Article

Molecular features driving cellular complexity of human brain evolution

https://doi.org/10.1038/s41586-023-06338-4

Received: 24 March 2022

Accepted: 16 June 2023

Published online: 19 July 2023

Check for updates

Emre Caglayan[1,2,9], Fatma Ayhan[1,2,9], Yuxiang Liu[1,2], Rachael M. Vollmer[1,2], Emily Oh[1,2], Chet C. Sherwood[3], Todd M. Preuss[4,5], Soojin V. Yi[6,7,8]✉ & Genevieve Konopka[1,2]✉

Human-specific genomic changes contribute to the unique functionalities of the human brain[1–5]. The cellular heterogeneity of the human brain[6,7] and the complex regulation of gene expression highlight the need to characterize human-specific molecular features at cellular resolution. Here we analysed single-nucleus RNA-sequencing and single-nucleus assay for transposase-accessible chromatin with sequencing datasets for human, chimpanzee and rhesus macaque brain tissue from posterior cingulate cortex. We show a human-specific increase of oligodendrocyte progenitor cells and a decrease of mature oligodendrocytes across cortical tissues. Human-specific regulatory changes were accelerated in oligodendrocyte progenitor cells, and we highlight key biological pathways that may be associated with the proportional changes. We also identify human-specific regulatory changes in neuronal subtypes, which reveal human-specific upregulation of *FOXP2* in only two of the neuronal subtypes. We additionally identify hundreds of new human accelerated genomic regions associated with human-specific chromatin accessibility changes. Our data also reveal that FOS::JUN and FOX motifs are enriched in the human-specifically accessible chromatin regions of excitatory neuronal subtypes. Together, our results reveal several new mechanisms underlying the evolutionary innovation of human brain at cell-type resolution.

Phenotypic differences between humans and our closest extant relatives, including chimpanzees and other great apes, are driven by a combination of regulatory and coding sequence changes[1]. These genomic underpinnings of human brain evolution can be explained by genome-wide comparisons with non-human primate species. Previous studies have profiled the transcriptome of brain tissues in bulk to identify human-specific gene expression changes[2–4]. The findings of these studies highlighted human-specific changes that included changes in synaptogenesis[3,5] and myelination[5,8]. However, brain tissue has tremendous cellular heterogeneity[6,9]. Therefore, single-cell genomics approaches are required to identify the full scope of human-specific gene regulatory changes. Although previous studies have explored comparisons of epigenome or transcriptome between humans and other species[7,10–15], a systematic identification of human-specific epigenomic and transcriptomic changes at cellular resolution is lacking. To address this gap of knowledge and assign changes to the human lineage, here we profiled the transcriptomes and epigenomes of adult tissue from posterior cingulate cortex from humans and chimpanzees (*Pan troglodytes*) and included rhesus macaques (*Macaca mulatta*) as an outgroup. Our single-nucleus RNA-sequencing (snRNA-seq) and single-nucleus assay for transposase-accessible chromatin with sequencing (snATAC-seq) results revealed significant changes in proportions of cells in the oligodendrocyte lineage, uncovering thousands of human-specific regulatory changes. We further assessed the association of these regulatory changes with the underlying human-specific substitutions, providing critical links between changes in DNA sequence and function in human brain evolution at cellular resolution. We also uncovered specific enrichment of motifs of the immediate early gene transcription factors (TFs) FOS and JUN in human-specific chromatin accessibility gains, indicating human specificity in activity-dependent gene regulation. These results shed light on previously unknown cellular dimensions of human brain evolution.

To identify molecular and cellular changes accompanying human brain evolution, we examined the evolution of Brodmann area 23 (BA23) by applying snRNA-seq and snATAC-seq approaches to the same samples. Notably, BA23 is part of the posterior cingulate cortex, a hub region in the default mode network[16] that is involved in higher-order cognitive processes such as theory of mind and has been implicated in schizophrenia[17]. Despite such importance, at present there is no detailed study of BA23 at single-cell resolution. We detected 148,540 nuclei using snRNA-seq (human: 41,397; chimpanzee: 53,539; macaque: 53,604) and 73,486 nuclei using snATAC-seq (human: 28,630; chimpanzee: 20,703; macaque: 24,153) after quality control (Methods, Extended Data Fig. 1 and Supplementary Table 1). We annotated major cell types

[1]Department of Neuroscience, UT Southwestern Medical Center, Dallas, TX, USA. [2]Peter O'Donnell Jr. Brain Institute, UT Southwestern Medical Center, Dallas, TX, USA. [3]Center for the Advanced Study of Human Paleobiology, Department of Anthropology, The George Washington University, Washington, DC, USA. [4]Division of Neuropharmacology and Neurologic Diseases, Yerkes National Primate Research Center, Emory University, Atlanta, GA, USA. [5]Department of Pathology, Emory University School of Medicine, Atlanta, GA, USA. [6]Department of Ecology, Evolution, and Marine Biology, University of California, Santa Barbara, Santa Barbara, CA, USA. [7]Department of Molecular, Cellular and Developmental Biology, University of California, Santa Barbara, Santa Barbara, CA, USA. [8]Neuroscience Research Institute, University of California, Santa Barbara, Santa Barbara, CA, USA. [9]These authors contributed equally: Emre Caglayan, Fatma Ayhan. ✉e-mail: soojinyi@ucsb.edu; genevieve.konopka@utsouthwestern.edu

(Extended Data Figs. 1 and 2) and neuronal subtypes (14 excitatory subtypes and 8 inhibitory subtypes; Extended Data Fig. 3) across species in both snRNA-seq and snATAC-seq (Methods).

Proportional changes in oligodendrocytes

Evolutionary changes can arise from proportional[13] and/or gene regulatory[2,5,10–12,18] changes of cell types. Compared to that of non-human primates, the human brain has prolonged myelination and altered gene regulation in the oligodendrocyte lineage[5,8,18], indicating possible changes in human-specific cell-type abundances. Assessing the proportional changes of the oligodendrocyte lineage in single-cell genomics can be particularly challenging as glia express fewer transcripts than neurons as evidenced by fewer unique molecular identifiers (UMIs; Extended Data Fig. 1e) and are thus more prone to filtering during empty-droplet removal with a UMI cutoff. To overcome this bias, we used a low UMI cutoff after empty-droplet removal (Methods) and calculated the percentage of mature oligodendrocytes (MOLs) and oligodendrocyte progenitor cells (OPCs) compared to all glia. We found a human-specific increase in OPC abundance and a human-specific decrease in MOL abundance whereas the abundances of astrocytes and microglia were not significantly altered (Fig. 1a and Supplementary Table 2). To confirm this finding using an independent method that preserves tissue anatomy, we carried out single-molecule fluorescence in situ hybridization (smFISH), which validated a significant increase of OPC and a significant decrease of MOL populations in humans compared to chimpanzees (Fig. 1b–d). We then examined data from other cortical regions that were previously profiled. Reanalysis of a snRNA-seq dataset from anterior cingulate cortex[11] yielded a concordant result with our finding (Fig. 1e and Extended Data Fig. 4a,b). We further validated this with smFISH using anterior cingulate cortical tissue from humans and chimpanzees (Fig. 1f–h). As we quantified the signal from all layers of the cortex, we also divided the columnar images into sections, which revealed similar trends in both cortical regions (Extended Data Fig. 4c–f), indicating that the result is not driven by uneven sampling of cortical layers. In addition, we examined a bulk RNA-sequencing dataset of the entire oligodendrocyte lineage in the dorsolateral prefrontal cortex[18]. Using deconvolution, we found a higher OPC/MOL ratio in humans, regardless of which species was used as reference (Fig. 1i and Extended Data Fig. 4g). Reanalysis of a comparative dataset[7] in primary motor cortex tissue yielded similarly increased proportions of OPCs and decreased proportions of MOLs in humans compared to marmosets (Fig. 1j) and a similar trend compared to rhesus macaques (Fig. 1k). Notably, we did not observe similar abundance changes in the caudate nucleus[11] or dentate gyrus[19] (Extended Data Fig. 4h,i). Together, these results show that adult human brain cortical regions have proportionally more OPCs and fewer MOLs compared to those of non-human primates.

Gene regulatory changes in OPCs

To understand the gene regulatory changes in the human lineage, we identified human-specific gene expression alterations (HS-Genes: HS-Up-Genes and HS-Down-Genes) and human-specific chromatin accessibility level alterations in *cis*-regulatory elements (HS-CREs: HS-Open-CREs and HS-Closed-CREs) per cell type (Methods, Extended Data Figs. 4j and 5a,b and Supplementary Tables 3 and 4). Focusing on the oligodendrocyte lineage, we found a greater relative abundance of human-specific (HS) changes compared to chimpanzee-specific (CS) changes in OPCs than in MOLs in both snRNA-seq and snATAC-seq (Fig. 2a). Applying a similar approach to the anterior cingulate cortex[11] revealed similarly accelerated evolution in OPCs (Fig. 2b), as well as a significant overlap with our results (Extended Data Fig. 4k,l).

Among the human-specific regulatory changes in OPCs, HS-Down-Genes are enriched in cytoskeletal activity (Fig. 2c and Supplementary Table 5), which is crucial for OPC migration and oligodendrocyte differentiation[20]. We posited that marker genes in committed oligodendrocyte progenitors (COPs) may also have been altered in human OPCs. We identified 15 COP marker genes that are common across all species in our dataset and in two additional human datasets[21–23]. Two COP markers, *SH3RF3* and *KIF21A*, were HS-Down-Genes in OPCs (Fig. 2d–f). In line with the enrichment of cytoskeletal genes, *KIF21A* encodes a kinesin motor protein that is involved in microtubule function, whereas *SH3RF3* encodes a SH3-domain-containing protein with ubiquitin ligase activity, a process also implicated in oligodendrocyte maturation[24]. We also identified an HS-Closed-CRE in OPCs near the transcription start site (TSS) of *SH3RF3*, potentially linked to the human-specific downregulation of this gene (Fig. 2g). Notably, snRNA-seq profiles for the frontal cortex of adult mice showed that most primate COP markers exhibit upregulation in COPs or newly formed oligodendrocytes compared to OPCs, except for *Sh3rf3*, indicating potential primate specificity (Fig. 2h–j). Together, these results highlight key regulatory changes in human OPCs that may underlie human-specific proportional changes in the oligodendrocyte lineage.

Neuronal subtype specificity of evolution

We identified 14 subtypes of excitatory neurons and 8 subtypes of inhibitory neurons across species in both snRNA-seq and snATAC-seq (Extended Data Fig. 3). Unlike the results for the oligodendrocyte lineage, our findings showed that the neuronal subtype abundances were largely conserved across species (Extended Data Fig. 3b,h and Supplementary Table 2). The rates of gene regulatory changes were similar between human and chimpanzee lineages across most subtypes (Extended Data Fig. 5c,d). However, a few neuronal subtypes exhibited signatures of human-specific acceleration in the epigenome (for example, L2-3_1) or the transcriptome (for example, L5-6 FEZF2_1; Extended Data Fig. 5c,d).

We observed a high heterogeneity of HS changes among neuronal subtypes (Extended Data Fig. 5e). As most previous comparative studies lacked cellular resolution at the subtype level, we assessed reproducibility between the previous bulk comparisons[12,18] and the subtype-resolved comparisons. Although we found an overall enrichment between the species-specific genes across different studies (Extended Data Fig. 6a–c), bulk studies consistently showed low overlap with the more subtype-specific HS changes (Extended Data Fig. 6d,e). Notably, when we pooled the excitatory subtypes, our power to detect subtype-specific HS changes were also substantially reduced (Extended Data Fig. 6f,g). Therefore, most neuronal HS changes are not shared by more than a few subtypes and are masked in bulk approaches.

Subtype-specific evolution of *FOXP2*

We examined human-specific expression of TFs that are altered in only a few subtypes and found that *FOXP2*, which encodes a key TF known for its roles in the development of cortical–striatal circuits related to speech and language and human brain evolution[25,26], showed human-specific upregulation in two excitatory subtypes (Fig. 3a). This contrasted with the previous comparative studies of adult cortex that did not find a significant difference in the *FOXP2* expression level between human and chimpanzee neurons[11,12,14,18]. Among these two subtypes, the L5-6_THEMIS_1 subtype (the most abundant *THEMIS*$^+$ subtype, also marked by *C1QL3*; Extended Data Fig. 3e) exhibited low levels of *FOXP2* expression in non-human primates (Fig. 3a). We used smFISH to independently validate this finding in intact tissues and confirmed both more *FOXP2* and *THEMIS* co-positive cells in humans compared to chimpanzees (Fig. 3b,c), and more *FOXP2*$^+$ puncta in human *THEMIS*$^+$ cells but not in *THEMIS*$^-$ cells (Fig. 3b,d). A recent study found similar *FOXP2* levels across species in all neuronal subtypes of

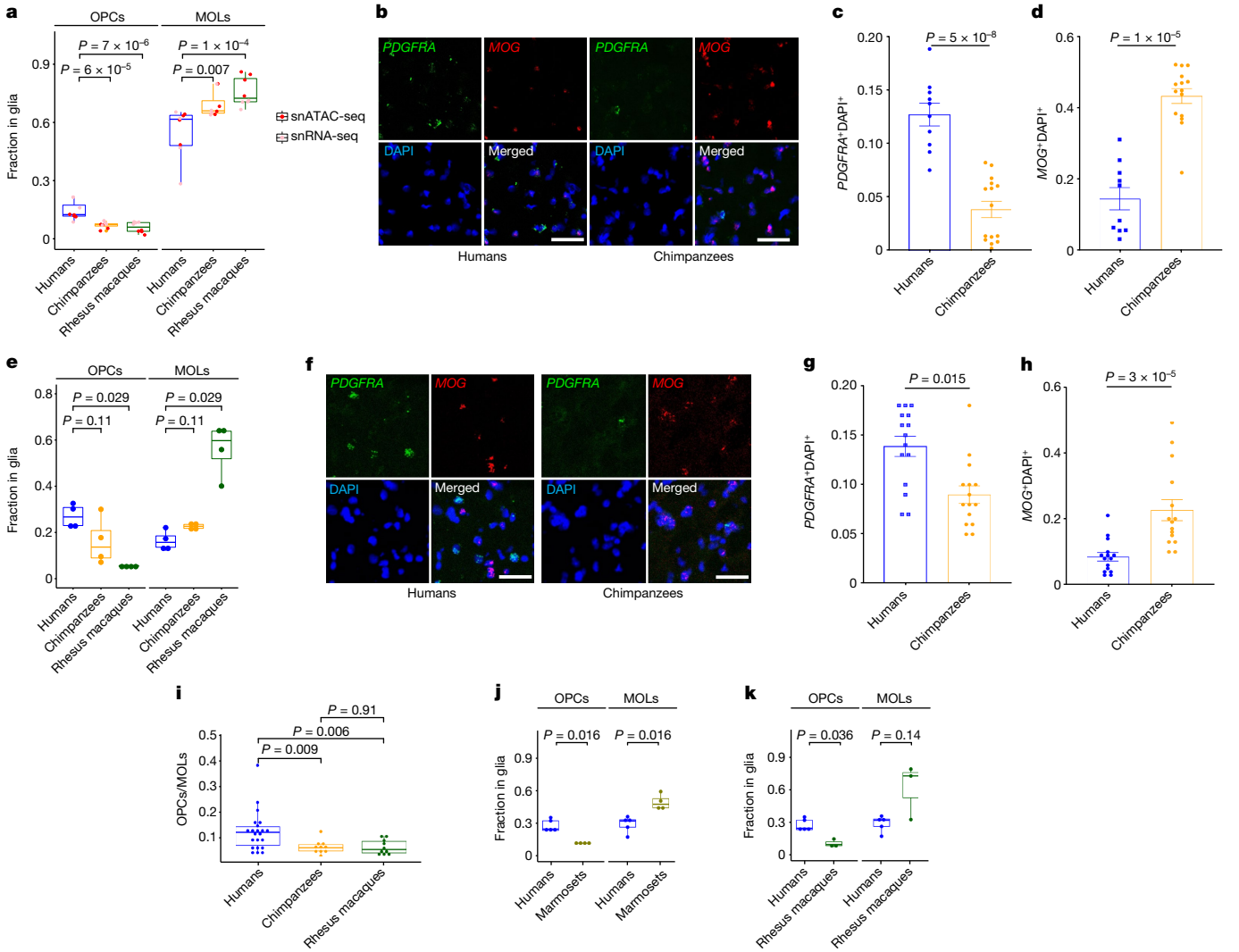


Fig. 1 | The oligodendrocyte lineage is proportionally altered in human evolution. a, The fractions of OPCs and MOLs in glia from posterior cingulate cortex (BA23) are altered in humans. Each dot represents a sample (red: snATAC-seq, pink: snRNA-seq; $n = 4$ individuals per species per assay; P value: likelihood ratio test, two-sided; Methods). **b**–**d**, smFISH shows increased *PDGFRA* (OPCs) and decreased *MOG* (MOLs) signals in humans compared to chimpanzees (region: posterior cingulate cortex). DAPI, 4′,6-diamidino-2-phenylindole. **b**, Representative images. Scale bars, 100 µm. **c**,**d**, Quantification of the fraction of OPCs and MOLs. Each data point is the average of all sub-areas in a section (2–4 sub-areas in each of 5 sections per individual; human: 10 sections; chimpanzee: 15 sections; Methods). The P value is the main effect of species from a linear mixed model (random effect: individual, two-sided). Error bars represent s.e.m. **e**, The fraction of OPCs or MOLs in glia based on snRNA-seq from anterior cingulate cortex ($n = 4$ individuals per species, P value: Wilcoxon rank sum test, two-sided). **f**, as in **b**, but for anterior cingulate cortex. **g**,**h**, Quantification of the fraction of OPCs and MOLs as in **c**,**d**, but for anterior cingulate cortex (human: 15 sections, chimpanzee: 15 sections; Methods). **i**, Deconvoluted proportions of cells from OLIG2 expressing bulk RNA-seq dataset for dorsolateral prefrontal cortex (reference dataset: human snRNA-seq from this study, $n = 22$ (human), 10 (chimpanzee), 10 (rhesus macaque) individuals; P value: Wilcoxon rank sum test, two-sided). **j**,**k**, Fraction of OPCs or MOLs in glia per species in the primary motor cortex. **j**, Human–marmoset comparison, **k**, human–rhesus macaque comparison. $n = 5$ (human), 4 (marmoset), 3 (rhesus macaque) individuals. P value: Wilcoxon rank sum test, two-sided. Boxplots represent median and interquartile range in **a**,**e**,**i**–**k**.

the dorsolateral prefrontal cortex[14]. Corroborating this result, we also found significantly lower levels of *FOXP2* in the *THEMIS*$^+$*C1QL3*$^+$ neurons of prefrontal cortex and anterior cingulate cortex in an independent dataset[23] (Fig. 3e,f). These results indicate that subtype-specific upregulation of *FOXP2* is also brain region specific. Notably, some of the experimentally validated *FOXP2* downstream targets (*VLDLR*, *SRPX2*, *CNTNAP2*, *MET* and *DISC1*)[25] are not human-specifically altered in these two subtypes, indicating potentially distinct FOXP2 gene regulation among neuronal subtypes in the cortex (Fig. 3a and Supplementary Table 3). Two previously identified FOXP2 targets, *CNTNAP2* and *MET*, are human-specifically upregulated in layer 4 subtypes (Fig. 3a). These results indicate a previously unappreciated neuronal subtype heterogeneity of key functional regulators in human brain evolution.

