

Cell-type-specific 3D epigenomes in the developing human cortex

<https://doi.org/10.1038/s41586-020-2825-4>

Received: 1 September 2019

Accepted: 23 July 2020

Published online: 14 October 2020

 Check for updates

Michael Song^{1,2,15}, Mark-Phillip Pebworth^{3,4,15}, Xiaoyu Yang^{1,15}, Armen Abnousi⁵, Changxu Fan^{6,7}, Jia Wen⁸, Jonathan D. Rosen⁹, Mayank N. K. Choudhary^{6,7}, Xiekui Cui¹, Ian R. Jones¹, Seth Bergenholtz¹, Ugomma C. Eze^{4,10}, Ivan Juric⁵, Bingkun Li¹, Lenka Maliskova¹, Jerry Lee¹, Weifang Liu⁹, Alex A. Pollen^{4,11}, Yun Li^{8,9,12}, Ting Wang^{6,7,13}, Ming Hu^{5,15}, Arnold R. Kriegstein^{4,11} & Yin Shen^{1,11,14}✉

Lineage-specific epigenomic changes during human corticogenesis have been difficult to study owing to challenges with sample availability and tissue heterogeneity. For example, previous studies using single-cell RNA sequencing identified at least 9 major cell types and up to 26 distinct subtypes in the dorsal cortex alone^{1,2}. Here we characterize cell-type-specific *cis*-regulatory chromatin interactions, open chromatin peaks, and transcriptomes for radial glia, intermediate progenitor cells, excitatory neurons, and interneurons isolated from mid-gestational samples of the human cortex. We show that chromatin interactions underlie several aspects of gene regulation, with transposable elements and disease-associated variants enriched at distal interacting regions in a cell-type-specific manner. In addition, promoters with increased levels of chromatin interactivity—termed super-interactive promoters—are enriched for lineage-specific genes, suggesting that interactions at these loci contribute to the fine-tuning of transcription. Finally, we develop CRISPRview, a technique that integrates immunostaining, CRISPR interference, RNAscope, and image analysis to validate cell-type-specific *cis*-regulatory elements in heterogeneous populations of primary cells. Our findings provide insights into cell-type-specific gene expression patterns in the developing human cortex and advance our understanding of gene regulation and lineage specification during this crucial developmental window.

The human cortex undergoes extensive expansion during development, a process which is markedly different and features distinct cell types from mouse cortical development. Much of its diversity arises from cortical stem cells known as radial glia (RG), which give rise to intermediate progenitor cells (IPCs) and excitatory neurons that undergo radial migration until they reach the cortical plate^{3,4}. Meanwhile, interneurons migrate tangentially into the dorsal cortex through the marginal and germinal zones⁵. Dynamic changes in the epigenomic landscape have been shown to influence development and cell fate commitment—for example, through the rewiring of physical chromatin loops between promoters and distal regulatory elements including enhancers⁶. These interactions are of particular interest as their dysregulation has been linked to alterations in gene expression and complex disorders and traits^{7,8}. Although previous studies have investigated bulk tissues, including the cortical plate and germinal zone⁹, detailed characterizations are missing for specific cell types.

Here we describe an approach for isolating RG, IPCs, excitatory neurons, and interneurons from mid-gestational human cortex samples, enabling a comparison of their 3D epigenomes. Furthermore, we develop CRISPRview, a sensitive technique for validating cell-type-specific distal regulatory elements in single cells. Our results identify key mechanisms that underlie gene regulation and lineage specification during human corticogenesis, and provide a framework for the understanding of diverse processes in development and disease.

Sorting cell types from the developing human cortex

To isolate cell types from human cortex samples between gestational weeks 15 to 22 (Supplementary Table 1), we expanded upon an established approach for isolating RG from human cortical samples using fluorescence-activated cell sorting (FACS)¹⁰. Germinal zone and cortical plate samples were dissociated, stained using antibodies for EOMES,

¹Institute for Human Genetics, University of California, San Francisco, San Francisco, CA, USA. ²Pharmaceutical Sciences and Pharmacogenomics Graduate Program, University of California, San Francisco, San Francisco, CA, USA. ³Biomedical Sciences Graduate Program, University of California, San Francisco, San Francisco, CA, USA. ⁴The Eli and Edythe Broad Center of Regeneration Medicine and Stem Cell Research, University of California, San Francisco, San Francisco, CA, USA. ⁵Department of Quantitative Health Sciences, Lerner Research Institute, Cleveland Clinic Foundation, Cleveland, OH, USA. ⁶Department of Genetics, Washington University School of Medicine, St Louis, MO, USA. ⁷The Edison Family Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St Louis, MO, USA. ⁸Department of Genetics, University of North Carolina, Chapel Hill, NC, USA. ⁹Department of Biostatistics, University of North Carolina, Chapel Hill, NC, USA. ¹⁰Medical Scientist Training Program, University of California, San Francisco, San Francisco, CA, USA. ¹¹Department of Neurology, University of California, San Francisco, San Francisco, CA, USA. ¹²Department of Computer Science, University of North Carolina, Chapel Hill, NC, USA. ¹³McDonnell Genome Institute, Washington University School of Medicine, St Louis, MO, USA. ¹⁴Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, CA, USA. ¹⁵These authors contributed equally: Michael Song, Mark-Phillip Pebworth, Xiaoyu Yang, Ming Hu. ✉e-mail: hum@ccf.org; arnold.kriegstein@ucsf.edu; yin.shen@ucsf.edu

