strong positive (E[Dp]) selection at GATA3 motifs bound in human neurons. error bars = standard error, * = $p < 0.05$.

(G) Estimation of divergence under weak negative selection (E[Dw]) at GATA3 motifs bound in human neurons.

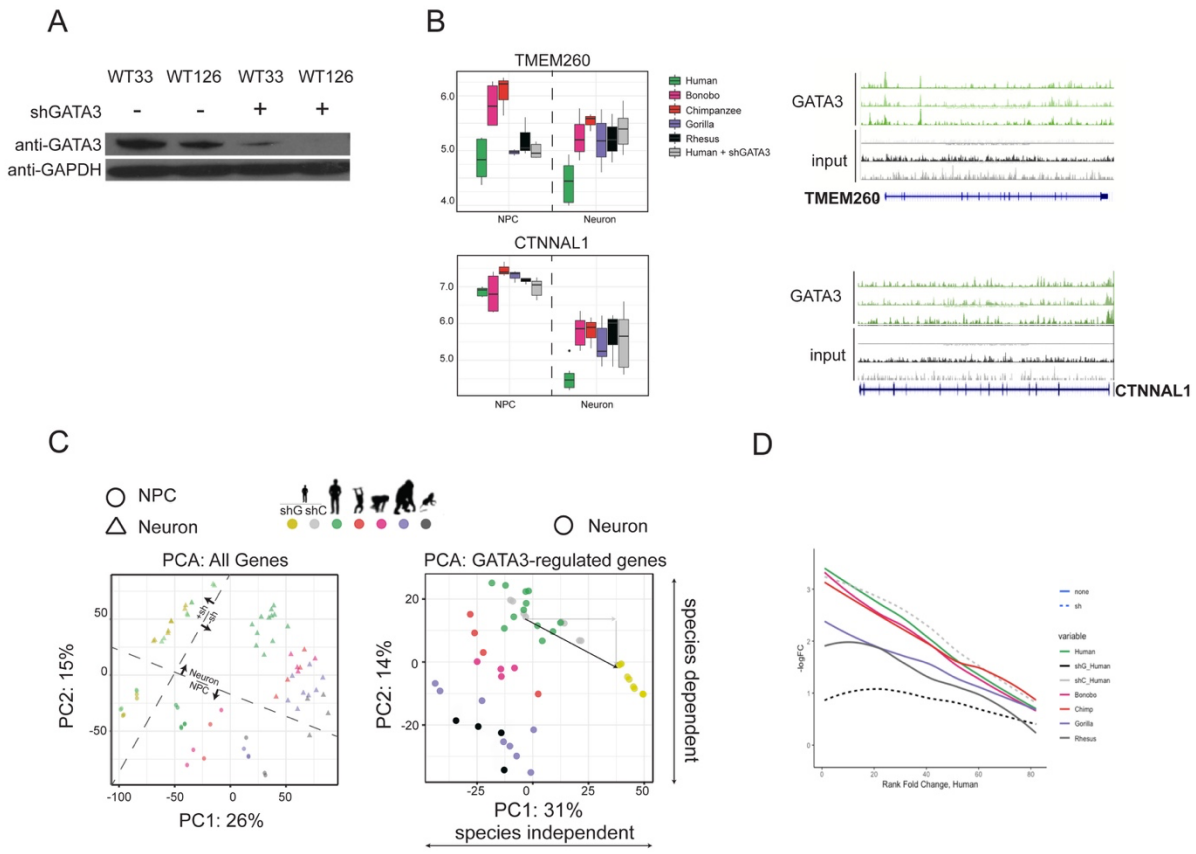


Figure S4. GATA3 knockdown in human neurons, related to Figures 3 and 4

(A) Western blot of GATA3 following shGATA3 knockdown in human samples (WT33 and WT126).

(B) Example genes with human specific expression patterns that are expressed more similarly to nonhuman primate samples after shGATA3 treatment (left) and which are directly bound by GATA3 (right).

(C) PCA of all samples either using all genes (left) or restricted to genes identified as regulated by GATA3 (right). Arrows indicate vector of the shift of shGATA3 samples in comparison to controls.

(D) logFC between expression NPCs and neurons (combined) of genes directly bound by GATA3 in ChIP-seq and modulated by GATA3 knockdown (y-axis) versus the rank of the logFC of the gene in human samples (x-axis).