Coevolution of chromatin and RNA

We then investigated the overall association between chromatin accessibility changes and gene expression changes. We found that the association between human-specific gene expression and chromatin accessibility changes was the strongest at promoters and declined with the distance from the TSS (Extended Data Fig. 7a). This trend was observed only among the gains and losses that are concordant between the genome and the transcriptome (HS-Up-Gene and HS-Open-CRE/HS-Down-Gene and HS-Closed-CRE) but not in the discordant overlaps (for example, HS-Up-Gene and HS-Closed-CRE; Extended Data Fig. 7a). Overlaps for concordant, but not discordant, gains or losses were significant for nearly all subtypes (Extended Data Fig. 7b–d). Together these

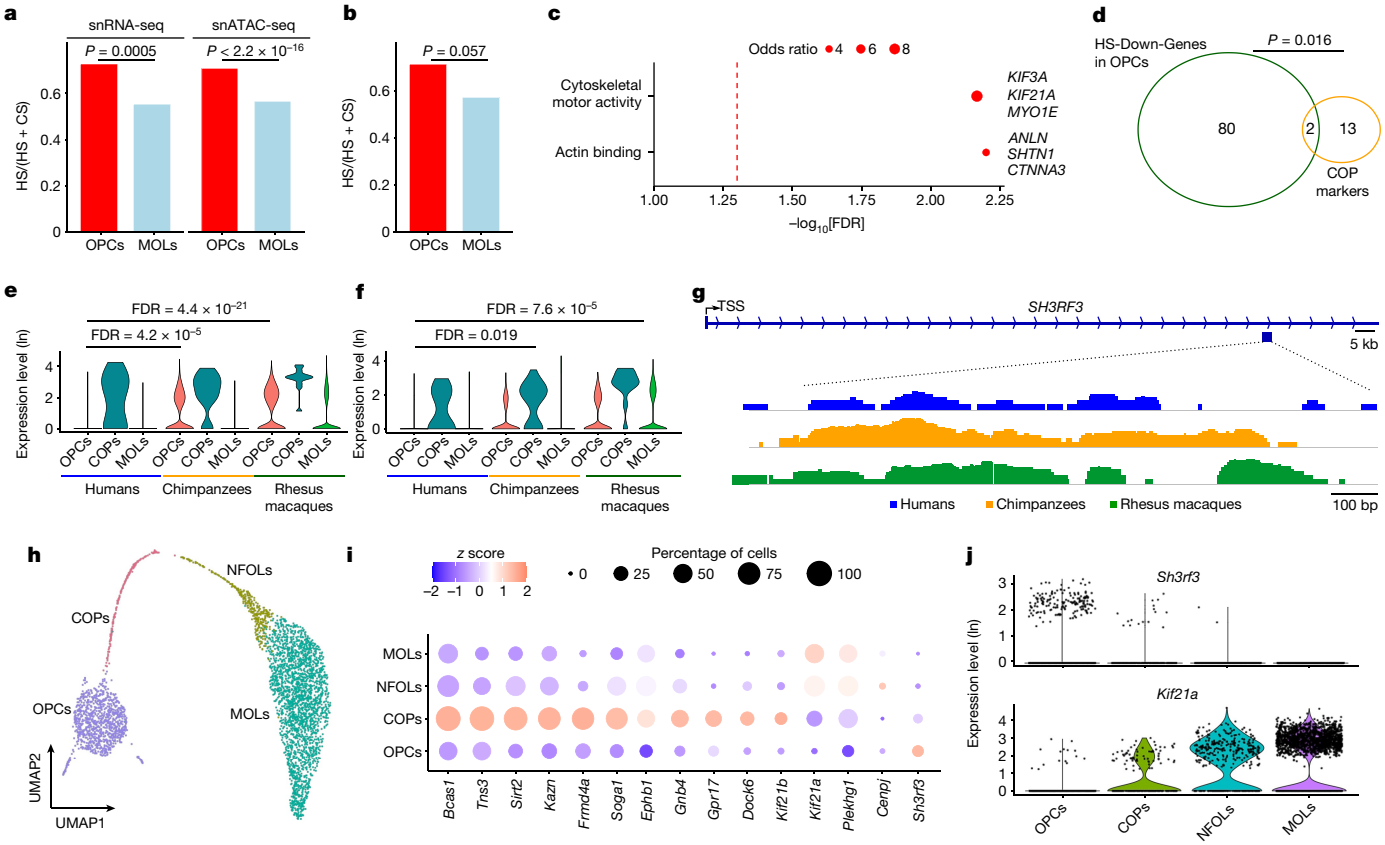


Fig. 2 | Human-specific gene regulatory changes in the oligodendrocyte lineage. a, The proportions of HS regulatory changes to HS + CS regulatory changes are higher in OPCs than in MOLs (left: snRNA-seq, right: snATAC-seq) for posterior cingulate cortex. *P* value: chi-square test, two-sided. **b**, The same as in **a**, except using the anterior cingulate cortex snRNA-seq dataset. **c**, GO enrichment for HS-Down-Genes in OPCs highlights altered cytoskeletal function (*P* value: Fisher's exact test, one-sided). **d**, Overlap of HS-Down-Genes in OPCs and primate-conserved COP markers reveals two COP markers with loss of function in human OPCs (*P* value: Fisher's exact test, one-sided). **e**,**f**, Expression levels of *SH3RF3* (**e**) and *KIF21A1* (**f**) across cell types in human, chimpanzee and rhesus macaque. False discovery rate (FDR)-corrected *P* values compare the expression levels in OPCs between species (Supplementary Table 3). **g**, snATAC-seq coverage plots of the Human-DOWN-CRE near *SH3RF3* in OPCs. Track scales are the same in all species. bp, base pairs. **h**, Uniform manifold approximation and projection plot of oligodendrocyte lineage cells in the mouse adult frontal cortex dataset. NFOLs, newly formed oligodendrocytes. **i**, Expression pattern of primate-conserved COP markers across mouse oligodendrocyte lineage cell types. Only *Sh3rf3* expression is decreased in COPs or newly formed oligodendrocytes compared to OPCs. **j**, Violin plots of *Sh3rf3* and *Kif21a* expression in mouse oligodendrocyte lineage cell types.

results show that HS-CREs are significantly associated with HS-Genes and that this association is stronger if the former is near the TSS and both are altered in the same direction.

To further refine associations between CREs and HS-Genes, we scanned the 500-kilobase (kb) vicinity of each HS-Gene for HS-CREs that are altered in the same direction in the same cell type. This analysis assigned at least one HS-CRE to 26% of HS-Genes across cell types (Supplementary Table 6). Focusing on the *FOXP2* gene and surrounding genomic regions, we identified four HS-Open-CREs in the L5-6_THEMIS_1 subtype. Two of these CREs are also close to another HS-Up-Gene (*MDFIC*; Fig. 3g and Supplementary Table 6). Among the other two, one resides within a *FOXP2* intron, whereas the other one is about 244 kb away from the nearest *FOXP2* TSS. To identify putative targets of human-specific *FOXP2* upregulation, we then retained HS-CREs that have a FOXP2 motif and are associated with an HS-Gene in the same subtype. This analysis yielded 47 genes for the L5-6_THEMIS_1 subtype and 14 genes for the L4-6_RORB_2 subtype (Extended Data Fig. 7e,f and Supplementary Table 6). We note that the *FOXP2* upregulation in L5-6_THEMIS_1 is greater than in L4-6_RORB_2 (log[fold change (FC)] = 0.8 and 0.4, respectively), and our analysis identified 3.35-fold more putative FOXP2 targets in L5-6_THEMIS_1 than in L4-6_RORB_2 (human-specific changes are only 1.8-fold more in L5-6_THEMIS_1 than L4-6_RORB_2). We further highlighted 7 genes that are not altered in the other 12 subtypes, similar to *FOXP2* itself (Extended Data Fig. 7e). Together, these results provide a list of potential epigenomic alterations associated with transcriptomic alterations in human brain evolution.

Human accelerated regions in brain

A goal of comparative genomic studies is to connect the changes at genomic sequences to functional changes. We therefore focused on human accelerated regions (HARs)[27], which are genomic regions that have significantly accelerated sequence evolution in the human lineage[28]. We found that 30% of published HARs overlapped the CREs in our dataset (about 2.5-fold excess compared to randomized background, *P* value < 0.05; Extended Data Fig. 8a), reaffirming the significance of these regions in human brain evolution[10,29,30]. Published HARs within CREs also showed modest but significant enrichment in HS-CREs in several cell types (Fig. 4a). However, these published HARs use sequence evolution without consideration of a specific tissue. Leveraging the snATAC-seq dataset, we reasoned that we could find many accelerated genomic regions by carrying out HAR analysis restricted to the CREs we identified (Methods). The odds ratio of published HAR and HS-CRE association is about 1.4, which was achieved in our analysis with an

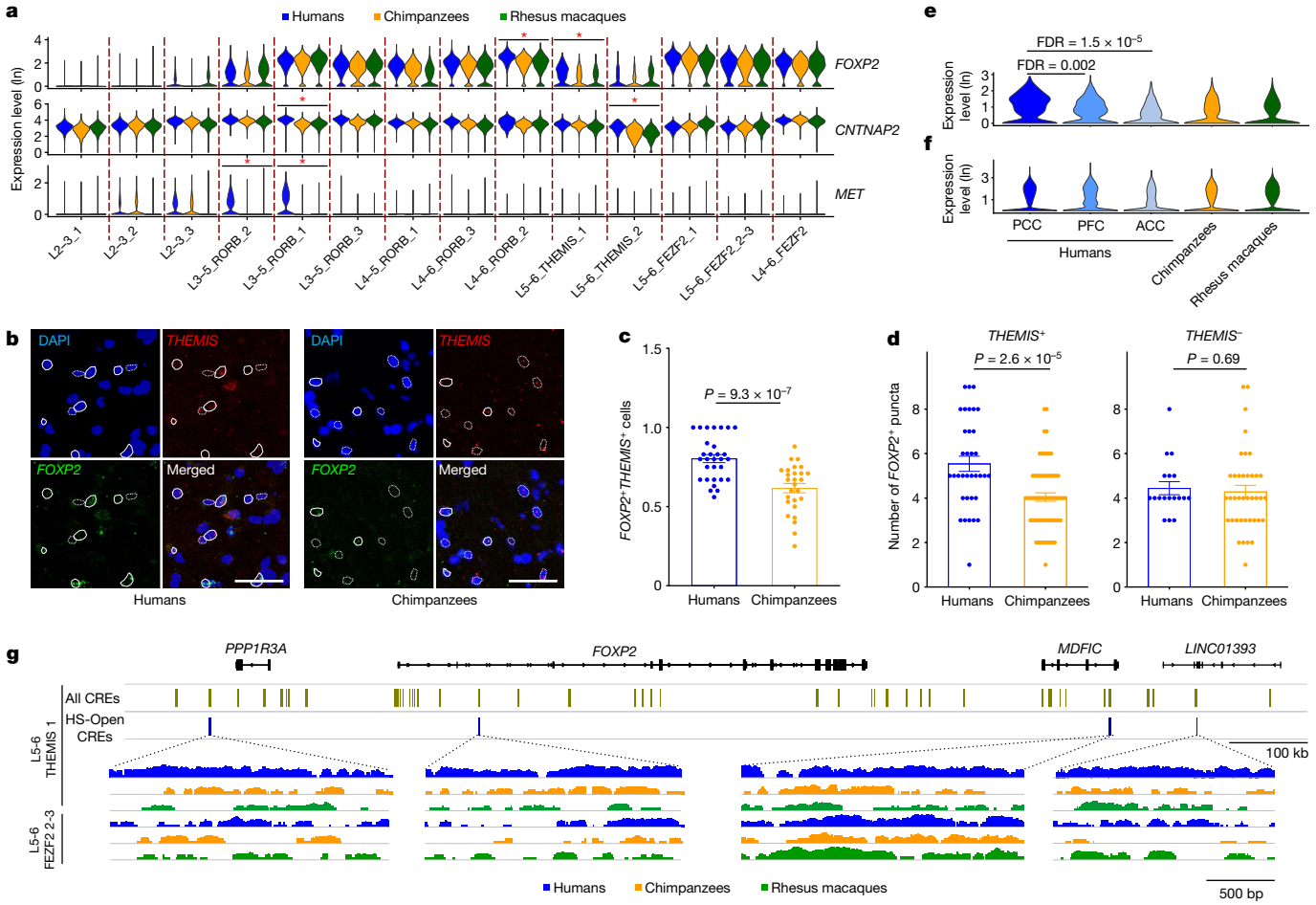


Fig. 3 | Subtype- and cortical region-specific upregulation of *FOXP2* in human neurons. a, Expression levels of *FOXP2, CNTNAP2 and MET* in the posterior cingulate cortex show subtype-specific expression changes in the human brain. Human-specific expression is labelled with a red asterisk; *x* axis denotes species and excitatory subtypes. **b**–**d**, smFISH of *FOXP2* and *THEMIS* in anterior cingulate cortex shows a greater number of *FOXP2*$^+$*THEMIS*$^+$ cells in humans compared to chimpanzees. **b**, Representative images. Solid outlines show cells in which *FOXP2* and *THEMIS* overlap; dashed outlines show *THEMIS*$^+$ cells without *FOXP2* expression. Scale bars, 50 µm. **c**,**d**, Quantification of *FOXP2*$^+$*THEMIS*$^+$ cells (**c**) and *FOXP2*$^+$ puncta per cell in *THEMIS*$^+$ cells (**d**, left) and *THEMIS*$^-$ cells (**d**, right). The *P* value is the main effect of species from a linear mixed model (random effect: individual, two-sided) with each data point representing a sub-area per image per individual. Error bars represent s.e.m. $n = 3$ individuals per species. **e**,**f**, *FOXP2* is upregulated in the posterior cingulate cortex (PCC) compared to prefrontal cortex (PFC) and anterior cingulate cortex (ACC) in *THEMIS*$^+$ neurons (**e**) but not among all excitatory neurons (**f**). *y* axes: normalized and log-transformed expression levels. **g**, snATAC-seq coverage plots of HS-Open-CREs near *FOXP2* in L5-6_THEMIS_1 and L5-6_FEZF2_2-3 neurons. The HS-Open-CREs shown have human-specific chromatin accessibility in L5-6_THEMIS_1 neurons but not in L5-6_FEZF2_2-3 neurons. Track scales are the same in all species.

unadjusted *P*-value cutoff of 0.001 (Fig. 4b and Supplementary Table 7). We note that, in contrast to previous genome-wide approaches, this focused approach to define HARs allows us to relax statistical criteria (unadjusted $P < 0.001$) without reducing the effect sizes observed in published HARs, while simultaneously enhancing validity by linking substitution changes to functional changes (that is, HS-CREs). We named these segments cortical HARs, as the cellular composition of cortical brain regions is similar and we found that CREs from other cortical regions show a high degree of overlap with our dataset (Extended Data Fig. 8b). Many published HARs are also cortical HARs (Extended Data Fig. 8c) and we identified >3-fold more HS-CREs overlapping a cortical HAR than overlapping a published HAR (Extended Data Fig. 8d). Cortical HARs were also significantly enriched in HS-CREs from most cell types (Fig. 4c), and we highlight some notable examples of HS-CRE-associated HARs that are important for synaptic (*CELF4*)[31] or oligodendrocyte (*NRG3*)[32] function (Extended Data Fig. 8e,f). Together, these results demonstrate a significant association between sequence divergence and chromatin accessibility in human evolution and provide hundreds of new HARs accompanying chromatin accessibility change at cell-type resolution in the human brain.

Chromatin evolution in modern humans

Comparison of the genomes of anatomically modern humans to those of archaic humans permits the identification of 'modern human-specific' variants with unknown functional consequences[33]. We thus investigated the associations between modern human-specific variants and chromatin changes in the brain. In total, we identified 12,161 modern human-specific variants associated with HS-CREs (Supplementary Table 8), which was a significant enrichment ($P = 0.007$; Methods). Among the cell types, we found a significant enrichment only in upper-layer excitatory neurons (Fig. 4d).

To compare the enrichments of modern human-specific variants to those that are specific to the entire human lineage (henceforth termed human specific), we first identified about 1.5 million human-specific substitutions within the CREs (Supplementary Table 9). Similar to the

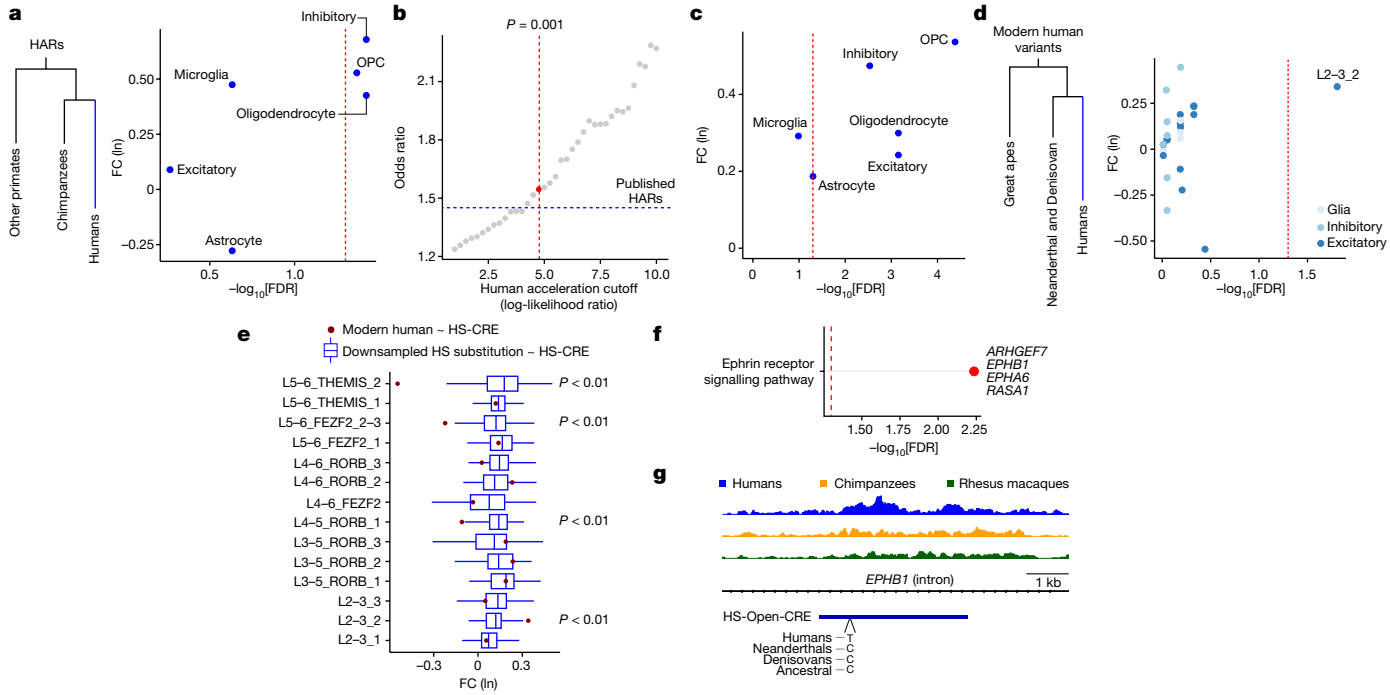


Fig. 4 | Significant association of chromatin accessibility changes with sequence divergence. a, Enrichment of publicly available HARs within the HS-CREs. Enrichment is tested by a logistic regression model with CRE length and evolution of the CRE as the predictor variables (HS-CRE or not HS-CRE) and HAR as the response variable (HAR or not HAR, *P* value: likelihood ratio test, two-sided). **b**, Odds ratio of HAR and HS-CRE association compared to HAR and NS-CRE (non-significant CREs; all CREs that are not HS-CREs) for published HARs (dashed blue line) and HARs identified in this study per log-likelihood cutoff (*x* axis). The red dashed line indicates the log-likelihood cutoff that corresponds to $P = 0.001$ (one-sided). The red dot indicates the odds ratio that corresponds to $P = 0.001$ cutoff. **c**, Cortical HAR analysis reveals stronger association with HS-CREs (same enrichment analysis as **a** between cortical HARs and HS-CREs). **d**, Modern human-specific variant enrichment within the HS-CREs. Enrichment is tested by a negative binomial regression model with CRE length and evolution of the CRE as the predictor variables (HS-CRE or not HS-CRE) and the number of modern human-specific variants as a response variable (*P* value: likelihood ratio test, two-sided). **e**, Log[FC] of substitution and HS-CRE association for substitutions on the human (blue) and modern human (red) lineage per excitatory subtype. Human lineage-specific substitutions were randomly downsampled 100 times to 12,161 (the number of modern human-specific variants) for comparison. Empirical *P* value (two-sided) is reported for conserved (L5-6_THEMIS_2, L5-6_FEXF2_2-3 and L4-5_RORB_1) and accelerated (L2-3_2) subtypes in modern humans. Boxplots represent median and interquartile range. **f**, GO enrichment of HS-Open-CREs with modern human-specific variants in the L2-3_2 subtype. **g**, snATAC-seq coverage plots of Human-Open-CREs near *EPHB1* in L2-3_2 neurons, which has a modern human variant. Track scales are the same in all species.