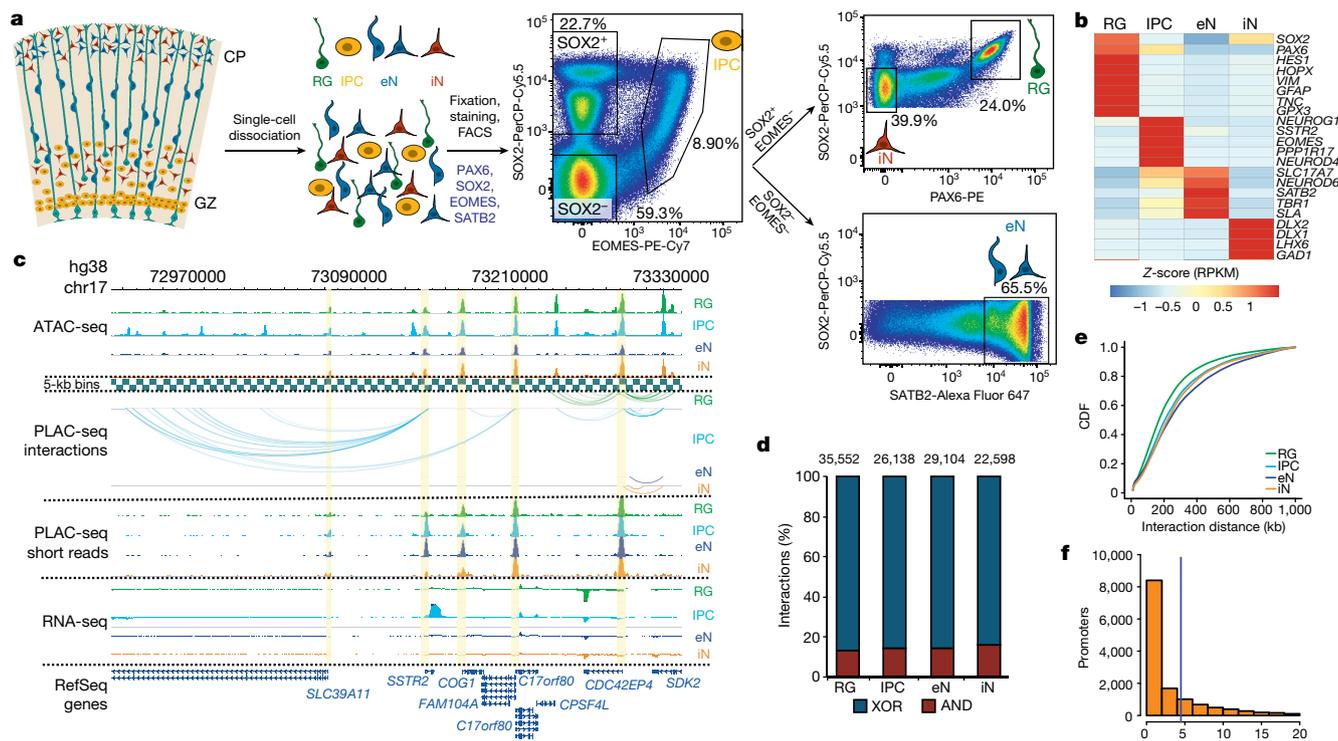


Fig. 1 | Experimental design and features of 3D epigenomes during human corticogenesis. **a**, Schematic of the sorting strategy. Microdissected germinal zone (GZ) and cortical plate (CP) samples were dissociated into single cells before being fixed, stained with antibodies for PAX6, SOX2, EOMES and SATB2, and sorted using FACS. **b**, Heat map displaying the expression of key marker genes for each cell type. **c**, WashU Epigenome Browser snapshot displaying a region (chr17: 72,970,000–73,330,000) with cell-type-specific interactions

linked to *SSTR2* expression in IPCs. **d**, Bar graph of interaction counts for each cell type, with the proportions of anchor to anchor (red) and anchor to non-anchor (blue) interactions highlighted. ‘AND’ or ‘XOR’ interactions are classified based on whether both or only one of the interacting 5-kb bins overlap anchor bins, respectively. **e**, Cumulative distribution function (CDF) plots of interaction distances for each cell type. **f**, Histogram displaying the numbers of interactions for interacting promoters across all cell types.

SOX2, PAX6 and SATB2, and partitioned into their constituent populations using FACS (Fig. 1a, Extended Data Fig. 1). IPCs were isolated as the EOMES⁺ population, whereas excitatory neurons were isolated from the EOMES⁻ and SOX2⁻ population based on high SATB2 expression¹. RG were isolated based on high SOX2 and high PAX6 expression, and interneurons were isolated based on medium SOX2 and low PAX6 expression. The gene expression profiles of the sorted cell populations were both highly consistent with cellular identity and reproducible between individuals (Fig. 1b, Extended Data Fig. 2a, b).

Characterizing cell-type-specific 3D epigenomes

We used histone 3 lysine 4 trimethylation (H3K4me3) proximity ligation-assisted chromatin immunoprecipitation sequencing (PLAC-seq)¹¹ to identify chromatin interactions at active promoters and assay for transposase-accessible chromatin using sequencing (ATAC-seq) to profile open chromatin peaks for the sorted cell populations (Fig. 1c, Supplementary Table 2). After confirming that the samples cluster by cellular identity (Extended Data Fig. 2c, d), we applied the model-based analysis of PLAC-seq (MAPS) pipeline¹² to call significant H3K4me3-mediated chromatin interactions at a resolution of 5 kb. We identify 35,552, 26,138, 29,104, and 22,598 interactions in RG, IPCs, excitatory neurons, and interneurons, respectively, with approximately 85% of the interactions classified as anchor to non-anchor, and the remaining interactions classified as anchor to anchor (Fig. 1d, Extended Data Fig. 3a, b). The median interaction distance was between 170 and 230 kb (Fig. 1e), with an average of 4 to 5 interactions per promoter (Fig. 1f), and the majority of interactions occurred within topologically

associated domains in the germinal zone or cortical plate (Extended Data Fig. 3c).

Chromatin interactions influence transcription

We characterized the extent to which H3K4me3-mediated chromatin interactions influence cell-type-specific transcription. First, the sorted cell populations cluster by developmental age on the basis of their interaction strengths across all interacting loci (Fig. 2a). This is consistent with interneurons at this age possessing progenitor-like characteristics, including high *SOX2* expression. Meanwhile, genes that participate in cell-type-specific interactions are enriched for biological processes linked to their respective cell types, including cell proliferation for RG and IPCs and neuron projection development for IPCs and excitatory neurons (Extended Data Fig. 4a, Supplementary Table 3). Interaction strength and gene expression are positively correlated (Fig. 2a, b, Extended Data Fig. 4b), suggesting that chromatin interactions orchestrate transcription in a manner that is distinctly cell-type-specific. Next, we leveraged the enrichment of open chromatin peaks at distal interacting regions (Fig. 2c, Extended Data Fig. 4c) and performed transcription-factor motif enrichment analysis for distal interacting regions in each cell type¹³ (Fig. 2d, Supplementary Table 4). The motifs for PAX6, EOMES, and *TBR1* are enriched in RG, IPCs, and excitatory neurons, respectively, recapitulating their sequence of expression along this developmental trajectory¹⁴. The motifs for *DLX1*, *DLX2*, *DLX6*, *GSX2*, and *LHX6* are enriched in interneurons, in accordance with their roles in the maturation and function of interneurons¹⁵. Finally, we detect motifs that are enriched in distal interacting regions