HARs, human-specific substitutions were significantly enriched in HS-CREs, and we noted the example of *GRIK4*, which encodes a glutamate receptor subunit implicated in brain disease[34] (Extended Data Fig. 8g,h). As expected, human-specific substitutions also encompassed about 88% of previously identified modern human-specific variants (Extended Data Fig. 8i). To reduce the confounding effects of sample sizes, we randomly downsampled human-specific substitutions to match the number of modern human-specific variants and calculated their association with HS-CREs per cell type. This analysis revealed greater associations between modern human-specific variants and upper-layer HS-CREs compared to the substitutions along the entire human lineage (Fig. 4e and Extended Data Fig. 8j). Gene ontology (GO) enrichment analysis of HS-CREs with modern human-specific variants revealed the ephrin receptor signalling pathway as the only ontological enrichment (Fig. 4f,g). These results indicate that modern human-specific variants are associated with human-specific CRE changes.

Activity-response elements in human CREs

TFs are key components in evolution and disease. We found enrichments of diverse TF-binding motifs in HS-Open-CREs across neuronal subtypes (Extended Data Fig. 9a,b and Supplementary Table 10). Notably, we observed significant enrichments for FOS::JUN motifs in the upper-layer excitatory neurons and for FOX motifs in the lower-layer excitatory neurons (Fig. 5a,b). We further identified TFs that may be functional at these HS-CRE target sites by examining the accessibility of each enriched TF within each family (Fig. 5a,b and Extended Data Fig. 9c,d).

FOS::JUN TFs are immediately transcribed following neuronal depolarization and target hundreds of CREs[35,36]. As FOS::JUN TFs respond to environmental stimuli, we tested whether FOS::JUN TF enrichment in HS-Open-CREs is driven by environmental factors. We first investigated whether greater postmortem interval (PMI) in human tissues compared to chimpanzee and rhesus macaque tissues, a limitation in many similar studies[12,14,18], is driving this enrichment. To test this, we substituted our human snATAC-seq dataset (named PMI_24) with a surgical human dataset from middle temporal gyrus that has no PMI (PMI_0)[37]. We similarly found all excitatory subtypes in this dataset and identified HS-CREs that exhibited highly significant overlaps with the PMI_24 HS-CREs (Extended Data Fig. 9e–g), as well as enrichments of similar motifs (Extended Data Fig. 9h). Similar to the PMI_24 dataset, HS-Open-CREs in the upper-layer excitatory neurons were highly enriched in FOS::JUN motifs (Extended Data Fig. 9i). These results show that FOS::JUN enrichments in upper-layer excitatory HS-Open-CRE are not driven by PMI differences.

To provide an orthogonal test for a possible environmental effect on FOS::JUN motif enrichments, we investigated whether HS-Open-CREs with FOS::JUN motifs also contain signatures of accelerated evolution. If FOS::JUN motif enrichments in HS-Open-CREs are driven by

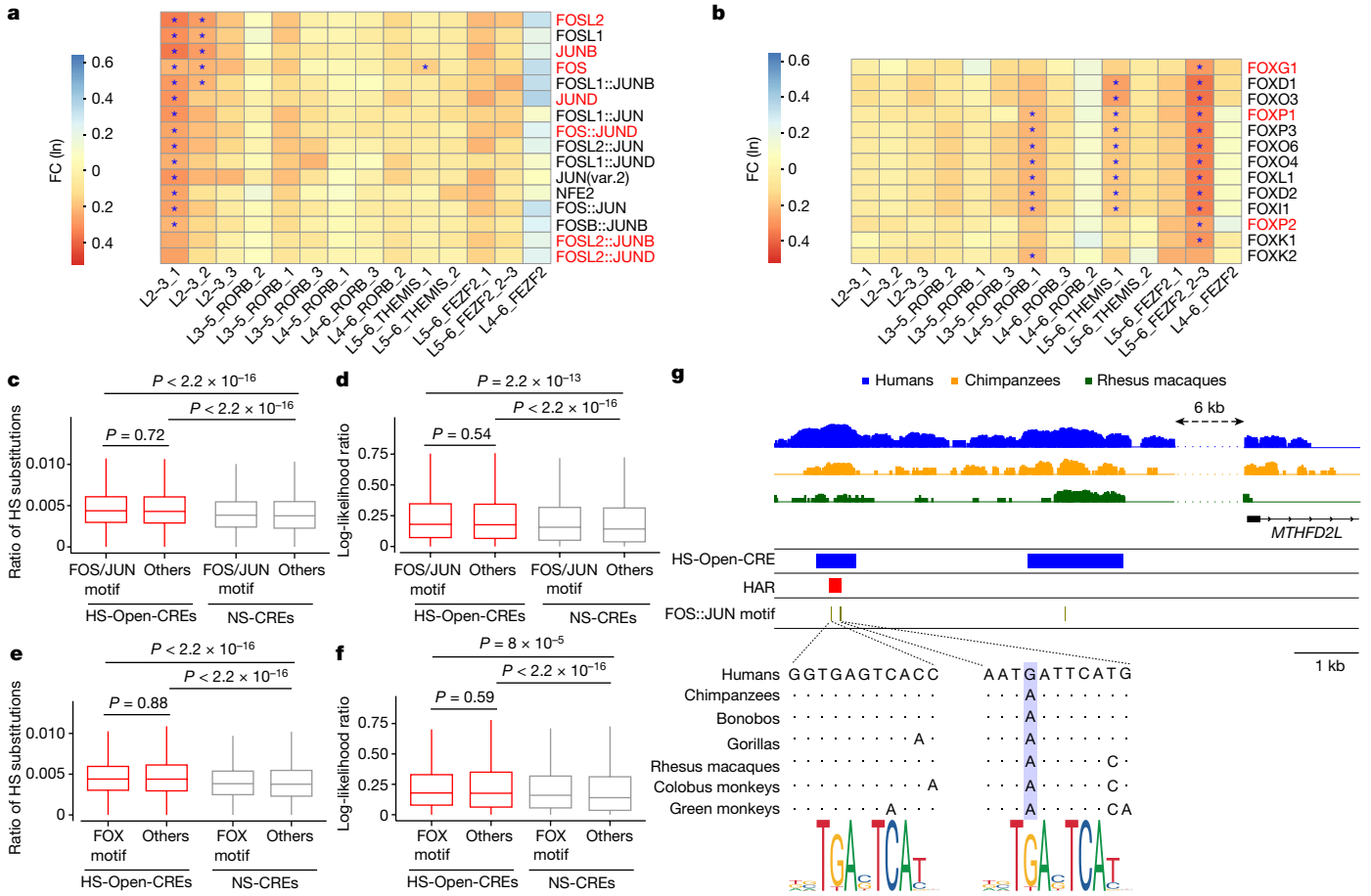


Fig. 5 | Accessibility changes highlight subtype-specific TF target evolution in the human brain. a,**b**, FOS::JUN (**a**) and FOX (**b**) family TF enrichments in HS-Open-CREs. The heatmaps show the log[FC] of TF motif enrichment within HS-Open-CREs per TF motif and per subtype (blue asterisks indicate FDR < 0.05; red font highlights distinctly more accessible TFs within these subtypes). **c**, Ratio of HS substitutions by CRE length per group of CREs. Groups from left to right: HS-Open-CREs that contain at least one motif, HS-Open-CREs that do not contain a motif, NS-CREs (non-significant CREs) that contain a motif, NS-CREs that do not contain a motif. Only the enriched subtypes per TF group and the highlighted TFs in **a** per TF family were used for these comparisons (n = 1,519, 2,894, 38,486 and 109,452 CREs left to right). Boxplots represent median and interquartile range. P value: Wilcoxon rank sum test (two-sided). **d**, The same CREs as in **c** except using the mean log-likelihood ratios per CREs computed in the HAR analysis. **e**,**f**, The same comparison as in **c**,**d** except using FOX motifs (n = 1,338, 3,075, 38,774 and 109,164 CREs left to right). **g**, Track plot of an HS-Open-CRE with a HAR and a human-specific gain of a FOS::JUN motif. Identical sequences with respect to the human sequence are shown with dots. The motif representation on the bottom is a FOS motif.

environmental factors, human-specific substitutions within the HS-Open-CREs with FOS::JUN motif occurrences should be depleted compared to those within other HS-Open-CREs. Contrary to this expectation, we found a significant excess of HS substitutions and accelerated evolution when HS-Open-CREs with FOS::JUN motifs were compared to the nonspecific (NS) CREs (Fig. 5c,d). FOX targets were also more divergent in humans compared to NS-CREs with or without FOX motifs (Fig. 5e,f). An example of a human-specific gain of a FOS::JUN motif within a HAR is shown near *MTHFD2L* (Fig. 5g), which encodes a key enzyme in the one-carbon metabolism associated with neurotransmitter synthesis[38]. Taken together, these results do not support a possible environmental cause.

In summary, we have uncovered proportional and gene regulatory changes in human brain evolution using single-cell genomics and have linked human-specific DNA sequence divergence, chromatin accessibility and gene expression at cellular resolution.

Discussion

In this study, we delineated epigenomic and transcriptomic features of human brain evolution at cell-type resolution. We found that the adult human cortex had an increased proportion of OPCs and a decreased proportion of MOLs compared to non-human primates. Focusing on neurons, we showed that many human-specific changes were found in only a few neuronal subtypes, and demonstrated human-specific upregulation of *FOXP2* in two neuronal subtypes. We also associated genomic sequence changes with HS-CREs at cellular resolution and identified hundreds of new HARs that were associated with open chromatin in the adult brain. Furthermore, we identified increased FOS::JUN TF targets among the HS-Open-CREs in the upper-layer excitatory neurons, emphasizing a previously unappreciated temporal dimension of human-specific molecular traits.

Previous studies showed prolonged myelination in brain development in humans compared to chimpanzees and rhesus macaques[5,8]. Correspondingly, the production of myelinating oligodendrocytes reaches a plateau in individuals older than about 40 years old in grey matter[39]. Notably, we observed proportionally higher OPCs in humans compared to chimpanzees and rhesus macaques even though individuals in our dataset are all in their mid to late adulthood (humanized age; Supplementary Table 1). We also found that COPs, cells that denote active oligodendrocyte generation[40], are extremely rare (only 74 nuclei in all species), indicating low levels of oligodendrocyte generation in all species in our samples. We reason that the higher proportion of OPCs and lower proportion of MOLs can contribute to neural plasticity

in the human brain by altering myelination patterns. Non-canonical functions of OPCs such as pruning axonal branches and contributing to synaptic function have been described recently[41], indicating that increased numbers of OPCs in the human brain may serve functions other than providing a reservoir for MOLs. We note that a recent study found more divergence between species for MOLs compared to OPCs by comparing the gene expression correlations[42]. The discrepancy with our results could be due to differences in the brain regions analysed, sorting strategy (NeuN-sorted versus not sorted) or analytical pipeline (for example, correlations versus differential gene expression (DGE)). We also note that our approach separates human-specific changes from chimpanzee-specific changes as a measure of human specificity, making it better tailored to highlight the changes in human lineage.

Single-cell sequencing facilitates characterization of regulatory changes in all cell types. However, we recently discovered that neuronal ambient RNAs contaminate glial cell types and require rigorous removal before identification of differentially expressed genes[21], as such contamination can skew the DGE results[43]. We found that a recent study[14] shows evidence of human-specific differences in the level of ambient RNA contamination in glial cell types, indicating the importance of ambient contamination removal (Extended Data Fig. 10a,b). Among the studies with human–chimpanzee comparisons so far[11,14], to our knowledge, only our study has removed ambient RNA contamination. Following removal of ambient RNAs, we uncovered that cytoskeletal activity and ubiquitin ligase activity through *SH3RF3* are specifically decreased in human OPCs (Fig. 2c–g). Both biological processes are linked to oligodendrocyte maturation, indicating that such functions might be linked to the human-specific OPC increase[20,24]. These results also suggest that an evolutionary modification in human brain may have been achieved through a loss of function in OPCs[44].

We found a subtype- and human-specific upregulation of *FOXP2* that may be unique to the posterior cingulate cortex. We also note that most FOX TF motifs are enriched in the HS-Open-CREs in *THEMIS*$^{+}$*C1QL3*$^{+}$ neurons (Fig. 5b), and although the FOXP2 motif enrichment itself was not significant, this could be ascribed to possible variations of FOXP2-binding sites in different tissues. Indeed, we previously showed that FOXP2 can act both as a repressor or activator through heterodimerization with other TFs at distinct DNA motifs[45]. In addition, a recent study identified human-specific *FOXP2* upregulation in microglia[14], with a trend similar to that in our dataset (Supplementary Table 3), suggesting a previously undescribed potential role of FOXP2. These results provide further insights into the role of FOXP2 in human brain evolution.

Notably, association between human-specific gene expression changes and chromatin accessibility changes was significant only between the concordant changes but not between discordant changes (Extended Data Fig. 7). Although human-specific upregulated repressors may exist, the activity of such repressors is probably manifested as closed accessibility of CREs via other epigenetic mechanisms[46]. Consequenctly, we may still observe a concordant change.

We discovered a new enrichment of FOS and JUN family motifs in specifically cortical upper-layer excitatory HS-Open-CREs. Late activity-regulated genes are known to be evolutionarily divergent[38,47–49] and exhibit high cell-type specificity[50]. Along with the previous studies, our results underscore the need for more direct experiments to understand how adult human cortical cells respond to neuronal activity, and the underlying evolutionary trajectories. We also note that some of our analyses are limited to BA23 and future comparative studies from other brain regions are needed. Overall, our results provide a comprehensive roadmap for delineating functional regulatory mechanisms of human brain evolution at cellular resolution.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-023-06338-4.

1. King, M. C. & Wilson, A. C. Evolution at two levels in humans and chimpanzees. *Science* **188**, 107–116 (1975).
2. Konopka, G. et al. Human-specific transcriptional networks in the brain. *Neuron* **75**, 601–617 (2012).
3. Liu, X. et al. Extension of cortical synaptic development distinguishes humans from chimpanzees and macaques. *Genome Res.* **22**, 611–622 (2012).
4. Sousa, A. M. M. et al. Molecular and cellular reorganization of neural circuits in the human lineage. *Science* **358**, 1027–1032 (2017).
5. Zhu, Y. et al. Spatiotemporal transcriptomic divergence across human and macaque brain development. *Science* https://doi.org/10.1126/science.aat8077 (2018).
6. Hodge, R. D. et al. Conserved cell types with divergent features in human versus mouse cortex. *Nature* **573**, 61–68 (2019).
7. Bakken, T. E. et al. Comparative cellular analysis of motor cortex in human, marmoset and mouse. *Nature* **598**, 111–119 (2021).
8. Miller, D. J. et al. Prolonged myelination in human neocortical evolution. *Proc. Natl Acad. Sci. USA* **109**, 16480–16485 (2012).
9. Jakel, S. et al. Altered human oligodendrocyte heterogeneity in multiple sclerosis. *Nature* **566**, 543–547 (2019).
10. Jeong, H. et al. Evolution of DNA methylation in the human brain. *Nat. Commun.* **12**, 2021 (2021).
11. Khrameeva, E. et al. Single-cell-resolution transcriptome map of human, chimpanzee, bonobo, and macaque brains. *Genome Res.* **30**, 776–789 (2020).
12. Kozlenkov, A. et al. Evolution of regulatory signatures in primate cortical neurons at cell-type resolution. *Proc. Natl Acad. Sci. USA* **117**, 28422–28432 (2020).
13. Krienen, F. M. et al. Innovations present in the primate interneuron repertoire. *Nature* **586**, 262–269 (2020).
14. Ma, S. et al. Molecular and cellular evolution of the primate dorsolateral prefrontal cortex. *Science* https://doi.org/10.1126/science.abo7257 (2022).
15. Mendizabal, I. et al. Comparative methylome analyses identify epigenetic regulatory loci of human brain evolution. *Mol. Biol. Evol.* **33**, 2947–2959 (2016).
16. Li, W., Mai, X. & Liu, C. The default mode network and social understanding of others: what do brain connectivity studies tell us. *Front. Hum. Neurosci.* **8**, 74 (2014).
17. Wang, D. et al. Altered functional connectivity of the cingulate subregions in schizophrenia. *Transl. Psychiatry* **5**, e575 (2015).
18. Berto, S. et al. Accelerated evolution of oligodendrocytes in the human brain. *Proc. Natl Acad. Sci. USA* **116**, 24334–24342 (2019).
19. Franjic, D. et al. Transcriptomic taxonomy and neurogenic trajectories of adult human, macaque, and pig hippocampal and entorhinal cells. *Neuron* **110**, 452–469 (2022).
20. Brown, T. L. & Verden, D. R. Cytoskeletal regulation of oligodendrocyte differentiation and myelination. *J. Neurosci.* **37**, 7797–7799 (2017).
21. Caglayan, E., Liu, Y. & Konopka, G. Neuronal ambient RNA contamination causes misinterpreted and masked cell types in brain single-nuclei datasets. *Neuron* https://doi.org/10.1016/j.neuron.2022.09.010 (2022).
22. Lake, B. B. et al. Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat. Biotechnol.* **36**, 70–80 (2018).
23. Velmeshev, D. et al. Single-cell genomics identifies cell type-specific molecular changes in autism. *Science* **364**, 685–689 (2019).
24. Fumagalli, M. et al. The ubiquitin ligase Mdm2 controls oligodendrocyte maturation by intertwining mTOR with G protein-coupled receptor kinase 2 in the regulation of GPR17 receptor desensitization. *Glia* **63**, 2327–2339 (2015).
25. den Hoed, J., Devaraju, K. & Fisher, S. E. Molecular networks of the FOXP2 transcription factor in the brain. *EMBO Rep.* **22**, e52803 (2021).
26. Konopka, G. et al. Human-specific transcriptional regulation of CNS development genes by FOXP2. *Nature* **462**, 213–217 (2009).
27. Doan, R. N. et al. Mutations in human accelerated regions disrupt cognition and social behavior. *Cell* **167**, 341–354 (2016).
28. Franchini, L. F. & Pollard, K. S. Human evolution: the non-coding revolution. *BMC Biol.* **15**, 89 (2017).
29. Capra, J. A., Erwin, G. D., McKinsey, G., Rubenstein, J. L. & Pollard, K. S. Many human accelerated regions are developmental enhancers. *Philos. Trans. R. Soc. Lond. B* **368**, 20130025 (2013).
30. Girskis, K. M. et al. Rewiring of human neurodevelopmental gene regulatory programs by human accelerated regions. *Neuron* https://doi.org/10.1016/j.neuron.2021.08.005 (2021).
31. Wagnon, J. L. et al. CELF4 regulates translation and local abundance of a vast set of mRNAs, including genes associated with regulation of synaptic function. *PLoS Genet.* **8**, e1003067 (2012).
32. Lundgaard, I. et al. Neuregulin and BDNF induce a switch to NMDA receptor-dependent myelination by oligodendrocytes. *PLoS Biol.* **11**, e1001743 (2013).
33. Prufer, K. et al. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43–49 (2014).
34. Arora, V. et al. Increased Grik4 gene dosage causes imbalanced circuit output and human disease-related behaviors. *Cell Rep.* **23**, 3827–3838 (2018).
35. Kim, T. K. et al. Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182–187 (2010).
36. Yap, E. L. & Greenberg, M. E. Activity-regulated transcription: bridging the gap between neural activity and behavior. *Neuron* **100**, 330–348 (2018).
37. Berto, S. et al. Gene-expression correlates of the oscillatory signatures supporting human episodic memory encoding. *Nat. Neurosci.* **24**, 554–564 (2021).
38. Ducker, G. S. & Rabinowitz, J. D. One-carbon metabolism in health and disease. *Cell Metab.* **25**, 27–42 (2017).

39. Yeung, M. S. et al. Dynamics of oligodendrocyte generation and myelination in the human brain. *Cell* **159**, 766–774 (2014).
40. Marques, S. et al. Oligodendrocyte heterogeneity in the mouse juvenile and adult central nervous system. *Science* **352**, 1326–1329 (2016).
41. Buchanan, J. et al. Oligodendrocyte precursor cells ingest axons in the mouse neocortex. *Proc. Natl Acad. Sci. USA* **119**, e2202580119 (2022).
42. Jorstad, N. L. et al. Comparative transcriptomics reveals human-specific cortical features. Preprint at *bioRxiv* https://doi.org/10.1101/2022.09.19.508480 (2022).
43. Berg, M. et al. FastCAR: Fast Correction for Ambient RNA to facilitate differential gene expression analysis in single-cell RNA-sequencing datasets. Preprint at *bioRxiv* https://doi.org/10.1101/2022.07.19.500594 (2022).
44. McLean, C. Y. et al. Human-specific loss of regulatory DNA and the evolution of human-specific traits. *Nature* **471**, 216–219 (2011).
45. Hickey, S. L., Berto, S. & Konopka, G. Chromatin decondensation by FOXP2 promotes human neuron maturation and expression of neurodevelopmental disease genes. *Cell Rep.* **27**, 1699–1711 (2019).
46. Yang, C. C. et al. Discovering chromatin motifs using FAIRE sequencing and the human diploid genome. *BMC Genomics* **14**, 310 (2013).
47. Ataman, B. et al. Evolution of osteocrin as an activity-regulated factor in the primate brain. *Nature* **539**, 242–247 (2016).
48. Pruunsild, P., Bengtson, C. P. & Bading, H. Networks of cultured iPSC-derived neurons reveal the human synaptic activity-regulated adaptive gene program. *Cell Rep.* **18**, 122–135 (2017).
49. Qiu, J. et al. Evidence for evolutionary divergence of activity-dependent gene expression in developing neurons. *Elife* https://doi.org/10.7554/eLife.20337 (2016).
50. Hrvatin, S. et al. Single-cell analysis of experience-dependent transcriptomic states in the mouse visual cortex. *Nat. Neurosci.* **21**, 120–129 (2018).

Methods

Specific details of all analyses can be found https://github.com/konopkalab/Comparative_snATAC_snRNA.

Sampling strategy for snRNA-seq and snATAC-seq

All human tissue was obtained from the University of Texas Neuropsychiatry Research Program (Dallas Brain Collection). Chimpanzee and macaque tissues were obtained from Yerkes National Primate Research Center. BA23 (part of the posterior cingulate cortex) was dissected from frozen postmortem tissue slabs. Humanized age (calculated as described before[18]) and sex were matched between species to minimize the effect of demographics. In total, four individuals were sequenced from each species (Extended Data Fig. 1a and Supplementary Table 1).

snRNA-seq library preparation

Nuclei for snRNA-seq were isolated from human, chimpanzee and macaque BA23 brain tissue. Briefly, the tissue was homogenized using a glass Dounce homogenizer in 2 ml of ice-cold lysis buffer (10 mM Tris-HCl, 10 mM NaCl, 3 mM $MgCl_2$ and 0.1% Nonidet P40 Substitute) and was incubated on ice for 5 min. Nuclei were centrifuged at 500*g* for 5 min at 4 °C, washed with 4 ml ice-cold lysis buffer and incubated on ice for 5 min. Nuclei were centrifuged at 500*g* for 5 min at 4 °C. After centrifugation, the nuclei were resuspended in 500 µl of nucleus suspension buffer (NSB) containing 1× PBS, 1% BSA (no. AM2618, Thermo Fisher Scientific) and 0.2 U μl^{-1} RNAse inhibitor (no. AM2694, Thermo Fisher Scientific). The nucleus suspension was filtered through a 70-µm Flowmi cell strainer (no. H13680-0070, Bel-Art). Debris was removed with density gradient centrifugation using Nuclei PURE 2 M sucrose cushion solution and Nuclei PURE sucrose cushion buffer from the Nuclei PURE prep isolation kit (no. NUC201-1KT, Sigma Aldrich). Nuclei PURE 2 M sucrose cushion solution and Nuclei PURE sucrose cushion buffer were first mixed in a 9:1 ratio. A 500 µl volume of the resulting sucrose solution was added to a 2-ml Eppendorf tube. A 900 µl volume of the sucrose buffer was added to 500 µl of isolated nuclei in NSB. A 1,400 µl volume of nucleus suspension was layered to the top of the sucrose buffer. This gradient was centrifuged at 13,000*g* for 45 min at 4 °C. The pellet of nuclei was resuspended, washed once in NSB and filtered through a 70-µm Flowmi cell strainer (no. H13680-0070, Bel-Art). The concentration of nuclei was determined using 0.4% trypan blue (no. 15250061, Thermo Fisher Scientific), and was adjusted to a final concentration of 1,000 nuclei per microlitre with NSB.

Droplet-based snRNA-seq libraries were prepared using Chromium Single Cell 3′ v3.1 (1000121, 10x Genomics) according to the manufacturer's protocol[51]. Libraries were sequenced using an Illumina NovaSeq 6000.

snATAC-seq library preparation

For snATAC-seq, nuclei were isolated from human, chimpanzee and macaque BA23 tissue as previously described (https://www.protocols.io/view/isolation-of-nuclei-from-frozen-tissue-for-atac-se-6t8herw). Briefly, tissue pieces neighbouring the tissue used for snRNA-seq were cut and homogenized using a glass Dounce homogenizer in ATAC-seq homogenization buffer (0.25 M sucrose, 25 mM KCl, 5 mM $MgCl_2$, 20 mM tricine-KOH (pH 7.8), 1 mM dithiothreitol, 0.5 mM spermidine, 0.15 mM spermine, 0.3% NP40, protease inhibitors). The nuclei were filtered through a 70-µm Flowmi cell strainer (no. H13680-0070, Bel-Art) and were pelleted by centrifugation for 5 min at 4 °C at 350*g* in a 2-ml Eppendorf tube. The supernatant was discarded, and the nuclei were resuspended in 400 µl of homogenization buffer. A 400 µl volume of 50% iodixanol solution was added to the nucleus suspension and was mixed by pipetting. A 600 µl volume of 30% iodixanol solution was layered under the 25% mixture. A 600 µl volume of 40% iodixanol solution was then layered under the 30% mixture. This gradient was then centrifuged for 20 min at 4 °C at 3,000*g*. After centrifugation, the nuclei were recovered at the 30%–40% interface. The nuclei were transferred to a new Eppendorf tube and resuspend in 200 µl ATAC-RSC-Tween buffer (10 mM Tris-HCl pH 7.5, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% Tween-20). The concentration of nuclei was determined using 0.4% trypan blue (no. 15250061, Thermo Fisher Scientific). Nucleus integrity was tested by staining with ethidium homodimer 1 (catalogue no. E1169, Invitrogen). Droplet-based snATAC-seq libraries were prepared using the Chromium Single Cell ATAC Library kit (1000110, 10x Genomics) according to the manufacturer's protocols. Libraries were sequenced using an Illumina NovaSeq 6000.

snRNA-seq preprocessing and annotation

Bcl files were converted to fastq using cellranger mkfastq. Barcode correction and reference genome alignment were carried out using cellranger count with default parameters (software: 10x Genomics Cell Ranger 3.1.0). For the alignment, the genome builds GRCh38, panTro5 (Pan_tro 3.0) and rheMac10 (Mmul_10) were used as reference genomes for humans, chimpanzees and macaques, respectively. The BAM output from cellranger count was further processed to keep only uniquely mapped reads using samtools (-q 255)[52]. As chimpanzee and macaque gene annotation files (gtf) are less accurate than those for humans, chimpanzee and macaque reads were then mapped to human coordinates using CrossMap[53]. featureCount was used to count reads mapping to gene bodies[54], and umi_tools[55] was used to create the count matrix (gene by cell barcode; per sample, the top 50,000 cell barcodes with the highest UMI counts were pre-filtered for faster computation).

To remove ambient RNA contamination, we used CellBender on the unnormalized count matrix per sample[56]. We note that without ambient RNA removal, glial cells were shown to be conspicuously contaminated with neuronal ambient RNAs[21].

Empty-droplet-filtered output from CellBender was further processed to retain only the protein-coding and orthologous genes (between *Homo sapiens*, *P. troglodytes* and *M. mulatta*), similar to the approach of ref. 18. An orthologous gene list was obtained from Ensembl version 103 (ref. 57). For quality control, we kept only nuclei with >200 UMIs and percentage of reads mapping to mitochondria of <5. We then clustered nuclei for further analysis. The following methods from Seurat v3 (ref. 58) were used to carry out and visualize clustering (a similar approach was followed for each new clustering procedure; details are available in the publicly available code): normalization (SCTransform), dimensionality reduction (RunPCA), batch correction (RunHarmony, default parameters), *k*-nearest neighbours (FindNeighbors) on batch-corrected dimensions and cluster identification by shared nearest neighbours (FindClusters). Uniform manifold approximation and projection embedding was then computed for visualization in two-dimensional space (RunUMAP). We removed clusters with an unusually high number of detected genes accompanied with a high level of expression of at least two typically distinct marker genes as potential nucleus doublets. We re-clustered the nuclei and repeated this process if needed until no such clusters were found. We then used canonical marker genes (for example, *GAD1* for inhibitory neurons) and a reference dataset[6] (using label transfer; see next paragraph) to broadly annotate nuclei in each species. Major cell types were defined as: excitatory neurons, inhibitory neurons, astrocytes, oligodendrocytes, OPCs and microglia. After broad annotation, we extracted each broad category (for example, excitatory neurons) from all species and integrated them across species using the default approach in Seurat v3 across all samples (SelectIntegrationFeatures, PrepSCTIntegration, FindIntegrationAnchors). We then clustered the nuclei on the integrated matrix per cell type and further removed potential doublets with the same criteria as above. We additionally removed clusters with high enrichment in previously identified ambient RNA markers as previously described[21].

To annotate neuronal subtypes, we used a previous study[6] as a reference to annotate our clusters through label transfer

(FindTransferAnchors, TransferData). We assigned each cluster an annotation label based on the layer and marker gene of predominant predicted annotations per cluster. These annotations were also verified with known marker genes that separate certain neuronal subtypes (Extended Data Fig. 3).

We note that endothelial cells were removed from the analysis as we did not detect a distinct cluster of endothelial cells in snATAC-seq.

snATAC-seq preprocessing and annotation

Bcl files were converted to fastq using cellranger mkfastq. Barcode correction and reference genome alignment were carried out using cellranger-atac count with default parameters (Software: 10x Genomics Cell Ranger ATAC 1.1.0). For the alignment, the genome builds GRCh38, panTro5 (Pan_tro 3.0) and rheMac10 (Mmul_10) were used as reference genomes for humans, chimpanzees and macaques, respectively. The BAM output from cellranger count was further processed to keep only uniquely mapped and properly paired reads using samtools[52]. Read duplicates were removed using MarkDuplicates from Picard tools[59]. Peak calling was carried out using macs2 (ref. 60) with the following parameters: --nomodel, --keep-dup all, extsize 200, --shift −100 to enrich for the cut sites. To obtain peaks concordant across samples, peak calling was carried out by pooling all samples from each species, as well as from each sample. Peaks from pooled samples were kept for further analysis only if they overlap >50% with peaks per sample in 3/4 of samples. This yielded a list of consensus peaks for each species.

To obtain a final set of peaks from consensus peaks, chimpanzee and macaque peaks were converted to human coordinates using liftOver[61]. All peaks were then merged using bedtools[62], resulting in merged peaks with a minimum distance of 200 between them (-d 200). To keep peaks with a reliable level of conservation across all species, merged peaks were reciprocally mapped to chimpanzee or macaque genomes and any peaks with more than twofold change in size, multi-mapped or less than 50% conserved (-minMatch = 0.5) were discarded in each liftOver operation. Merged peaks were then filtered for the peaks that reciprocally mapped to both chimpanzee and macaque by requiring >50% overlap between a peak in the merged peak set and reciprocally mapped peak set (bedtools intersect -f 0.5 -F 0.5). Despite being conservative, this approach kept >93% of the initial peaks, indicating that sequence identity of most open chromatin peaks is reliably conserved across species and allows direct comparisons between species in downstream analysis (for example, differential accessibility). After this stage, peaks were also referred to as CREs.

To obtain the peak–cell count matrix, reads were counted in each species' own coordinates using custom functions on bed files. To keep high-quality cells, only the barcodes with >3,000 reads in peaks, <100,000 reads in peaks and >15% fraction of reads in peaks were kept for further analysis. Barcode multiplets[63] were additionally removed using cellranger's, clean_barcode_multiplets_1.1.py tool. Resulting matrices were processed separately for each species. The following methods from Seurat v3 (ref. 58) were used to carry out and visualize each clustering procedure (details can be found in the publicly available code). Dimensionality reduction was carried out with latent semantic indexing (using the functions RunTFIDF and RunSVD in Signac[64]). Batch correction was achieved with harmony on latent semantic indexing dimensions (RunHarmony[65]). Batch-corrected dimensions were then used to compute *k*-nearest neighbours (FindNeighbors) and identify clusters by shared nearest neighbours (FindClusters). Uniform manifold approximation and projection embedding was computed for visualization in two-dimensional space (RunUMAP).

To annotate snATAC-seq cells, correspondence between gene accessibility and gene expression is required. To achieve this, a gene activity matrix was calculated using Cicero[66] for each species. Only CREs with more than 1% accessibility were retained for analysis, and CREs in protein-coding genes (gene body + 3 kb upstream) were used to annotate CREs to genes (annotate_cds_by_site), and these data were further processed to build the unnormalized gene activity matrix (build_gene_activity_matrix). Both major cell types (for example, excitatory) and subtypes (for example, L2-3_1) in snATAC-seq were annotated through label transfer with the corresponding snRNA-seq dataset as reference. All snRNA-seq to snATAC-seq label transfers were carried out separately for each species. Clusters with mixed annotation accompanied with unusually high number of reads in peaks and mixed marker gene activity (typically distinct marker genes highly accessible in the same cluster) were removed as potential doublets. An annotation label was assigned per cluster depending on the dominant annotation for each cluster. All cell types found in snRNA-seq were distinctly found in snATAC-seq and thus annotated with the same names.

Cell-type-fraction comparisons

For comparison of cell-type ratios, we calculated the fraction of glial cell types within all glia, the fraction of excitatory subtypes within all excitatory cells and the fraction of inhibitory subtypes within all inhibitor cells for each individual in both snRNA-seq and snATAC-seq. To determine whether the fraction differences were significant between species, we calculated the *P* value using a log-likelihood ratio on two nested models:

H0: fraction ~ assay (snRNA−seq or snATAC−seq)

H1: fraction ~ assay (snRNA−seq or snATAC−seq) + species,

(for example, human and chimpanzee)

This was carried out for each pairwise species comparison per cell type. The statistics are available in Supplementary Table 2.

smFISH

See Supplementary Table 1 for sample demographics. Cortical BA23 (posterior cingulate cortex) and anterior cingulate cortex samples from all species were postmortem, flash-frozen tissues that were embedded in OCT (optimal cutting temperature) compound. The tissue was sectioned at −20 °C to 20 µm on Superfrost Plus Microscope slides. smFISH was carried out using RNAScope Multiplex v2 Fluorescent assays. Protease was applied for 30 min, and all subsequent steps, including probe application, tyramide signal amplification, channel development and fluorophore application, were carried out according to the manufacturer's instructions for fresh frozen tissue except with the addition of Sudan Black B. Sudan Black B (0.05%) was added to the tissue after application of DAPI to quench autofluorescence. Probes for *MOG* (human: 543181-C2, chimpanzee: 1076431-C2; Advanced Cell Diagnostics), *PDGFRA* (human: 604488, chimpanzee: 1120031; Advanced Cell Diagnostics), *THEMIS* (human: 407261; Advanced Cell Diagnostics) and *FOXP2* (human: 551661-C2; Advanced Cell Diagnostics) were incubated with the tissue and hybridized with their target genes. Opal fluorophores 570 (NC1601878, Akoya Biosciences, 1:750) and 620 (NC1612059, Akoya Biosciences, 1:750) were used to label the gene-specific probes after signal amplification. A 3-plex human (320861, Advanced Cell Diagnostics), and non-human primate (320901, Advanced Cell Diagnostics) positive control probe was used for each species alongside a primate negative control probe (320871, Advanced Cell Diagnostics).

To separate fluorophore signals, multispectral imaging was carried out on a Zeiss LSM 880 in UT Southwestern's Quantitative Light Microscopy Core. Final imaging was carried out on the Zeiss LSM 710 and Zeiss LSM 880 confocal microscope at ×20 magnification in the UT Southwestern Neuroscience Microscopy Facility on chimpanzee and human samples.