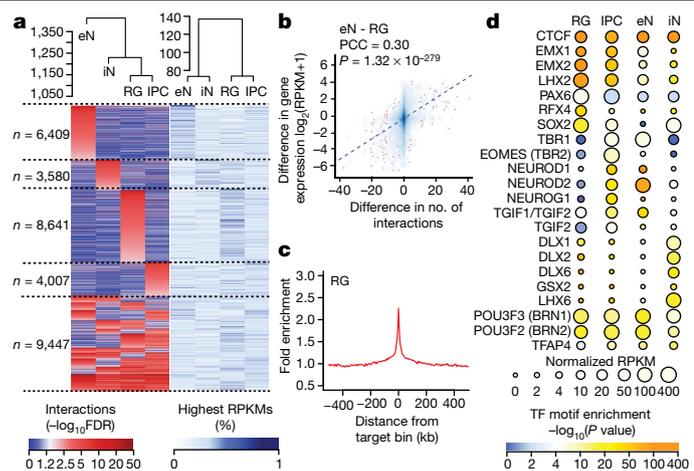


Fig. 2 | H3K4me3-mediated chromatin interactions influence cell-type-specific transcription. **a**, Heat maps showing interaction strengths (left) and gene expression (right) for anchor to non-anchor interactions grouped according to their cell type specificity. Interaction strengths are based on the $-\log_{10}$ -transformed false discovery rate (FDR) from the MAPS pipeline. **b**, Scatter plot showing the correlation between the difference in the number of interactions for each promoter and the difference in the expression of the corresponding genes for RG and excitatory neurons. $P = 1.32 \times 10^{-279}$, two-tailed Pearson product-moment correlation coefficient; $n = 13,996$ anchor bins with promoters. The trend line from linear regression is shown. RPKM, reads per kilobase of transcript, per million mapped reads. **c**, Fold enrichment of open chromatin peaks compared with distance-matched background regions in 1-Mb windows around distal interacting regions in RG. **d**, Transcription-factor (TF) motif enrichment analysis for open chromatin peaks at cell-type-specific distal interacting regions in each cell type. We analysed 4,203, 1,412, 3,088, and 949 regions in RG, IPCs, excitatory neurons (eNs), and interneurons (iNs), respectively. Colours represent enrichment scores based on the P value from HOMER, and sizes represent the gene expression of the corresponding TFs.

for co-expression modules in the developing human cortex¹ (Supplementary Table 5). Our results identify key lineage-specific transcription factors while linking them to their interacting genes, enabling insights into gene regulatory networks during human corticogenesis.

SIPs are enriched for lineage-specific genes

The number of chromatin interactions at H3K4me3-mediated anchor bins is only modestly correlated with gene expression (Extended Data Fig. 5a). A potential explanation for this is that individual genes are expressed to varying degrees in the contexts of their diverse cellular functions, and a subset of regulatory elements may be better described as fine-tuning rather than independently inducing or silencing transcription. Multiple regulatory interactions can also exert synergistic or non-linear effects on gene regulation. Cell-type-specific genes tend to form more chromatin interactions than shared genes across all four cell types (Extended Data Fig. 5b). By ranking anchor bins according to their cumulative interaction scores, we delineate a subset of promoters with increased levels of chromatin interactivity, termed super-interactive promoters (SIPs) (Fig. 3a, Extended Data Fig. 5c). We identify 755, 765, 638, and 663 SIPs in RG, IPCs, excitatory neurons, and interneurons, respectively (Extended Data Fig. 5d, Supplementary Table 6). SIPs are enriched for key lineage-specific genes including *GFAP* and *HES1* for RG, *EOMES* for IPCs, *SATB2* for excitatory neurons, and *DLX5*, *DLX6*, *GADI1*, *GAD2*, and *LHX6* for interneurons. We also observe forebrain-specific SIPs including *FOXG1* in all four cell types, progenitor-specific SIPs including *SOX2* in RG, IPCs, and interneurons, and cortical neuron-specific SIPs including *TBR1* in IPCs and excitatory neurons. Numerous promoters for long intergenic non-coding RNAs

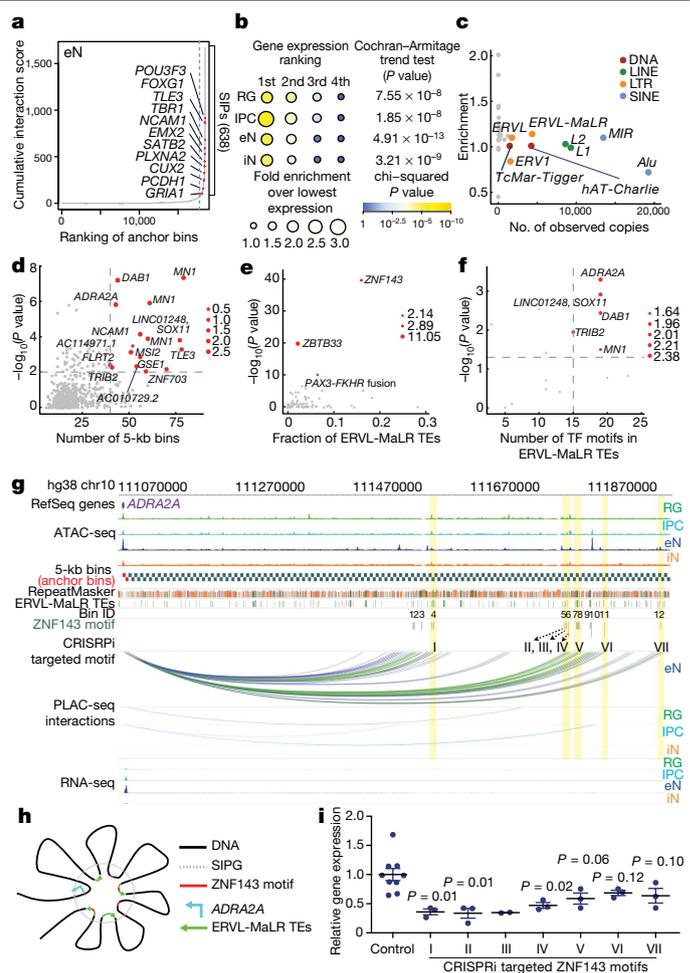


Fig. 3 | SIPs are enriched for lineage-specific genes. **a**, Anchor bins were ranked according to their cumulative interaction scores in excitatory neurons. SIPs are located past the point in each curve at which the slope is equal to 1. **b**, The number of SIPs was divided by the total number of anchor bins (both SIPs and non-SIPs) associated with genes with the first, second, third, and fourth highest expression among all four cell types ($n = 13,996$ anchor bins with promoters). Fold enrichment was calculated relative to the group with the lowest expression among all four cell types. **c**, Scatter plot showing the enrichment and number of observed copies for TE families in SIPGs for excitatory neurons. TE families occupying more than 1% of the genome are coloured. **d**, Scatter plot showing the enrichment of ERVL-MaLR TEs and number of distal interacting regions for each of the SIPGs in excitatory neurons ($n = 638$ SIPGs). The 16 SIPGs with significant enrichment ($P < 0.01$, one-tailed hypergeometric test) and 40 or more distal interacting regions are highlighted. **e**, Scatter plot showing the enrichment of TF motifs in ERVL-MaLR TEs for the 16 SIPGs highlighted in **d**. Enrichment P values are from HOMER. **f**, Scatter plot showing the enrichment of ZNF143 motifs in ERVL-MaLR TEs for the 16 SIPGs highlighted in **d** (Poisson distribution; Methods). **g**, Interactions between the *ADRA2A* promoter and 12 distal interacting regions containing ERVL-MaLR TE-localized ZNF143 motifs. **h**, Proposed mechanism for the contribution of TEs to SIP formation. **i**, *ADRA2A* expression was significantly downregulated for three out of seven regions relative to control sgRNAs ($P < 0.05$, two-tailed two-sample t -test; $n = 3$ for all regions except region III, which has $n = 2$). Data are mean and s.e.m.