To determine the composition of OPCs and MOLs in both BA23 (human: $n = 2$, chimpanzee: $n = 3$) and anterior cingulate cortex (human: $n = 3$, chimpanzee: $n = 3$), we sampled 2–4 vertical bins (layers 1–6) of

cortex from each individual and evenly divided each bin from the apical to basal boundary into 5 sections, and then we randomly selected 2–4 sub-areas (456 × 456 pixels) in each section to quantify the number of cells (DAPI, 405 nm), OPCs (*PDGFRA*, 488 nm) and MOLs (*MOG*, 555 nm) by using self-generated ImageJ Macro code and R script in Fiji and R, respectively (https://github.com/konopkalab/Comparative_snATAC_snRNA). Maximum intensity projection images were generated from 13 slices of *Z* stack. OPCs were defined as *PDGFRA* and DAPI double-positive cells, whereas MOLs were defined as *MOG* and DAPI double-positive cells. Data were analysed using a linear mixed model with species as the fixed factor and individual as the random factor per comparison (*lme4*[67] package in R, with REML = F).

To compare the expression of *FOXP2* in *THEMIS*$^+$ neurons between humans ($n = 3$) and chimpanzees ($n = 3$), we quantified the fraction of *FOXP2*$^+$ neurons in *THEMIS*$^+$ neurons, and the number of fluorescent puncta as a proxy for *FOXP2* expression levels in BA23. We sampled 2–3 images from the deep layers of each individual, and then we randomly selected 2–3 sub-areas of each image to quantify the fraction of DAPI (405 nm), *FOXP2* (488 nm) and *THEMIS* (555 nm) triple-positive neurons in DAPI and *THEMIS* double-positive neurons. Data were analysed using a mixed linear model using species, image and sub-area as the fixed factors. For quantification of puncta, we used the same images as for quantification of fraction but selected only the cells with individually distinguishable puncta. This resulted in the quantification of 3–11 *THEMIS*$^+$ neurons and 3–9 *THEMIS*$^-$ neurons per image. Data were analysed using a linear mixed model with species as the fixed factor and individual as the random factor per comparison (lme4[67] package in R, with REML = F).

snRNA-seq DGE and identification of species-specific gene expression

We carried out DGE analysis using two approaches: a single-cell-based DGE approach and a pseudobulk-based DGE approach. We retained the pseudobulk DGE results for all analyses as both the HS-Genes and CS-Genes were more reproducible with previous studies[12,18] compared to the single-cell-based DGE method (Extended Data Fig. 10c–e).

For the pseudobulk DGE method, we aggregated all cells per cell type and species using sumCountsAcrossCells from scuttle[68] and only retained the genes that were detected in all samples (UMIs > 0) of at least one species. DGE analysis was carried out using edgeR QLRT approach[69] and differentially expressed genes were determined with FDR (<0.05) and log[FC] (log[FC] > 0.3 or log[FC] < −0.3) cutoffs. DGE analysis was carried out with the following covariates: humanized age, sex and library batch. Humanized age was calculated as described before by linear modelling of life traits between species[70]. Genes with species-specific expression were determined as before[18]. Briefly, HS-Genes were determined as differentially expressed genes that are H > C = M or H < C = M (C = M was determined if FDR > 0.1; in which, H represents humans, C represents chimpanzees and M represents rhesus macaques). The same criteria were used for CS-Genes. Genes that are consistently different between macaque–human and macaque–chimpanzee were referred to as macaque versus human–chimpanzee genes.

For the single-cell DGE method, genes were tested for DGE using MAST[68]. The same covariates were used as in the pseudobulk method, except for cngeneson as recommended by the MAST approach[68]. Genes with FDR < 0.05 and absolute average log (ln) FC > 0.25 were considered significant. Genes with species-specific expression were determined as described for the pseudobulk method above.

snATAC-seq differential CRE accessibility and identification of species-specifically accessible CREs

As for DGE, we carried out differential CRE accessibility using two approaches: a single-cell-based approach and a pseudobulk-based approach. We retained the pseudobulk method results for all analyses as both the HS-CREs and CS-CREs were more reproducible with the previous study[12] compared to the single-cell based method (Extended Data Fig. 10f).

For the pseudobulk method, we used the edgeR QLRT approach, which is widely used for differential accessibility analysis[71], similar to the DGE analysis. We aggregated all cells per cell type and species and retained only the CREs that were detected in all samples (total detected reads > 3) of at least one species, and among the top 100,000 CREs by accessibility per cell type. Differentially accessible CREs were determined with FDR (< 0.05) and log[FC] (log[FC] > 0.3 or log[FC] < −0.3) cutoffs. Differentially accessible CRE analysis was carried out with the following covariates: humanized age and sex. Species-specifically accessible CREs were determined in the same manner as the species-specifically expressed genes described above.

For the single-cell method, CRE accessibility was used as the response variable and logistic regression was used to fit two models of covariates with or without species identity per comparison. Then, a log-likelihood ratio test was used to determine the *P* value, which was later adjusted with FDR correction. The covariates were: humanized age, sex and total gene activity (as a measure of cell depth and quality, calculated using Cicero[66]). To determine an effect-size cutoff, we first calculated a mean accessibility ratio among tested CREs per pairwise comparison (MeanAccChimp/MeanAccHuman) and used this to normalize accessibility of one species to another. This was then used to compute delta accessibility (HumanAcc − ChimpAccNormalized) for all CREs, which followed a normal distribution around zero. This calculation was carried out for each pairwise comparison per cell type and 1.5 s.d. away from the mean was used as the cutoff. Therefore, only CREs with FDR < 0.05 and s.d. > 1.5 were considered significant. Species-specifically differential CREs were determined with the same criteria used for species-specifically expressed genes as described above.

Analysis of previously published datasets

The data for ref. 11 were analysed through publicly available fastq files (Gene Expression Omnibus (GEO) accession: GSE127898). Preprocessing (until count matrix) was carried out in a way similar to that for our own dataset, including ambient RNA correction by CellBender[56]. As species were mixed in the same library, we assigned cell barcodes to a given species (humans, chimpanzees, bonobos, rhesus macaques) by counting reads with no mismatch (carried out for each species) and assigning the cell barcode to the species with the most counts. Our annotation corresponded with the original publication for >99.9% of the cell barcodes annotated in the original study[11]. We then used 200 UMIs as the cutoff, rather than 500 UMIs in the original study, as we are interested in the ratios of OPCs and MOLs, and glial cells have an overall lower number of UMIs (Extended Data Fig. 1e). We then used canonical markers to identify the major cell type and define the ratio of OPC and MOL nuclei per sample.

The data for ref. 12 were analysed through its supplementary tables. The overlap of CREs was tested for statistical significance with a Fisher's exact test. For overlap of species-specifically accessible CREs, we used the number of all CREs as the background.

The OLIG2 dataset from ref. 18 was deconvoluted using MuSiC[72] as previously carried out[73] except that the reference single-cell study was used from this dataset (humans, chimpanzees and macaques were used separately for comparisons).

The ref. 23 raw count matrix was filtered to contain only L5/6 CC *THEMIS*$^+$ neurons from the healthy controls. L5/6 CC was further subclustered and filtered to contain only *C1QL3*$^+$ subclusters. We also only retained the orthologous protein-coding genes initially identified for the original comparative analyses. DGE analysis was carried out between posterior cingulate cortex and prefrontal cortex and between posterior cingulate cortex and anterior cingulate cortex using the pseudobulk DGE (edgeR QLRT) approach as described before.

The data in ref. 7 were analysed for the proportional changes in the oligodendrocyte lineage. We obtained the metadata associated with the final count (NEMO identifier: dat-ek5dbmu) and computed the fraction of OPCs and MOLs in all glia per individual. For species with both single-nucleus chromatin accessibility and mRNA expression sequencing (SNARE-seq) and single-nucleus transcriptome (humans and marmosets) datasets, both datasets were used.

Epigenome–transcriptome associations

To test overlap of epigenomic and transcriptomic changes, we expanded the CREs on both sides of the TSS, either with increasing distance (Extended Data Fig. 7a) or for 500 kb to identify potential HS-CREs associated with HS-Genes. We used 500 kb as most physical interactions between enhancers and promoters are within 500 kb distance[74]. We determined that an HS-CRE and HS-Gene are associated if the following conditions are true: they are both found in the same cell type (neuronal subtypes are treated as different cell types); they are altered in the same direction (for example, HS-Open-CRE and HS-Up-Gene); the HS-CRE is within 500 kb on either side of the TSSs per HS-Gene.

Gene set enrichment analyses

GO enrichment for HS-Genes was carried out using the clusterProfiler package in R[75]. HS-Up-Genes and HS-Down-Genes were tested separately with all genes tested for differential expression used as the background. Background was calculated separately for each cell type. Only the GO enrichments with FDR < 0.05 and FC > 1.3 were considered significant.

Modern-variant-associated HS-CREs were first divided into HS-Open-CREs and HS-Closed-CREs. Then the nearby genes were annotated using annotatr[76]. Background genes were similarly identified by annotating all accessible CREs to their genes with annotr[76]. As for HS-Genes, GO enrichment was carried out using the clusterProfiler package in R. As only the L2-3_2 subtype showed enrichment, we carried out modern-variant GO enrichment only for this subtype.

Motif enrichment analysis

Non-redundant motifs for humans were downloaded from the JASPAR 2018 database[77]. A binary CRE motif matrix (CREs in the rows, motifs in the columns) was created using Signac, which calculates the motif matrix using motifmatchr[78]. We then tested the enrichment of motifs in HS-Open-CREs per cell type using a log-likelihood ratio test on two nested binomial linear regression models (Evolution: HS-Open-CRE or not):

H0: Evolution ~ CRE length

H1: Evolution ~ CRE length + Motif occurrence

CRE length was added as a covariate as longer CREs will include more motifs. To avoid capturing the motifs divergent between species in general, and to highlight the motifs divergent only in the human evolution, the background was selected as all evolutionarily divergent CREs for the given cell type. Motifs with FDR < 0.05 and log[FC] > 0 were considered as significantly enriched. Motif enrichments were clustered and visualized using pheatmap[79].

Visualization of CREs

To visualize CREs across all species, we converted raw chimpanzee and rhesus macaque snATAC-seq reads to human coordinates using CrossMap[53]. This was carried out separately for reads counted in each cell type. For more accurate comparisons of the track plots between the species, reads were randomly downsampled to the lowest number of reads detected per species for the given subtype. The reads were then converted to bigwig format and visualized using Integrated Genome Viewer[80]. Tracks were log transformed and are presented at the same scale for all comparisons.

Identification of cortical HARs

To identify HARs within the CREs in our dataset, we followed a similar approach to a previous study[81]. We first segmented each CRE into bins of 150 base pairs (the size of the smallest CRE in our dataset). We then used the following 15 primate species from a 30-species alignment from the University of California, Santa Cruz[82]: *H. sapiens*, *P. troglodytes*, *Gorilla gorilla*, *Nomascus leucogenys*, *M. mulatta*, *Macaca fascicularis*, *Macaca nemestrina*, *Cercocebus atys*, *Chlorocebus sabaeus*, *Mandrillus leucophaeus*, *Colobus angolensis*, *Callithrix jacchus*, *Saimiri boliviensis*, *Cebus capucinus* and *Aotus nancymaae*, and retained the CRE segments that have no gaps in at least 8 species and no gaps between humans, chimpanzees and rhesus macaques (the species used in our single-cell genomics experiments) using the rphast package[83].

To estimate the neutral substitutions per CRE, we padded each CRE by 25 kb upstream and 25 kb downstream, and ran the phyloFit function with the following parameters on the phylogenetic tree of all species: subst.mod = SSREV, EM = T, nrates = 4. To test acceleration in the human lineage, we then used the phyloP function on each CRE segment using its corresponding neutral model with the following parameters: method = LRT, mode = ACC, branches = hg38. The final list of HARs was determined using the cutoff P value < 0.001.

Identification of modern human variants

The original publication of modern human variants lists 321,820 human-specific substitutions that contain an ancestral allele either in the Altai Neanderthal or in the Altai Denisovan genome[33]. Since the original publication, two additional high-quality Neanderthal genomes have been reported[84,85]. We have therefore updated this original list of human-specific substitutions and retained only the substitutions that are different from in all reported high-quality archaic genomes (3 Neanderthals and 1 Denisovan). This resulted in 98,550 human-specific substitutions. The original publication had only retained the substitutions that are present in >90% of present-day humans using the human polymorphism dataset[33]. Since then, the human polymorphism dataset expanded from 1,092 individuals from 14 populations[86] to 2,504 individuals from 26 populations[87]. Therefore, we updated this cutoff with the most recent 1000 Genomes Project Phase 3 dataset[87], which reduced the number of human-specific substitutions to 91,488. As we were mainly interested in assessing the modern-variant enrichment in HS-CREs compared to all CREs, we further filtered for the modern human variants overlapping the CREs in our dataset, which resulted in 12,161 variants. Out of 12,161 variants, 1,920 variants (15.7%) overlapped HS-CREs.

Identification of human-specific substitutions

Our main objective to identify human-specific substitutions was to compare them with the modern human variants. As modern-variant analysis used only chimpanzees and gorillas as outgroup species[33], we also limited our comparison to apes. We used the 30-species genome-wide alignment and extracted the alignment for humans, chimpanzees, gorillas and gibbons. We excluded orangutans because the associated alignment was not based on synteny (it was based on reciprocal blast) and showed more missing elements in the alignment compared to other species. Using this four-way alignment, we then identified single nucleotides that are only different in humans and map to the CREs identified in this dataset and referred to them as human-specific substitutions. As for the modern human variants, we further filtered human-specific substitutions for presence in at least 90% of modern-day humans according to the 1000 Genomes Project Phase 3 database[87].

Analyses of HS-CRE enrichment in HARs and modern variants

For a full list of published HARs, we merged the bed files of a compendium of HARs[27] and another HAR study based on accessibility patterns

of chromatin[81] using bedtools[62]. To test the overlap of all CREs with published HARs, we generated background genomic regions of similar GC content and length using genNullSeqs from the gkmSVM package with default parameters[88]. We then randomly selected the same number of regions as the entire CRE list ($n = 100$) and tested for significantly higher overlap of HARs with the observed CREs compared to the randomized background using an empirical P value.

Enrichment of HARs in HS-CRE was tested by logistic regression. Predictor variables were CRE length and CRE evolution (HS or NS (non-significant)), and the response variable was whether the CRE contains a HAR or not. The effect of CRE evolution was tested with a likelihood ratio test. The test was carried out for each major cell type separately (excitatory neurons, inhibitory neurons, MOLs, OPCs, astrocytes, microglia).

To test which cell types evolved more recently after the split of modern humans from other ancient human species (Neanderthals and Denisovans)[33], we carried out a negative binomial regression. Predictor variables were CRE length and CRE evolution and the response variable was the number of overlapping modern variants. The effect of CRE evolution was tested with a likelihood ratio test. The test was carried out for each major cell type separately (excitatory neurons, inhibitory neurons, MOLs, OPCs, astrocytes, microglia). We also tested the overall enrichment of modern variants by considering HS-CREs as a CRE that is an HS-CRE in at least one cell type.

Reported P values were FDR adjusted in all enrichments. CRE length was used as a covariate in both enrichments as larger CREs tend to have more variants and a better chance to overlap HARs and modern variants.

Analysis of surgically resected human snATAC-seq

Raw fastq files were downloaded from the GEO database (accession number: GSE139914, brain region: BA38, middle temporal gyrus)[37]. We pre-processed the snATAC-seq as in the original publication; however, instead of carrying out peak calling to generate the peak–cell matrix, we counted the reads in the CREs identified in our dataset for direct comparison of accessibility on the same CREs. We extracted the excitatory cells as they were annotated in the original study and annotated the subtypes by co-clustering with the human snATAC-seq excitatory subtypes in this study. We then carried out differential accessibility analyses as described above, this time comparing the surgical human tissue and chimpanzee or rhesus macaque samples per excitatory subtype. Motif enrichment analyses were also carried out as before on the HS-Open-CREs per excitatory subtype.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Raw and processed data are available at National Center for Biotechnology Information GEO under the accession number GSE192774. Processed data associated with ref. 7 were accessed from https://assets.nemoarchive.org/dat-ek5dbmu. Other datasets were obtained using their GEO accession numbers (GSE127774, for ref. 11; GSE107638, GSE123936 and GSE139914, for ref. 18; GSE18653 for ref. 19).

Code availability

All analysis scripts are available at https://github.com/konopkalab/Comparative_snATAC_snRNA.

51. Zheng, G. X. et al. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
52. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
53. Zhao, H. et al. CrossMap: a versatile tool for coordinate conversion between genome assemblies. *Bioinformatics* **30**, 1006–1007 (2014).
54. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
55. Smith, T., Heger, A. & Sudbery, I. UMI-tools: modeling sequencing errors in unique molecular identifiers to improve quantification accuracy. *Genome Res.* **27**, 491–499 (2017).
56. Fleming, S. J., Marioni, J. C. & Babadi, M. Unsupervised removal of systematic background noise from droplet-based single-cell experiments using CellBender. Preprint at *bioRxiv* https://doi.org/10.1101/791699v2 (2019).
57. Howe, K. L. et al. Ensembl 2021. *Nucleic Acids Res.* **49**, D884–D891 (2021).
58. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902 (2019).
59. Picard Toolkit (Broad Institute, 2019); http://broadinstitute.github.io/picard/.
60. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
61. Kuhn, R. M., Haussler, D. & Kent, W. J. The UCSC genome browser and associated tools. *Brief. Bioinform.* **14**, 144–161 (2013).
62. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
63. Lareau, C. A., Ma, S., Duarte, F. M. & Buenrostro, J. D. Inference and effects of barcode multiplets in droplet-based single-cell assays. *Nat. Commun.* **11**, 866 (2020).
64. Stuart, T., Srivastava, A., Madad, S., Lareau, C. A. & Satija, R. Single-cell chromatin state analysis with Signac. *Nat. Methods* **18**, 1333–1341 (2021).
65. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
66. Pliner, H. A. et al. Cicero predicts cis-regulatory DNA interactions from single-cell chromatin accessibility data. *Mol. Cell* **71**, 858–871 (2018).
67. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *J. Statist. Softw.* **67**, 1–48 (2015).
68. Finak, G. et al. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* **16**, 278 (2015).
69. Chen, Y., Lun, A. T. & Smyth, G. K. From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. *F1000Res* **5**, 1438 (2016).
70. McCarthy, D. J., Campbell, K. R., Lun, A. T. & Wills, Q. F. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* **33**, 1179–1186 (2017).
71. Gontarz, P. et al. Comparison of differential accessibility analysis strategies for ATAC-seq data. *Sci. Rep.* **10**, 10150 (2020).
72. Wang, X., Park, J., Susztak, K., Zhang, N. R. & Li, M. Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. *Nat. Commun.* **10**, 380 (2019).
73. Mendizabal, I. et al. Cell type-specific epigenetic links to schizophrenia risk in the brain. *Genome Biol.* **20**, 135 (2019).
74. van Arensbergen, J., van Steensel, B. & Bussemaker, H. J. In search of the determinants of enhancer-promoter interaction specificity. *Trends Cell Biol.* **24**, 695–702 (2014).
75. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
76. Cavalcante, R. G. & Sartor, M. A. annotatr: genomic regions in context. *Bioinformatics* **33**, 2381–2383 (2017).
77. Khan, A. et al. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.* **46**, D260–D266 (2018).
78. Schep, A. motifmatchr: Fast motif matching in R. R version 1.4.0. (2018).
79. Kolde, R. pheatmap: Pretty heatmaps. R version 4.1.1. https://cran.r-project.org/web/packages/pheatmap/index.html (2012).
80. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
81. Gittelman, R. M. et al. Comprehensive identification and analysis of human accelerated regulatory DNA. *Genome Res.* **25**, 1245–1255 (2015).
82. Blanchette, M. et al. Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res.* **14**, 708–715 (2004).
83. Hubisz, M. J., Pollard, K. S. & Siepel, A. PHAST and RPHAST: phylogenetic analysis with space/time models. *Brief. Bioinform.* **12**, 41–51 (2011).
84. Mafessoni, F. et al. A high-coverage Neandertal genome from Chagyrskaya Cave. *Proc. Natl Acad. Sci. USA* **117**, 15132–15136 (2020).
85. Prufer, K. et al. A high-coverage Neandertal genome from Vindija Cave in Croatia. *Science* **358**, 655–658 (2017).
86. The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
87. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
88. Ghandi, M. et al. gkmSVM: an R package for gapped-kmer SVM. *Bioinformatics* **32**, 2205–2207 (2016).

Acknowledgements We thank C. Tamminga and K. Gleason at UT Southwestern for postmortem human tissues. We also thank H. Stroud and M. Chahrour for their critical comments on the manuscript. We also thank K. Luby-Phelps and S. Yamazaki for their help with imaging. G.K. is a Jon Heighten Scholar in Autism Research and Townsend Distinguished Chair in Research on Autism Spectrum Disorders at UT Southwestern. E.C. is a Neural Scientist Training Program Fellow in the Peter O'Donnell Brain Institute at UT Southwestern. This work was partially supported by the James S. McDonnell Foundation 21st Century Science Initiative in Understanding Human Cognition Scholar Award to G.K.; NHGRI (HG011641) to G.K., S.V.Y. and C.C.S.; National Science Foundation (SBE-131719 and EF-2021635) to S.V.Y. and C.C.S.; the NIMH (MH103517) to T.M.P., G.K. and S.V.Y.; NIH grants T32DA007290 and T32HL139438 to F.A., American Heart Association Postdoctoral Fellowship (915654) to Y.L. and NIMH grant MH126481 to R.M.V. and G.K. The National Chimpanzee Brain Resource was supported by NINDS (R24NS092988). Macaque tissue

collection and archiving was supported by funding from the NIH National Center for Research Resources (P51RR165; superseded by the Office of Research Infrastructure Programs (OD P51OD11132)) to the Yerkes National Primate Research Center. The Zeiss LSM 880 with Airyscan was supported by NIH grant 1S10OD021684-01 to K. Luby-Phelps.

Author contributions S.V.Y., G.K. and E.C. designed the study. T.M.P. and C.C.S. carried out tissue dissections. S.V.Y., T.M.P., C.C.S. and G.K. obtained funding. F.A. collected snRNA-seq and snATAC-seq data. Y.L., R.M.V. and E.O. carried out smFISH experiments. Y.L. carried out image quantification. E.C. carried out all analyses. T.M.P. edited the manuscript. E.C., S.V.Y. and G.K. wrote the manuscript.

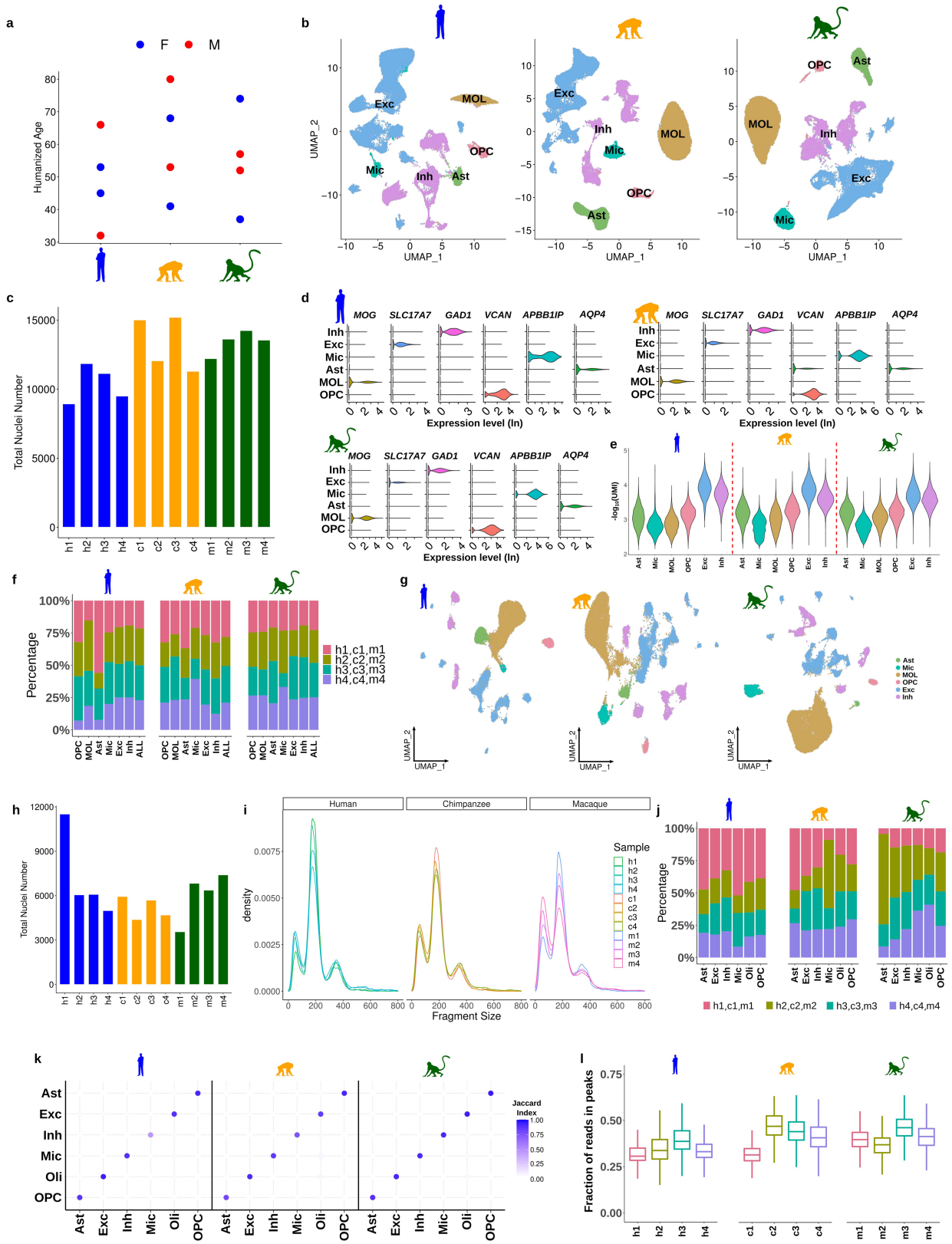
Competing interests The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-023-06338-4.

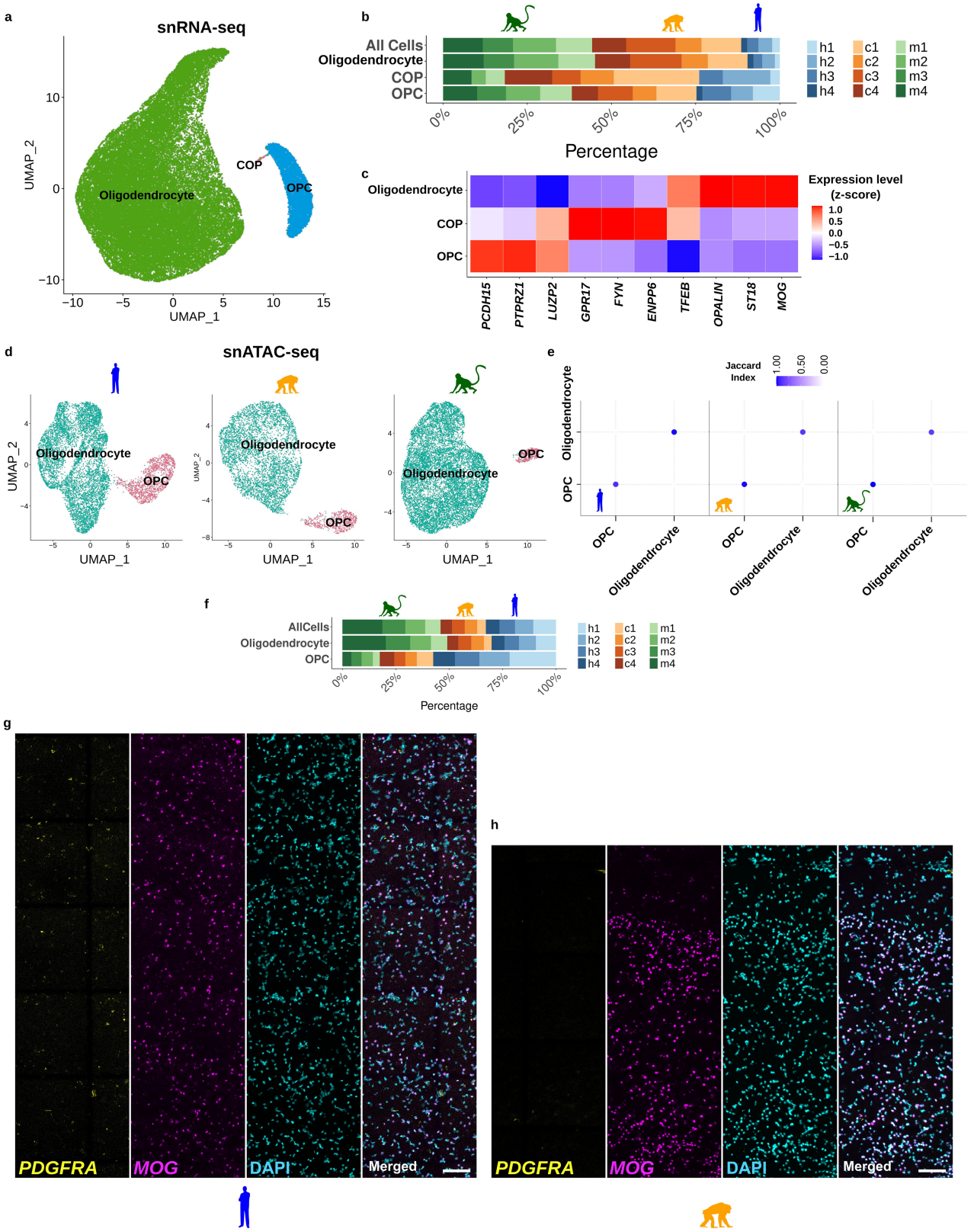
Correspondence and requests for materials should be addressed to Soojin V. Yi or Genevieve Konopka.

Peer review information *Nature* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

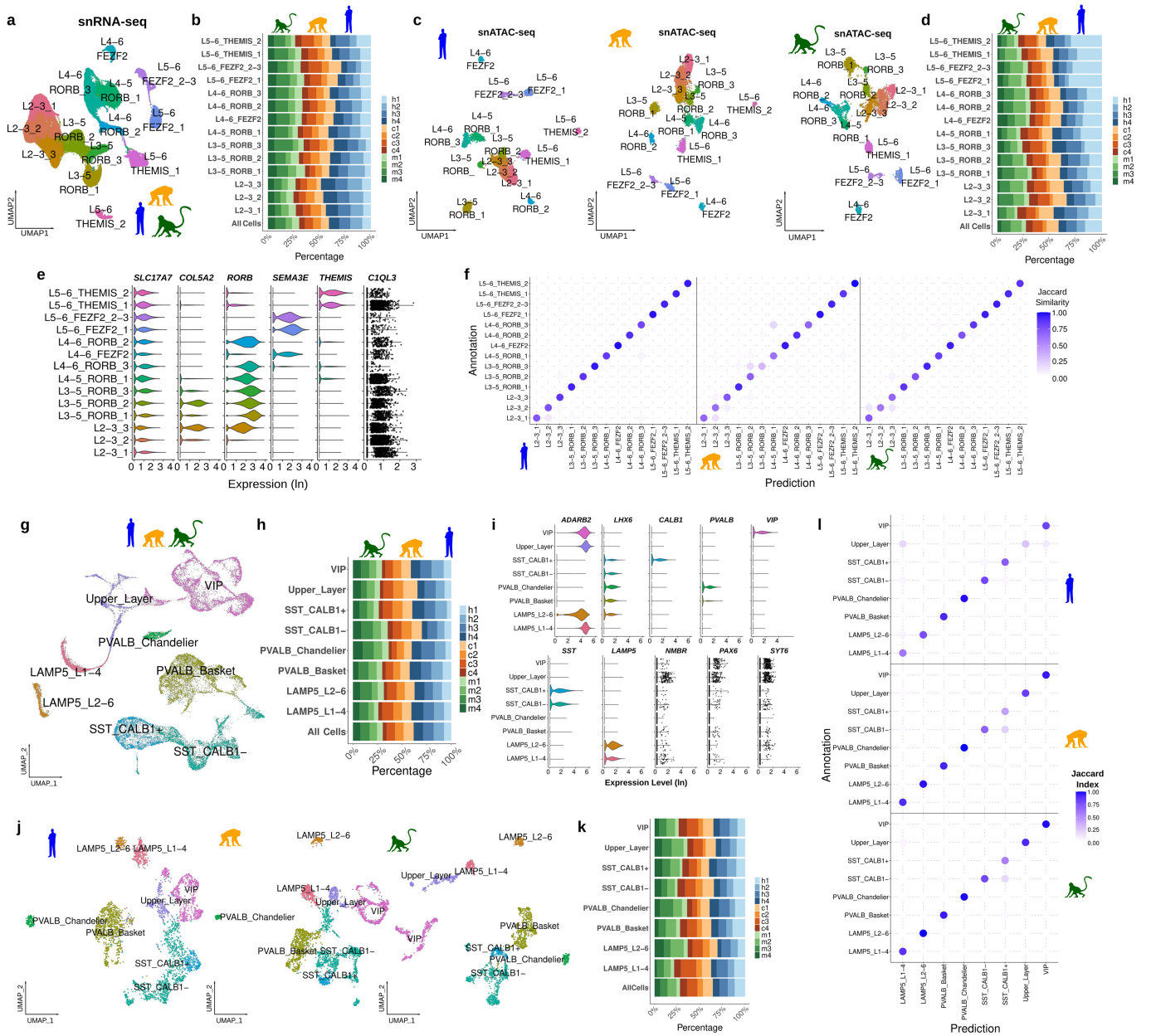


Extended Data Fig. 1 | See next page for caption.

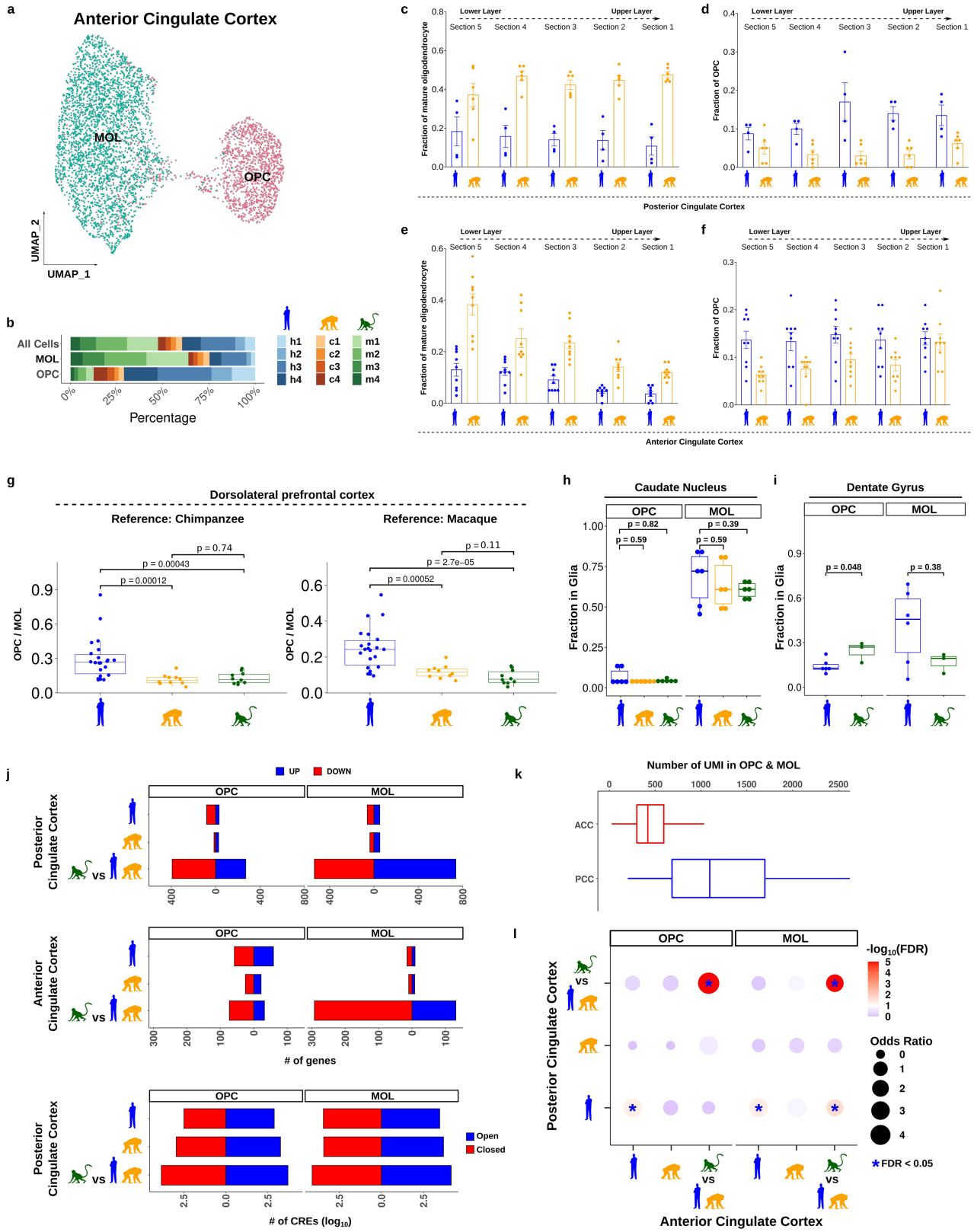
Extended Data Fig. 1 | Annotation and quality control of single-nuclei RNA-seq and single-nuclei ATAC-seq. (a) Distribution of sex and humanized age of samples. **(b)** Broadly annotated UMAP of nuclei per species. **(c)** Total nuclei number per sample after filtering. **(d)** Normalized, log (ln) transformed expression values of major cell type markers. **(e)** Violin plots of number of detected UMIs (log_{10} transformed) per major cell type. **(f)** Percentage of cells contributed per individual per species per major cell type. **(g)** Broad annotation of snATAC-seq data per species. **(h)** Total nuclei number per sample after quality control. **(i)** Nucleosome band pattern per sample; each line represents one sample. First, second and third peaks represent nucleosome free, mononucleosome and dinucleosome fractions, respectively. **(j)** Percentage of cells contributed per individual per species per major cell type. **(k)** Clarity of annotation transfer from snRNA-seq to snATAC-seq as displayed by Jaccard similarity index, which is the number of nuclei with the same final annotation and prediction (intersection) divided by the total number of nuclei with a given annotation or prediction (union). y-axis represents final annotation; x-axis represents the prediction which was assigned by label transfer per nucleus. Higher values indicate more similarity between final annotation and initial prediction. **(l)** Fraction of reads in peaks per sample (N = 9280, 5383, 5657, 4655, 5941, 4381, 5691, 4690, 3321, 6426, 5984, 6793 left to right). Boxplots represent median and interquartile range.