(lincRNAs) including *LINC00461* and *LINC01551* are annotated as SIPs, consistent with their expression in the developing cortex¹⁶. In general, SIPs are enriched in cell types with the highest expression of their linked genes, supporting their putative roles in lineage specification (Fig. 3b). Moreover, super-enhancers and DNA methylation valleys¹⁷ are enriched at SIPs (Extended Data Fig. 5e, f). Finally, SIPs based on

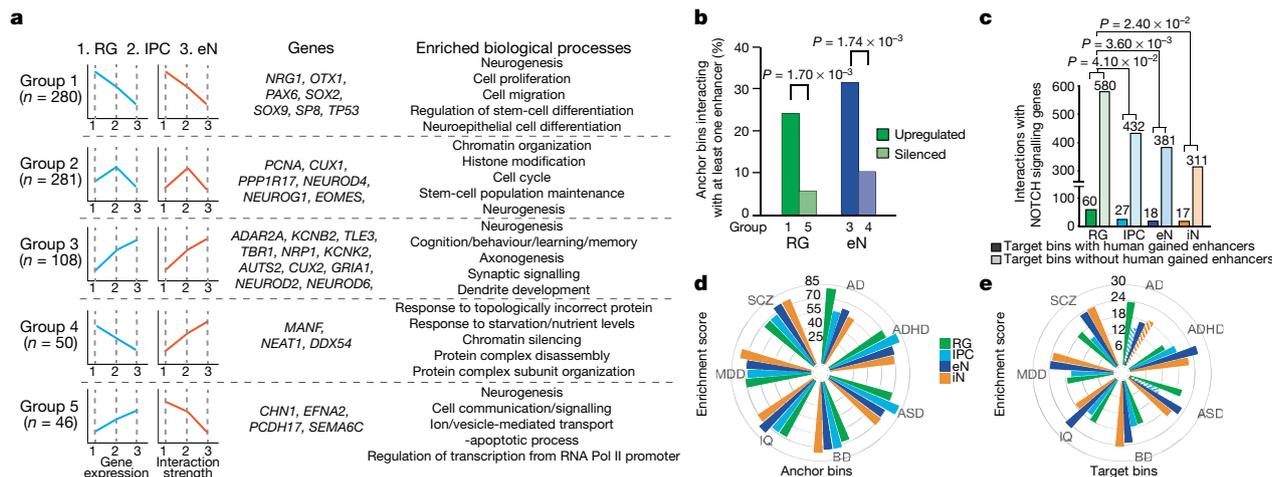


Fig. 4 | Features of cortical development and partitioning SNP heritability for complex disorders and traits. **a**, Genes categorized on the basis of their gene expression and chromatin interactivity from RG to excitatory neurons. Groups 1 to 5 represent RG-upregulated, IPC-upregulated, excitatory neuron-upregulated, excitatory neuron-silenced, and RG-silenced genes, respectively. Representative genes and biological processes are shown for each group. **b**, Groups 1 (75 out of 312 bins) and 3 (40 out of 127 bins) are enriched for interactions with enhancers relative to groups 4 (6 out of 58 bins) and 5 (3 out of 52 bins). *P* values determined by two-tailed chi-squared test. Only bins with at least one interaction were considered. **c**, Interaction counts from Notch

signalling genes to regions with and without HGEs in each cell type. *P* values determined by two-tailed chi-squared test. We observed 2,541, 1,854, 1,869, and 1,610 interactions with HGEs in RG, IPCs, excitatory neurons, and interneurons, respectively. **d, e**, LDSC enrichment scores for each disease and cell type, stratified by PLAC-seq anchor and target bins. Non-significant enrichment scores are shown as striped bars. AD, Alzheimer’s disease³⁶; ADHD, attention deficit hyperactivity disorder³⁷; ASD, autism spectrum disorder³⁸; BD, bipolar disorder³⁹; IQ, intelligence quotient⁴⁰; MDD, major depressive disorder⁴¹; SCZ, schizophrenia⁴².

Transposable elements in SIP formation

promoter capture Hi-C data in neutrophils, naive CD4⁺ T cells, monocytes, megakaryocytes, and erythroblasts¹⁸ are analogously enriched for cell-type-specific genes compared with shared genes (Extended Data Fig. 5g), indicating that SIPs present a generalized mechanism for maintaining the expression of key genes underlying cellular identity and function.

To explore mechanisms that underlie SIP formation, we evaluated the contributions of transposable elements (TEs), which are known to influence 3D chromatin architecture and propagate regulatory elements^{19–21}. We analysed the enrichment of TEs at the class, family, and subfamily levels in sequences defined by SIPs and their distal interacting regions, termed super-interactive promoter groups (SIPGs) (Fig. 3c, Extended Data Fig. 6a–c). We first observe that ERVL-MaLR TEs are enriched in SIPGs across all four cell types. We identify 16 SIPGs in excitatory neurons that exhibit significant enrichment for ERVL-MaLRs and have 40 or more distal interacting regions (*P* < 0.01, one-tailed hypergeometric test) (Fig. 3d). Transcription-factor motif enrichment analysis for ERVL-MaLRs reveals the highest enrichment for ZNF143, an architectural protein that mediates physical chromatin looping between promoters and distal regulatory elements²² (Fig. 3e). ERVL-MaLR TE subfamilies have also been linked to ZNF143 binding in 3T3 and HeLa cells²³. We find that ZNF143 motifs are broadly enriched in ERVL-MaLRs, SIPGs, and ERVL-MaLR TEs in SIPGs (Extended Data Fig. 6d–f). The *ADRA2A* SIPG is characterized by the highest enrichment of ERVL-MaLR TE-localized ZNF143 motifs (*P* = 5.1 × 10⁻⁴, one-sided Poisson test) (Fig. 3f) and is associated with elevated *ADRA2A* expression in excitatory neurons (Extended Data Fig. 6g). It spans 42 distal interacting regions, 25 of which contain ERVL-MaLRs, and 12 of which contain ERVL-MaLR-localized ZNF143 motifs (Fig. 3g, Extended Data Fig. 6h). These ZNF143 motifs can be aligned to the consensus sequences of their corresponding ERVL-MaLR TE subfamilies (Extended Data Fig. 6i, j), supporting a model in which ZNF143 motifs are coordinately expanded by the insertion of ERVL-MaLR TEs, promoting increased binding site redundancy and strengthened assembly of the *ADRA2A* regulatory

unit (Fig. 3h). CRISPR interference (CRISPRi) targeting of ERVL-MaLR TE-localized ZNF143 motifs in the *ADRA2A* SIPG resulted in significant downregulation of *ADRA2A* expression for three out of seven regions in excitatory neurons (*P* < 0.05, two-tailed two-sample *t*-test) (Fig. 3i), supporting the role of TEs in mediating the formation of higher-order chromatin features including SIPs²⁴.

Developmental trajectories in the human cortex

Because RG, IPCs, and excitatory neurons represent a developmental trajectory from dorsal cortical progenitors to mature functional neurons, we grouped genes on the basis of their expression and chromatin interactivity along this axis and identified genes linked to cell-type-specific processes in RG, IPCs, and excitatory neurons (groups 1 to 3) (Fig. 4a, Extended Data Fig. 7a, Supplementary Table 7). We similarly identified genes with anti-correlated gene expression and chromatin interactivity from RG to excitatory neurons (groups 4 and 5), which represent excitatory neuron-silenced and RG-silenced genes, respectively. Excitatory neuron-silenced genes are enriched for biological processes linked to chromatin remodelling and epigenetic regulation, whereas RG-silenced genes are enriched for excitatory neuron-specific signatures. Furthermore, genes in these groups are depleted for interactions with enhancers annotated using ChromHMM in the germinal matrix²⁵ while exhibiting enrichment for interactions with transcription factors containing domains associated with transcriptional repression (Fig. 4b, Extended Data Fig. 7b, Supplementary Table 8). Our results demonstrate that cell-type-specific 3D epigenomes can identify distinct modes of epigenetic regulation during development.

Human-specific aspects of cortical development

Human corticogenesis is markedly distinct from other mammals, driven largely by the increased diversity and proliferative capacity of cortical progenitors²⁶. Notch signalling genes in particular have been implicated in the clonal expansion of RG^{27,28}. Here, RG are enriched relative to other cell types for interactions involving Notch signalling genes²⁹ (Fig. 4c). Interactions in RG also target a significantly higher

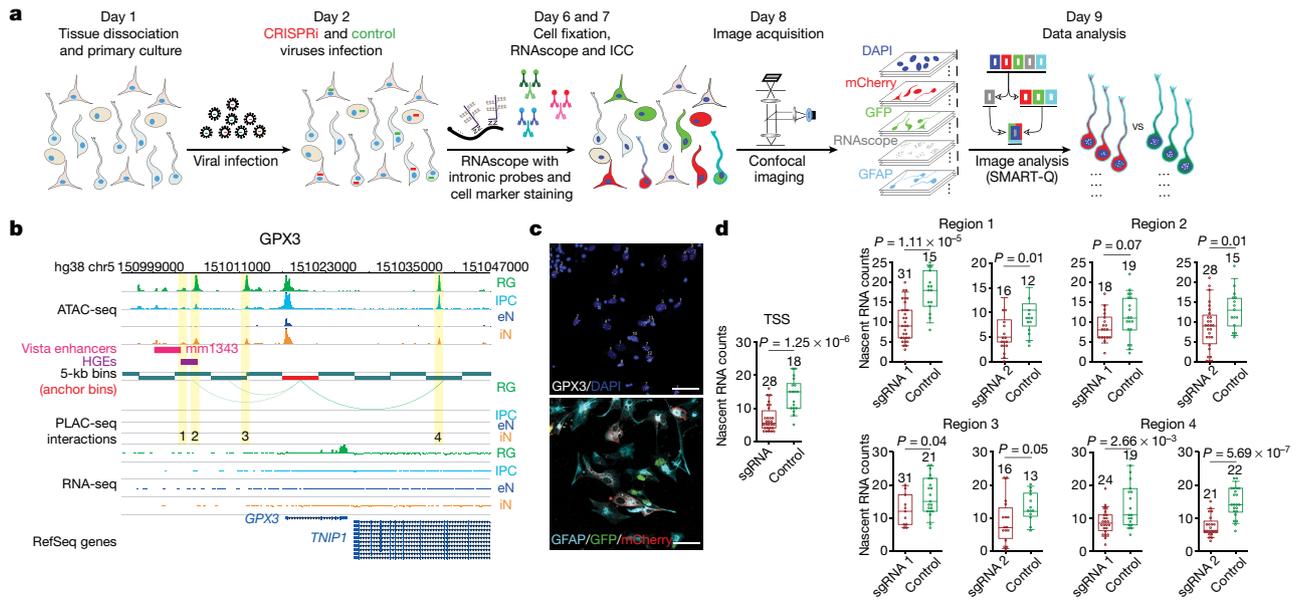


Fig. 5 | Validation of cell-type-specific distal regulatory elements using CRISPRview. **a**, Schematic of the CRISPRview workflow. Image analysis was performed using the SMART-Q pipeline. ICC, immunocytochemistry. **b**, Interactions between the *GPX3* promoter and distal interacting regions containing open chromatin peaks targeted for silencing are highlighted. Notably, region 1 overlaps both an HGE and a Vista enhancer element (mm1343), supporting its function as a putative enhancer. **c**, Representative

images show staining for intronic RNAscope probes (white), DAPI (blue), GFAP (light blue), GFP (green), and mCherry (red). Scale bar, 50 μ m. **d**, Box plots show results for experimental (red) and control (green) sgRNA-treated cells for each region. *P* values determined by two-tailed two-sample *t*-test. The median, upper and lower quartiles, and 10% to 90% range are indicated. Open circles represent single cells. Sample sizes are indicated above each box plot.

proportion of human-gained enhancers (HGEs)³⁰. This suggests that epigenetic modifications that surround Notch signalling genes in RG contribute to neurological differences between humans and other species. Additional biological processes that exhibit enrichment for interactions with HGEs include forebrain neuron fate commitment in RG, neuroblast proliferation in IPCs, forebrain neuron development in excitatory neurons, and GABA (γ -aminobutyric acid)-ergic interneuron development (Supplementary Table 9). We provide detailed annotations of genes interacting with HGEs and enhancer elements that have been validated *in vivo*³¹ in Supplementary Table 10.

Partitioning SNP heritability for complex disorders

Chromatin interactions present a unique resource for linking genome-wide association study (GWAS) variants to their target genes (Extended Data Fig. 7c, d, Supplementary Table 11). Furthermore, expression quantitative trait loci (eQTLs) from both fetal³² and adult³³ brains are enriched at chromatin interactions (Extended Data Fig. 8a–c). We leveraged linkage disequilibrium score regression (LDSC)^{34,35} to partition single nucleotide polymorphism (SNP) heritability for seven complex neuropsychiatric disorders and traits: Alzheimer’s disease³⁶, attention deficit hyperactivity disorder³⁷, autism spectrum disorder³⁸, bipolar disorder³⁹, intelligence quotient⁴⁰, major depressive disorder⁴¹ and schizophrenia⁴². First, conditioned on a baseline model⁴³, PLAC-seq anchor and target bins exhibit significant enrichment for all of the traits we tested, except for Alzheimer’s disease and autism spectrum disorder (Extended Data Fig. 8d). Anchor and target bins are also more informative than distal open chromatin peaks and cell-type-specific genes (Extended Data Fig. 8e, f), which is attributable to the utility of chromatin interactions for linking genes to distal regulatory sequences. Next, we used a joint model incorporating all four cell types to investigate cell-type-specific patterns of SNP heritability enrichment (Fig. 4d, e). Target bins exhibit more variability than anchor bins in their enrichment scores, reflecting the increased cell type specificity of distal regulatory elements compared to promoters. Furthermore, excitatory neurons