Extended Data Fig. 2 | Annotation of oligodendrocyte lineage cells. **(a)** UMAP visualization of integrated and annotated oligodendrocyte lineage nuclei in snRNA-seq. Oligodendrocyte: mature oligodendrocytes, COP: committer oligodendrocyte progenitor cells, OPC: oligodendrocyte progenitor cells. **(b)** Percentage of nuclei per sample for each subtype in snRNA-seq. **(c)** Normalized and scaled (z-scored) expression values of major oligodendrocyte lineage cell type markers. **(d)** UMAP visualization of annotated oligodendrocyte lineage nuclei in snATAC-seq per species. **(e)** Clarity of annotation transfer from snRNA-seq to snATAC-seq as displayed by Jaccard similarity index (similar to Extended Data Fig. 1k). **(f)** Percentage of cells contributed per individual per species per major cell type. **(g-h)** smFISH of *PDGFRA* (OPC) and *MOG* (MOL) in humans **(g)** and chimpanzees **(h)** (region: posterior cingulate cortex. Images span all cortical layers in both species. Scale bar is 100 µm). Similar results have been obtained for 4 bins across 2 humans and for 6 bins across 3 chimpanzees (see Extended Data Fig. 4c, d).

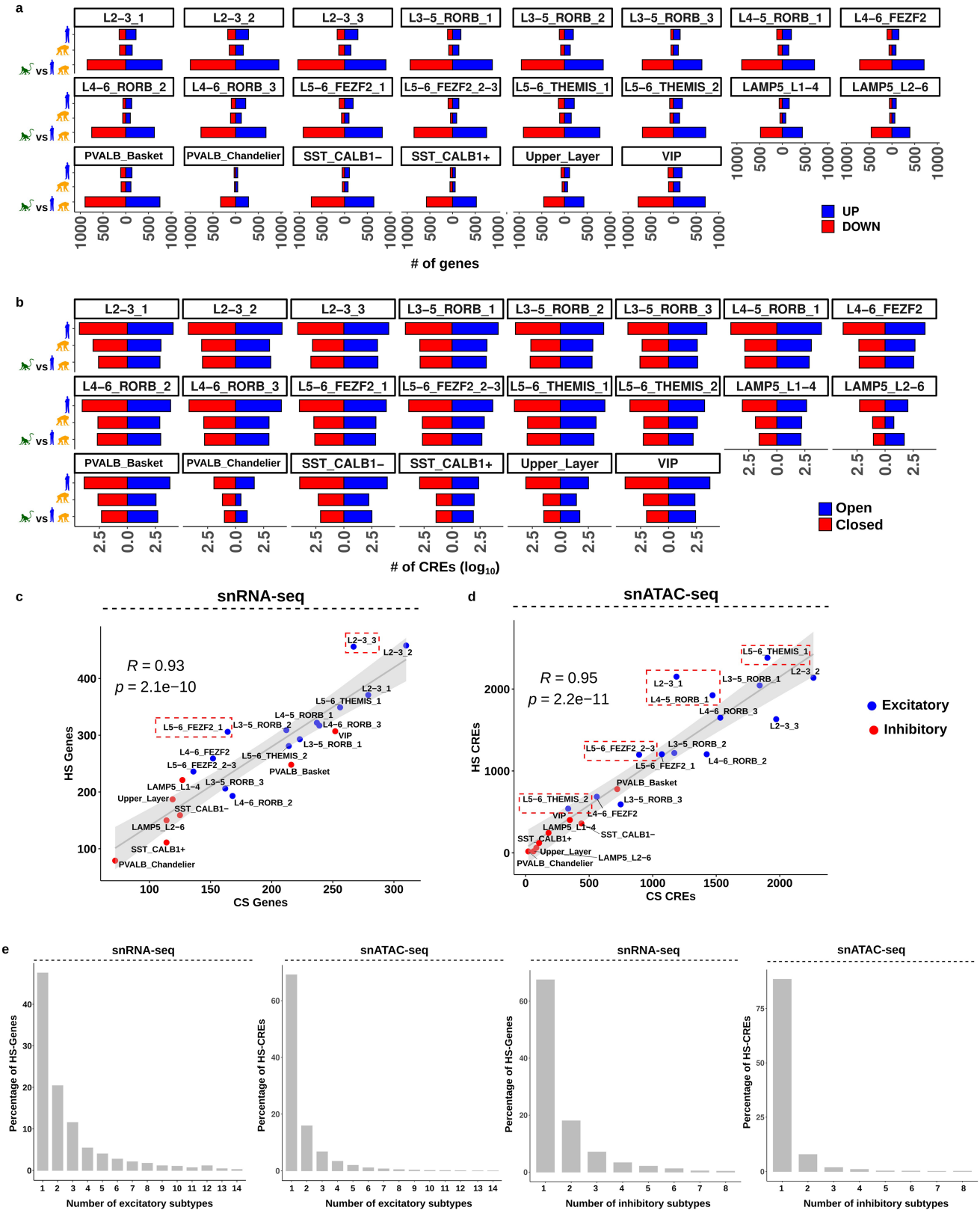


Extended Data Fig. 3 | Integration and annotation of neurons. **(a)** Annotated UMAP of excitatory neurons integrated across all species in snRNA-seq. **(b)** Percentage of nuclei per sample for each excitatory subtype in snRNA-seq. **(c)** Annotated snATAC-seq per species in the UMAP space. All 14 subtypes identified in snRNA-seq are also distinctly annotated in snATAC-seq for all species. **(d)** Percentage of nuclei per sample for each excitatory subtype in snATAC-seq. **(e)** Excitatory subtype markers for validation of annotation (expressions are normalized and log transformed). Note that the individual cells are plotted for *C1QL3* since the expression level is not sufficient for a violin plot. **(f)** Clarity of annotation transfer from snRNA-seq to snATAC-seq as displayed by Jaccard similarity index (similar to Extended Data Fig. 1k). **(g)** Annotated UMAP of inhibitory neurons integrated across all species in snRNA-seq. **(h)** Percentage of nuclei per sample for each inhibitory subtype in snRNA-seq. **(i)** Known inhibitory subtype markers for validation of annotation. Expression levels are normalized and log transformed. Note that the individual cells are plotted for *NMBR, PAX6, SYT6* since the expression level is not sufficient for a violin plot. **(j)** Annotated snATAC-seq per species in the UMAP space. All 8 subtypes identified in snRNA-seq are also distinctly annotated in snATAC-seq for all species. **(k)** Percentage of nuclei per sample for each inhibitory subtype in snATAC-seq. **(l)** Clarity of annotation transfer from snRNA-seq to snATAC-seq as displayed by Jaccard similarity index (similar to Extended Data Fig. 1k).

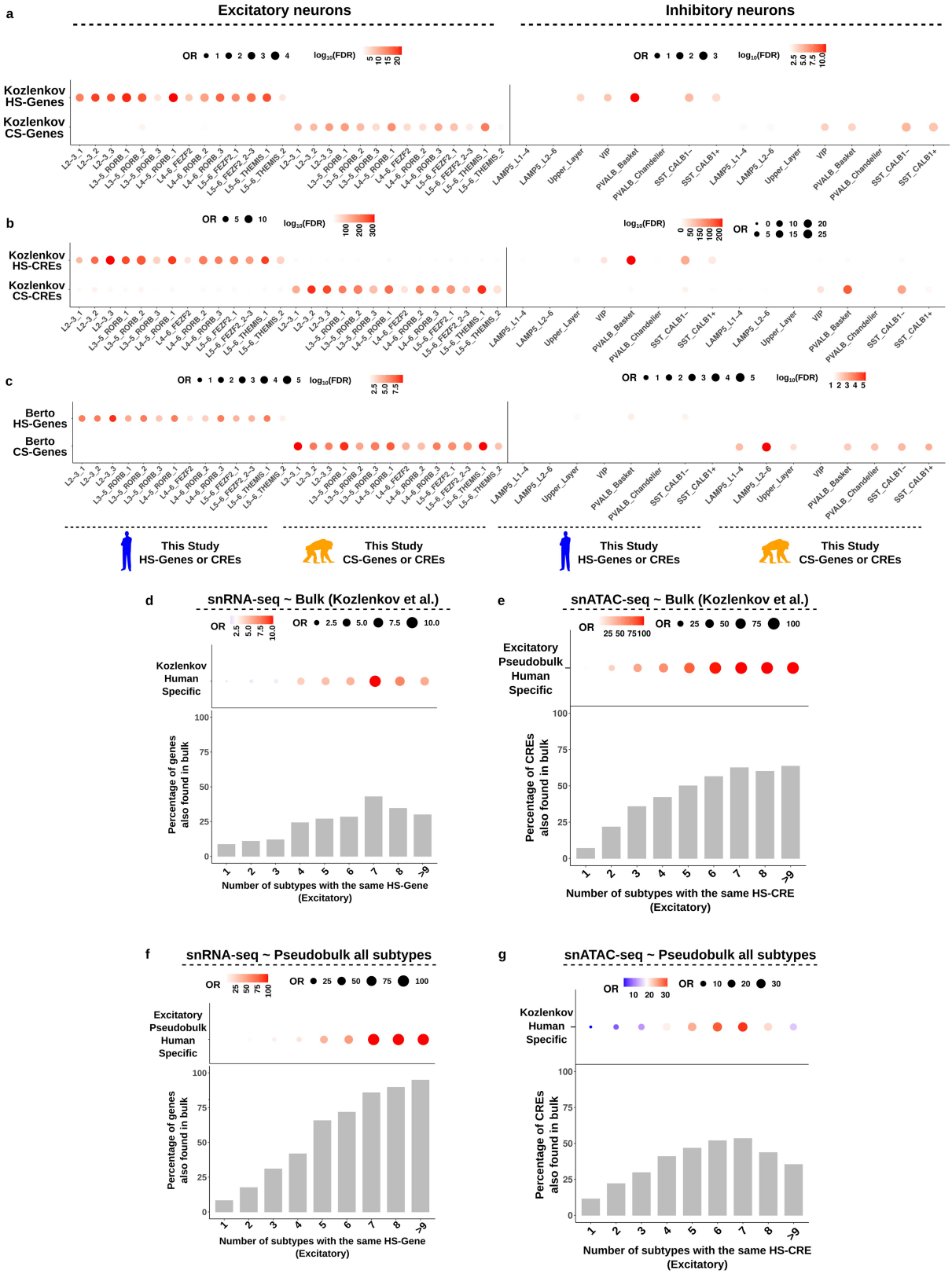


Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Additional analyses on the oligodendrocyte lineage. **(a)** UMAP of MOLs and OPCs in the anterior cingulate cortex (ACC). **(b)** Percentage of cells contributed per individual per species per cell type. **(c-f)** Fractions of MOLs and OPCs in smFISH experiments per section (see Fig. 1). Stitched column images encompassing all layers were divided into 5 equal parts from upper (Section 1) to lower layers (Section 5) in all images from human and chimpanzee. **(c-d)** are data from posterior cingulate cortex (PCC), and **(e-f)** are data from ACC. Each data point is a bin that contains sections from all layers. **c-d:** 4 bins from 2 humans, 6 bins from 3 chimpanzees. **e-f:** 9 bins from 3 humans and 3 chimpanzees. Data are represented as mean values +/− SEM. **(g)** Deconvoluted proportions from OLIG2+ bulk RNA-seq dataset (reference datasets: (left) chimpanzee, (right) rhesus macaque from this study). N = 22 (human), 10 (chimpanzee), 10 (rhesus macaque) individuals. P-value: Wilcoxon rank sum test, two-sided). **(h-i)** Fraction of OPCs or MOLs in glia in **(h)** caudate nucleus and **(i)** dentate gyrus per species. Each dot represents a sample (p-value: Wilcoxon rank sum test, two-sided. Caudate nucleus: N = 6 per species. Dentate gyrus: N = 6 for human, 3 for rhesus macaque). Box plots represent median and interquartile range in panels **g-i**. **(j)** Number of species-specific regulatory changes (PCC snRNA-seq (top), ACC snRNA-seq (middle), and PCC snATAC-seq (bottom, $\log_{10}$ transformed for better readability). **(k)** Distributions of UMIs (unique molecular identifiers) in ACC and PCC oligodendrocyte lineage nuclei (N = 12 individuals both for PCC and ACC). Box plots represent median and interquartile range. **(i)** Enrichment results between species-specifically expressed genes in ACC (x-axis) and PCC (this study, y-axis). Blue asterisk indicates a significant overlap (FDR < 0.05, Fisher's exact test, one-sided).

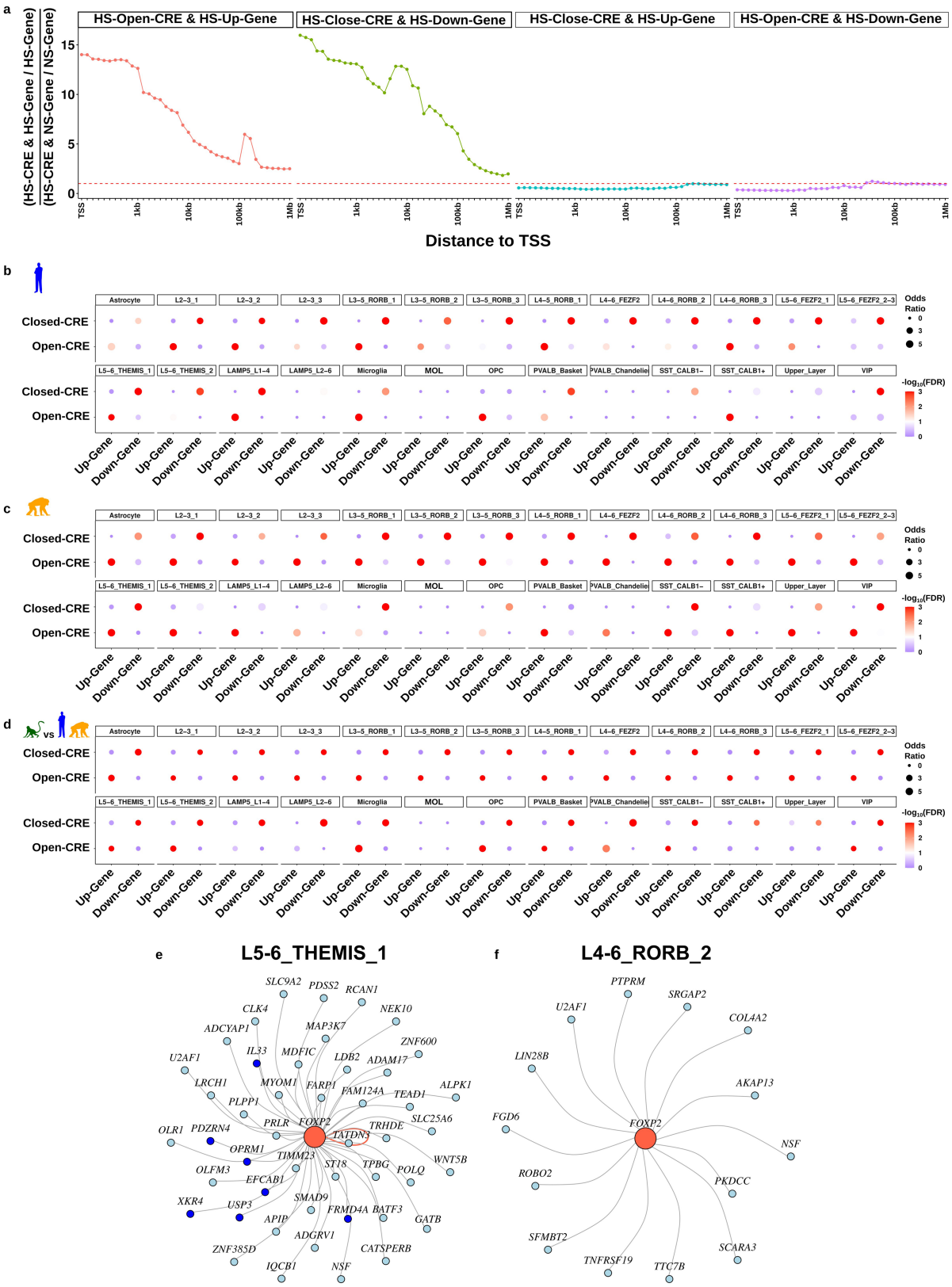


Extended Data Fig. 5 | Additional analyses of the regulatory changes in neuronal subtypes. (a-b) Number of regulatory changes that are human-specific, chimpanzee-specific or differential between rhesus macaque - human and rhesus macaque – chimpanzee in **(a)** snRNA-seq or **(b)** snATAC-seq (log_{10} transformed for better readability). **(c)** Scatter plots of number of HS-Genes and CS-Genes per neuronal subtype. Dashed rectangles indicate the subtypes with an excess number of human-specific regulatory gene expression changes (Two-sided chi-square test, FDR < 0.05). Shaded area indicates 95% confidence interval around the best fit (R indicates Spearman's rank correlation coefficient). **(d)** Same as **(c)** for HS-CREs and CS-CREs identified in snATAC-seq data. **(e)** Percentage distribution of excitatory HS-Genes that are found in only one subtype or shared among increasing number of subtypes (x-axis). Sum of all percentages equal 100. From left to right: in excitatory snRNA-seq, excitatory snATAC-seq, inhibitory snRNA-seq , inhibtory snATAC-seq.



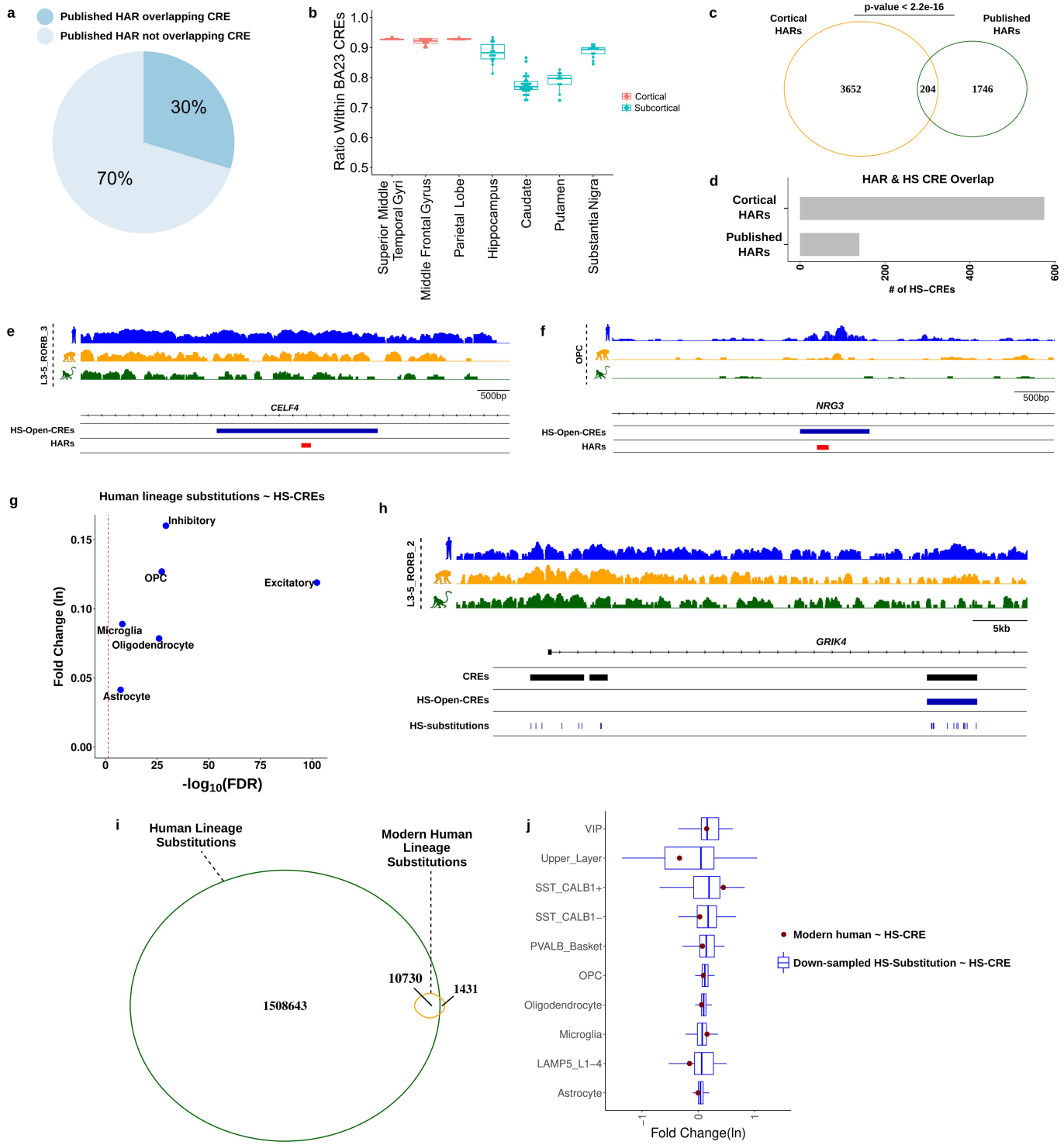
Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Comparisons of neuronal expression patterns between this dataset and previous comparative bulk datasets. **(a-c)** Enrichments of species-specific expression patterns between this study and previous bulk studies between excitatory neurons (left) and inhibitory neurons (right). **(a)** Transcriptomic changes between the Kozlenkov et al. dataset and this dataset, **(b)** epigenomic changes between the Kozlenkov et al. dataset and this dataset, **(c)** transcriptomic changes between the Berto et al. dataset and this dataset. FDR values are from a Fisher's exact test with multiple testing correction. **(d-e)** Subtype-specific changes are captured less in the bulk RNA-seq datasets. **(d)** Comparison of excitatory HS-Genes between a previous bulk analysis and this dataset. Top: odds ratio between the bulk dataset and this dataset with increasing subtype specificity of HS- Genes (from right to left). Bottom: percentage of HS- Genes that were also found in the bulk dataset. **(e)** Same comparison as **(d)** with HS-CREs. **(f-g)** Subtype-specific changes are captured less when the subtypes are combined within the same dataset. **(f)** Same comparison as **(d)** with HS-Genes but this time pseudobulk data results are obtained by combining the subtypes in this study. **(g)** Same comparison as **(f)** with HS-CREs.



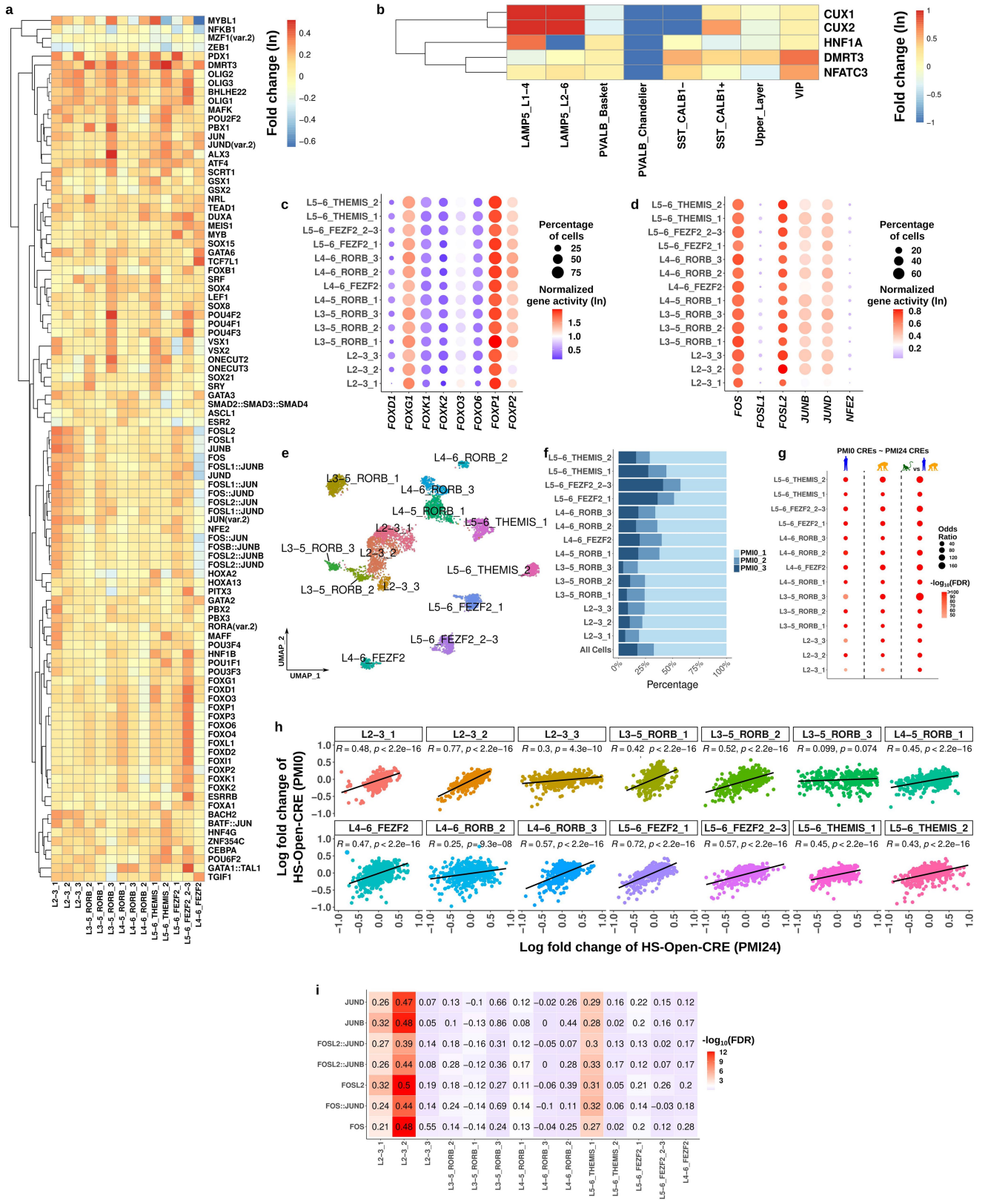
Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Associations between HS-Genes and HS-CREs. **(a)** The specificity of association between HS-Genes and HS-CREs decreases with increasing distance from the transcription start site (TSS). Y-axis shows the odds ratio, which is defined by the ratio of HS-Genes associated with HS-CREs divided by the ratio of not significant genes (NS-Genes) associated with HS-CREs. We calculated the odds ratio for increasing the distance from the TSS in both directions for four different associations (from left to right): HS-Open-CRE & HS-Up-Gene, HS-Open-CRE & HS-Down-Gene, HS-Close-CRE & HS-Up-Gene, HS-Open-CRE & HS-Down-Gene. The value for each observation was obtained by taking the mean across all cell types. **(b-d)** Enrichments between HS-CRE associated genes within a 10kb window from the TSS and HS-Genes per cell type. **(e-f)** Putative target genes of human-specific *FOXP2* upregulation in **(e)** L5-6_THEMIS_1 and **(f)** L4-6_RORB_2 cells. All genes show human-specific up / down regulation in their respective subtype and reside within 500kb of at least one human-specific chromatin accessibility change that has a FOXP2 motif. Dark blue circles indicate the genes that are not altered in the other 12 excitatory subtypes (similar to *FOXP2* itself). Red loop in (**a**) indicates that *FOXP2* itself is also identified with this analysis in the L5-6_THEMIS_1 subtype.



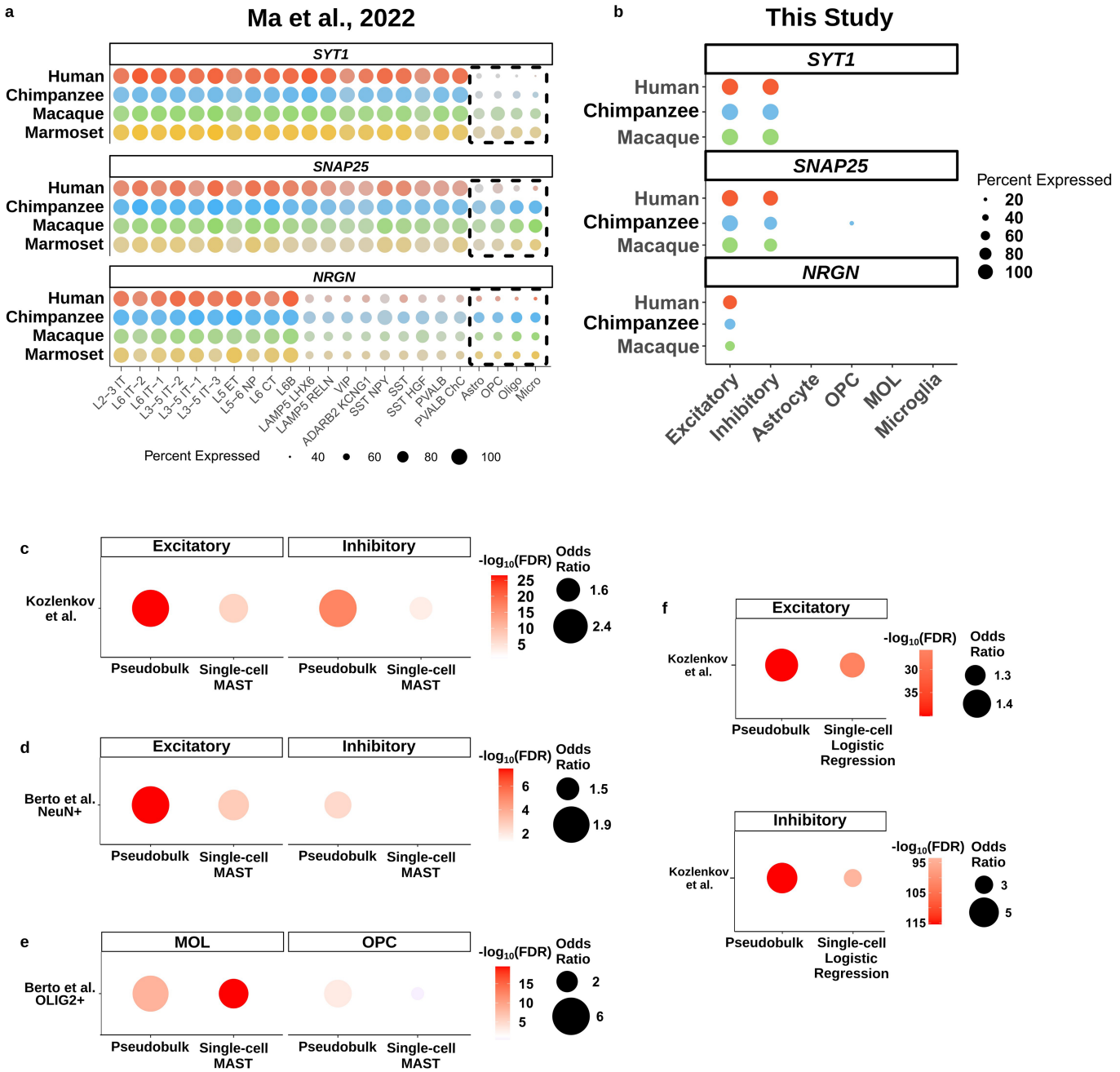
Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Further associations between human-specific substitutions and human-specific chromatin accessibility changes. **(a)** Pie-chart distribution of published HARs overlapping CREs in this dataset. **(b)** Ratio of non-BA23 CREs overlapping BA23 CREs (denominator: all CREs in BA23). Each dot represents an independent library prep. Red datasets indicate cortical regions, blue datasets indicate sub cortical regions. (Sample sizes; Superior Middle Temporal Gyri: 8, Middle Frontal Gyrus: 12, Parietal Lobe: 7, Hippocampus: 16, Caudate: 32, Putamen: 11, Substantia Nigra: 14. Box plots represent median and interquartile range). **(c)** Overlap between cortical HARs (identified in this study) and published HARs (p-value: One-sided chi-square test). **(d)** Number of HS-CREs associated with a cortical HAR or a published HAR. **(e-f)** Examples of HS-Open-CRE associated HARs. Bottom panel shows the multi-species alignment for *CELF4* HAR. Dots represent no change from the human (hg38) sequence. Human-specific changes conserved in other lineages are highlighted in shaded blue. **(g)** Enrichment of human-specific substitutions within the HS-CREs per major cell type. Enrichment is tested by a negative binomial regression model with CRE length and evolution of the CRE as the predictor variables (HS-CRE or not HS-CRE) and number of human-specific substitutions as a response variable(Significance: likelihood ratio test). **(h)** Example of an HS-Open-CRE with many human specific substitutions. **(i)** Overlap of substitutions that are specific to the human lineage (in comparison to chimpanzee, gorilla and gibbon) and previously identified modern human substitutions. **(j)** Log fold changes of substitution and HS-CRE association for substitutions on the human (blue boxplots) and modern human lineage (tile red dots) per cell type (except for excitatory cells). Human lineage-specific substitutions were randomly down sampled for 100 times to 12,161 (the number of modern human-specific variants) for comparison. Box plots represent median and interquartile range.



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Supplementary motif enrichment results. **(a-b)** Hierarchical clustering of motif enrichments (log-fold change) in HS-Open-CREs across **(a)** excitatory and **(b)** inhibitory neuronal subtypes. Transcription factors (TFs) associated with each motif enrichment are displayed in rows and the neuronal subtypes are displayed in columns. Only the motifs enriched in at least one subtype are displayed. **(c-d)** Accessibility of **(c)** FOX and **(d)** FOS / JUN family TFs. Accessibility is assessed by the normalized gene activity scores (calculated using Cicero[66]) per gene per subtype. **(e)** Annotated UMAP of excitatory neurons in snRNA-seq of surgically resected samples (referred to as PMI0 compared to postmortem BA23 human samples that are referred to as PMI24 in this figure). **(f)** Percentage of nuclei per sample for each excitatory subtype in snRNA-seq. **(g)** Enrichments of species-specifically expressed genes when PMI0 or PMI24 datasets were used as the human dataset in the comparative analyses. **(h)** Pearson correlations (test for p-value is two-sided) between the log fold changes of HS-Open-CRE motif enrichments when PMI0 or PMI24 datasets were used as the human dataset in the comparative analyses. **(i)** Heatmap of motif FOS / JUN motif enrichments per excitatory subtype in HS-Open-CREs. Colors correspond to $-\log_{10}$(FDR); numbers correspond to log fold change of enrichment.



Extended Data Fig. 10 | Comparisons with external datasets. (a) Expression levels of three ambient RNA markers highly expressed in neurons (*SYT1*, *SNAP25* and *NRGN*[21]) in the Ma et al. dataset[14]. The dot plot is generated through the interactive web tool linked to the original publication. Dashed square brackets indicate glial cell types, which show exceptionally low levels in the human dataset. Note that the smallest dot shows the presence of a transcript in 40% of the cells. **(b)** Same as **(a)** using this PCC dataset. Neuronal ambient RNA markers are detected at very low levels in glial cells across species after ambient RNA removal. **(c-e)** Enrichment of HS-Genes between the previous study (y-axis) and the current study (x-axis) with two alternative methods. **(f)** Enrichment of HS-CREs between the previous study (y-axis) and the current study (x-axis) with two alternative methods. For simplicity, we combined all HS-Genes from the subtypes of a major cell type (e.g. all excitatory neuronal subtypes were combined for the excitatory cell type comparisons). P-values were computed using a Fisher's exact test (one-sided) and false discovery rate (FDR) was calculated per panel.

nature research

Corresponding author(s): Genevieve Konopka and Soojin V. Yi

Last updated by author(s): May 4, 2023

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	R v4.0.1
Data analysis	Datasets were generated as detailed in the Methods. Barcode correction and alignment to each species own reference genome was done using cellranger software (namely cellranger mkfastq and cellranger count) and other tools detailed in the analysis to obtain the final gene-cell or peak-cell matrix (namely samtools, featureCounts, umi_tools, CellBender and custom scripts for snATAC-seq peak-cell matrix generation). Gene-cell and peak-cell matrices across human, chimpanzee and rhesus macaque were post-processed for custom analyses as detailed in the Methods. For each analyses, we explicitly report the tools and our approach. We also provide the full script for full transparency. Here is the list and versions of standalone tools we have used: Cellranger: 6.0.0 CellBender: 0.1.0, samtools: 1.6, featureCounts (subread): 1.6.3 umi_tools: 1.1.0 picard: 2.10.3 bedtools: 2.29.2 IGV: 2.4.14 Major R libraries used: Seurat (v3.9.9), Signac (v1.2.0), Cicero(v1.0.0), clusterProfiler (v4.6.2), harmony (v0.1), lme4 (v1.1-14), MAST (v1.0.1) edgeR (v3.40.2), scater (v1.26.1), MuSiC (v1.0.0), annotatr (v1.24.0), motifmatchr (v1.4.0), pheatmap (v1.0.12), rphast (v1.6.9), gkmSVM (v.81.0)

Custom R and python codes used for this analysis are deposited and available at: https://github.com/konopkalab/Comparative_snATAC_snRNA

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Raw and processed data are available at NCBI GEO under the accession number GSE192774. Reviewer token: yvwnyycgrzytjoj. Accession for other datasets are as follows: Bakken et al: https://assets.nemoarchive.org/dat-ek5dbmu. Khrameeva et al: GSE127774. Berto et al: GSE107638, GSE123936, GSE139914. Franjic et al: GSE186538

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For the datasets we generated, no statistical methods were used to predetermine sample size. However, our sample size of n=4 for single nuclei RNA-seq is similar to other published studies. (e.g. PMID 32424074 used n=3 and PMID 35322231 used n=2 or 3)
Data exclusions	Data that represented ambient RNA and/or doublets were excluded. This exclusion criteria was pre-established.
Replication	We replicated a key finding related to oligodendrocytes in 2 other available datasets as well as by smFISH.
Randomization	Samples were not randomized. All known technical and biological covariates were included in the statistical models, and therefore randomization is not relevant.
Blinding	Data collection and analysis were not performed blind to the conditions of the experiments. Data collection and analysis were carried out by different individuals. For collection, in order for samples to be assigned to the correct group (human, chimpanzee, rhesus macaque), knowledge of species could not be blind. For analysis, in order to align reads to the correct genome, we needed to know which samples belonged to each species.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	All relevant details of primate samples (Chimpanzee, rhesus macaque and human) are described and in the Source data file, Table S1.
Wild animals	This study did not involve wild animals.
Field-collected samples	This study did not involve field-collected samples.
Ethics oversight	All use of Non-human primate material was approved by UT Southwestern Medical Center Institutional Biosafety Committee (NHMSR-2021-012). UT Southwestern Medical Center Institutional Review Board has determined that as this research was conducted using postmortem specimens, the project does not meet the definition of human subjects research and does not require IRB approval and oversight.

Note that full information on the approval of the study protocol must also be provided in the manuscript.