and interneurons exhibit higher enrichment scores at target bins relative to RG and IPCs, which suggests the increased relevance of neuronal cell types for these traits. We used H-MAGMA⁴⁴ to identify biological processes that are enriched for genes interacting with non-coding variants (Extended Data Fig. 9, Supplementary Table 12). Our results recapitulate the roles of lipoprotein metabolism and transport in Alzheimer’s disease pathophysiology in RG⁴⁵. Meanwhile, IPCs and excitatory neurons are enriched across all diseases for interactions linking SNPs to genes associated with neural precursor cell proliferation, axon guidance and axonogenesis. Finally, our results for schizophrenia align with extensive evidence that the disruption of chromatin regulators is a major contributor to disease risk^{9,46}.

Cell-type-specific validation in primary cells

The validation of distal regulatory elements in primary cells has been challenging, with most experiments performed using cell lines or induced pluripotent stem (iPS) cell-derived cells. A major obstacle lies in the robust detection of transcriptional changes in complex, heterogeneous samples. We developed CRISPRview to validate cell-type-specific distal regulatory elements in single cells (Fig. 5a). Specifically, primary cultures of germinal zone or cortical plate samples were first infected with lentivirus expressing mCherry, dCas9-KRAB, and single guide RNAs (sgRNAs) that target open chromatin peaks that interact with a gene of interest along with lentivirus expressing GFP, dCas9-KRAB, and control sgRNAs. The cells were then fixed and stained using antibodies for mCherry, GFP, cell-type-specific markers, DAPI, and intronic RNAscope probes that target the gene of interest. Finally, we leveraged SMART-Q⁴⁷ to compare the number of nascent RNA transcripts between experimental and control sgRNA-treated cells. We validated four regions that interact with the *GPX3* promoter, all of which exhibit significant downregulation in terms of *GPX3* expression upon silencing (Fig. 5b–d). Meanwhile, silencing three regions that interact with the *IDH1* promoter in RG and excitatory neurons results in the significant downregulation of *IDH1* expression in the respective cell types

(Extended Data Fig. 10a, b). Finally, we characterized two additional RG-specific loci in *TNC* and *HES1*, both of which are annotated as SIPs (Extended Data Fig. 10c–h). The observation of small but significant changes in gene expression supports the hypothesis that multiple interactions frequently work together to titrate the expression of key genes underlying cellular identity and function.

Discussion

Single-cell RNA sequencing studies have highlighted the heterogeneity of the developing human cortex. Despite marked differences in lineage and maturation state, many of the cell types share intriguing similarities in their transcriptional landscapes. For example, interneurons express genes for transcription factors that are typically associated with RG proliferation, including SOX2, as well as with excitatory neuron differentiation, including ASCL1 and NPAS3¹. By isolating and characterizing specific cell types, we are able to distinguish nuanced regulatory programs that drive cell-type-specific differences during human corticogenesis. We identify SIPs that are enriched for key lineage-specific genes and represent distinct chromatin features from A/B compartments⁴⁸, topologically associated domains⁴⁹, frequently interacting regions⁵⁰, and highly interacting regions⁵¹. Furthermore, we uncover a mechanism in which TEs propagate binding sites for architectural proteins such as ZNF143, facilitating the formation of multi-interaction clusters that function to sustain transcription. Lastly, by developing CRISPRview, we achieve several emergent advantages for validating distal regulatory elements in primary cells. First, we are able to focus our analysis on specific cell types, circumventing averaging effects associated with bulk measurements in complex samples. Next, we are able to compare cells infected with experimental or control sgRNAs in the same population. Finally, we achieve enhanced sensitivity and statistical power based on the detection of nascent RNA transcripts in single cells. Future experiments using CRISPRview in live tissues should continue to reveal regulatory relationships in a manner that is truly representative of the complex *in vivo* environment.

Online content

Any methods, additional references, Nature Research 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/s41586-020-2825-4>.

1. Nowakowski, T. J. et al. Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science* **358**, 1318–1323 (2017).
2. Zhong, S. et al. A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. *Nature* **555**, 524–528 (2018).
3. Hansen, D. V., Lui, J. H., Parker, P. R. & Kriegstein, A. R. Neurogenic radial glia in the outer subventricular zone of human neocortex. *Nature* **464**, 554–561 (2010).
4. Pontius, A., Kowalczyk, T., Englund, C. & Hevner, R. F. Role of intermediate progenitor cells in cerebral cortex development. *Dev. Neurosci.* **30**, 24–32 (2008).
5. Anderson, S., Mione, M., Yun, K. & Rubenstein, J. L. Differential origins of neocortical projection and local circuit neurons: role of *Dlx* genes in neocortical interneuronogenesis. *Cereb. Cortex* **9**, 646–654 (1999).
6. Zheng, H. & Xie, W. The role of 3D genome organization in development and cell differentiation. *Nat. Rev. Mol. Cell Biol.* **20**, 535–550 (2019).
7. Li, Y., Hu, M. & Shen, Y. Gene regulation in the 3D genome. *Hum. Mol. Genet.* **27**(R2), R228–R233 (2018).
8. Schoenfelder, S. & Fraser, P. Long-range enhancer-promoter contacts in gene expression control. *Nat. Rev. Genet.* **20**, 437–455 (2019).
9. Won, H. et al. Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature* **538**, 523–527 (2016).
10. Thomsen, E. R. et al. Fixed single-cell transcriptomic characterization of human radial glial diversity. *Nat. Methods* **13**, 87–93 (2016).
11. Fang, R. et al. Mapping of long-range chromatin interactions by proximity ligation-assisted ChIP-seq. *Cell Res.* **26**, 1345–1348 (2016).
12. Juric, I. et al. MAPS: model-based analysis of long-range chromatin interactions from PLAC-seq and HiChIP experiments. *PLoS Comput. Biol.* **15**, e1006982 (2019).
13. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).

14. Englund, C. et al. Pax6, Tbr2, and Tbr1 are expressed sequentially by radial glia, intermediate progenitor cells, and postmitotic neurons in developing neocortex. *J. Neurosci.* **25**, 247–251 (2005).
15. Lim, L., Mi, D., Llorca, A. & Marin, O. Development and functional diversification of cortical interneurons. *Neuron* **100**, 294–313 (2018).
16. Liu, S. J. et al. Single-cell analysis of long non-coding RNAs in the developing human neocortex. *Genome Biol.* **17**, 67 (2016).
17. Luo, C. et al. Cerebral organoids recapitulate epigenomic signatures of the human fetal brain. *Cell Rep.* **17**, 3369–3384 (2016).
18. Javierre, B. M. et al. Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. *Cell* **167**, 1369–1384 (2016).
19. Choudhary, M. N. et al. Co-opted transposons help perpetuate conserved higher-order chromosomal structures. *Genome Biol.* **21**, 16 (2020).
20. Feschotte, C. Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.* **9**, 397–405 (2008).
21. Zhang, Y. et al. Transcriptionally active HERV-H retrotransposons demarcate topologically associating domains in human pluripotent stem cells. *Nat. Genet.* **51**, 1380–1388 (2019).
22. Bailey, S. D. et al. ZNF143 provides sequence specificity to secure chromatin interactions at gene promoters. *Nat. Commun.* **2**, 6186 (2015).
23. Ngondo-Mbongo, R. P., Myslinski, E., Aster, J. C. & Carbon, P. Modulation of gene expression via overlapping binding sites exerted by ZNF143, Notch1 and THAP11. *Nucleic Acids Res.* **41**, 4000–4014 (2013).
24. Sundaram, V. & Wang, T. Transposable element mediated innovation in gene regulatory landscapes of cells: re-visiting the “gene-battery” model. *BioEssays* **40**, (2018).
25. Davis, C. A. et al. The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res.* **46**(D1), D794–D801 (2018).
26. Miller, D. J., Bhaduri, A., Sestan, N. & Kriegstein, A. Shared and derived features of cellular diversity in the human cerebral cortex. *Curr. Opin. Neurobiol.* **56**, 117–124 (2019).
27. Suzuki, I. K. et al. Human-specific NOTCH2NL genes expand cortical neurogenesis through Delta/Notch regulation. *Cell* **173**, 1370–1384 (2018).
28. Rani, N. et al. A Primate lncRNA mediates Notch signaling during neuronal development by sequestering miRNA. *Neuron* **90**, 1174–1188 (2016).
29. Carbon, S. et al. AmiGO: online access to ontology and annotation data. *Bioinformatics* **25**, 288–289 (2009).
30. Reilly, S. K. et al. Evolutionary genomics. Evolutionary changes in promoter and enhancer activity during human corticogenesis. *Science* **347**, 1155–1159 (2015).
31. Visel, A., Minovitsky, S., Dubchak, I. & Pennacchio, L. A. VISTA Enhancer Browser—a database of tissue-specific human enhancers. *Nucleic Acids Res.* **35**, D88–D92 (2007).
32. Walker, R. L. et al. Genetic control of expression and splicing in developing human brain informs disease mechanisms. *Cell* **179**, 750–771 (2019).
33. Hoffman, G. E. et al. CommonMind Consortium provides transcriptomic and epigenomic data for schizophrenia and bipolar disorder. *Sci. Data* **6**, 180 (2019).
34. Bulik-Sullivan, B. K. et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
35. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
36. Jansen, I. E. et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer’s disease risk. *Nat. Genet.* **51**, 404–413 (2019).
37. Demontis, D. et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat. Genet.* **51**, 63–75 (2019).
38. Grove, J. et al. Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* **51**, 431–444 (2019).
39. Stahl, E. A. et al. Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat. Genet.* **51**, 793–803 (2019).
40. Savage, J. E. et al. Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat. Genet.* **50**, 912–919 (2018).
41. Howard, D. M. et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat. Neurosci.* **22**, 343–352 (2019).
42. Pardiñas, A. F. et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat. Genet.* **50**, 381–389 (2018).
43. Gazal, S. et al. Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection. *Nat. Genet.* **49**, 1421–1427 (2017).
44. Sey, N. Y. A. et al. A computational tool (H-MAGMA) for improved prediction of brain-disorder risk genes by incorporating brain chromatin interaction profiles. *Nat. Neurosci.* **23**, 583–593 (2020).
45. Andersen, O. M. & Willnow, T. E. Lipoprotein receptors in Alzheimer’s disease. *Trends Neurosci.* **29**, 687–694 (2006).
46. Akbarian, S. Epigenetic mechanisms in schizophrenia. *Dialogues Clin. Neurosci.* **16**, 405–417 (2014).
47. Yang, X. et al. SMART-Q: an integrative pipeline quantifying cell-type-specific RNA transcription. *PLoS ONE* **15**, e0228760 (2020).
48. Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
49. Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
50. Schmitt, A. D. et al. A compendium of chromatin contact maps reveals spatially active regions in the human genome. *Cell Rep.* **17**, 2042–2059 (2016).
51. Sobhy, H., Kumar, R., Lewerentz, J., Lizana, L. & Stenberg, P. Highly interacting regions of the human genome are enriched with enhancers and bound by DNA repair proteins. *Sci. Rep.* **9**, 4577 (2019).

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

Data reporting

No statistical methods were used to predetermine sample size. The experiments were not randomized, and investigators were not blinded to allocation during experiments and outcome assessment.

Ethics statement

De-identified tissue samples were collected with prior informed consent in strict observance of legal and institutional ethical regulations. All protocols were approved by the Human Gamete, Embryo, and Stem Cell Research Committee (GESCR) and Institutional Review Board (IRB) at the University of California, San Francisco.

Tissue dissociation

Tissue dissociation was performed as previously described¹. In brief, samples were first cut into small pieces in artificial cerebrospinal fluid before being added to pre-warmed papain dissociation medium (Worthington LK003150). The samples were incubated in dissociation medium for 45 min at 37 °C. Next, they were triturated, filtered through a 70- μ m nylon mesh, and centrifuged for 8 min at 300g. For individual germinal zone and cortical plate cultures, samples were first cut coronally into thin slices. As previously described, cell density drops markedly past the outer subventricular zone, enabling the clear identification of the outer filamentous zone and subplate. Samples were dissected along this boundary to separate the germinal zone from the cortical plate before dissociation.

Sample fixation

Mid-gestational human cortex samples between GW15 and GW22 were fixed in 2% paraformaldehyde prepared in PBS with gentle agitation for 10 min at room temperature. Glycine was added to a final concentration of 200 mM to quench the reactions, and the samples were centrifuged for 5 min at 4 °C and 500g. The samples were washed twice with PBS before being frozen at -80 °C for further processing.

Permeabilization and staining

The cell pellet was thawed on ice and resuspended in PBS containing 0.1% Triton X-100 for 15 min. The cells were then washed twice with PBS and resuspended in 5% BSA in PBS for staining. Staining proceeded for at least one hour with FcR Blocking Reagent (Miltenyi Biotech, 1:20 dilution), EOMES PE-Cy7 (Invitrogen, 25-4877-42, Clone WD1928, Lot 1923396, 1:10 dilution), PAX6 PE (BD Biosciences, 561552, clone O18-1330, Lot 8187686, 1:10 dilution), SOX2 PerCP-Cy5.5 (BD Biosciences, 561506, clone O38-678, Lot 8165744, 1:10 dilution), and SATB2 Alexa Fluor 647 (Abcam, ab196536, clone EPNCIR130A, Lot GR3208103-1 and GR228747-2, 1:100 dilution). After staining, the cells were centrifuged for 5 min at 500g, and the pellet was diluted into PBS. When sorting cells for RNA-seq, 1% RNasin Plus RNase Inhibitor (Promega) was added to all buffers, and acetylated BSA was used to prepare 5% BSA in PBS for staining.

FACS

AbC Total Antibody Compensation Beads (Thermo Fisher) were used to generate single-colour compensation controls before sorting. Sorting was conducted on the FACS Aria II, FACS Aria IIu, or FACS Aria Fusion instruments using a 70- μ m nozzle, and cells were collected in 5-ml tubes pre-coated with FBS. A sample of each sorted cell population was reanalysed on the same machine to assess purity. Cells were collected by centrifuging for 10 min at 500g, and the cell pellet was frozen at -80 °C for further processing. When sorting cells for RNA-seq, cells were collected in 5 ml tubes pre-coated with both FBS and RNAlater (Thermo Fisher).

Primary cell culture

Following dissociation, cells were plated onto Matrigel-coated coverslips in 48 well plates or chamber slides at a density of approximately

0.7×10^6 cells per well. All cell culture was handled in sterile conditions. The cells were infected with lentivirus the day after plating, and medium was changed every two days. Medium was composed of 96% DMEM/F-12 with GlutaMAX, 1% N-2, 1% B-27 and 1% penicillin-streptomycin. The cells were grown in 8% O₂ and 5% CO₂ and were collected four days following infection for CRISPRview. For qPCR at the *ADRA2A* locus, the cells were collected six days following infection.

PLAC-seq

PLAC-seq was performed as previously described¹¹. Between 1 million to 5 million cells were used to prepare each library. Digestion was performed using 100 U Mbol for 2 h at 37 °C, and chromatin immunoprecipitation was performed using Dynabeads M-280 sheep anti-rabbit IgG (Invitrogen 11203D) superparamagnetic beads bound with 5 μ g anti-H3K4me3 antibody (Millipore 04-745). Sequencing adapters were added during PCR amplification. Libraries were sent for paired-end sequencing on the HiSeq X Ten or NovoSeq 6000 instruments (150 bp paired-end reads). fastp was applied to trim reads to 100 bp for all downstream analysis.

MAPS

We used the MAPS pipeline to call significant H3K4me3-mediated chromatin interactions at a resolution of 5 kb on the basis of our PLAC-seq data. First, bwa mem was used to map raw reads to hg38. Unmapped reads and reads with low mapping quality were discarded, and the resulting read pairs were processed as previously described¹². To define PLAC-seq anchor bins, we took the union of peaks identified by MACS2 using the options '-nolambda-nomodel-extsize 147-call-summits -B-SPMR' and an FDR cut-off value of 0.0001 for all read pairs with interaction distance <1 kb in each cell type. Next, we classified read pairs as AND, XOR, or NOT interactions based on whether both, one, or neither of the interacting 5-kb bins overlapped anchor bins (Extended Data Fig. 3a). Because we were specifically interested in identifying long-range H3K4me3-mediated chromatin interactions, we retained only read pairs corresponding to intrachromosomal XOR and AND interactions with interaction distances between 10 kb and 1 Mb. We downsampled the number of read pairs separately for each chromosome to ensure that we started with the same number of read pairs for each cell type.

To call significant interactions, we used a zero-truncated Poisson regression-based approach to normalize systematic biases from restriction sites, GC content, sequence repetitiveness, and ChIP enrichment. We fitted models separately for AND and XOR interactions and calculated FDRs for interactions based on the expected and observed contact frequencies between interacting 5-kb bins. We grouped interactions whose ends were located within 15 kb of each other into clusters and classified all other interactions as singletons. We defined our significant H3K4me3-mediated chromatin interactions as interactions with 12 or more reads, normalized contact frequency (defined as the ratio between the observed and expected contact frequency) ≥ 2 , and FDR < 0.01 for clusters and FDR < 0.0001 for singletons. This was based on the reasoning that biologically meaningful interactions are more likely to appear in clusters, whereas singletons are more likely to represent false positives.

Reproducibility analysis

PCA was performed on the basis of the normalized contact frequencies for interacting 5-kb bins from our PLAC-seq data. We first extracted AND and XOR interactions based on cell-type-specific anchor bins for each of the replicates. Next, we applied zero-truncated Poisson regression adjusting for the same biases as the MAPS pipeline. We derived normalized contact frequencies on the basis of the ratios between the observed and expected contact frequencies for interacting 5-kb bins, with the expected contact frequencies being the fitted values from the zero-truncated Poisson regression. Normalized contact frequencies

Article

were then log-transformed and merged across all the replicates. The merged data were used to generate the PCA plots. We restricted our analysis to interacting 5-kb bins in both 300- and 600-kb windows for Extended Data Fig. 2d.

ATAC-seq

ATAC-seq was performed as previously described⁵² using the Nextera DNA Library Prep Kit (Illumina FC-121-1030). In brief, fixed cells were washed once with ice cold PBS containing 1× protease inhibitor before being resuspended in ice cold nuclei extraction buffer (10 mM Tris-HCl pH 7.5, 10 mM NaCl, 3 mM MgCl₂, 0.1% Igepal CA630, and 1× protease inhibitor) for 5 min. 50,000 cells were aliquoted, exchanged into 50 μl 1× buffer TD, and incubated with 2.5 μl TDE1 enzyme for 45 min at 37 °C with shaking. Following transposition, 150 μl reverse crosslinking solution (50 μl 1 M Tris pH 8.0, 100 μl 10% SDS, 2 μl 0.5 M EDTA, 10 μl 5 M NaCl, 800 μl water, and 2.5 μl 20 mg ml⁻¹ proteinase K) was added to each tube and incubated at 65 °C overnight. DNA was column purified, PCR amplified, and size-selected for fragments between 300 and 1,000 bp. Libraries were sent for paired-end sequencing on the NovaSeq 6000 instrument (150 bp paired-end reads). Raw reads were trimmed to 50 bp, mapped to hg38, and processed using the ENCODE pipeline (https://github.com/kundajelab/atac_dnase_pipelines) running the default settings. The optimal naive overlap peaks for each cell type were used for all downstream analyses.

RNA-seq

We extracted total RNA from the sorted cell populations using the RNastorm FFPE RNA extraction kit (Cell Data Sciences CD501) starting with 5×10^5 to 1.5×10^6 cells. The quality of the extracted RNA was checked by determining the percentage of RNA fragments with size >200 bp (DV200) from the Agilent 2100 Bioanalyzer. RNA samples with DV200 ≥ 40% were used for library construction. First, samples were depleted of ribosomal RNA using the KAPA RNA HyperPrep Kit with RiboErase (HMR KK8560). Next, we performed first and second strand synthesis, dA-tailing, and sequencing adaptor ligation. cDNA was cleaned up and sequencing adapters were added via PCR amplification. Libraries were sent for paired-end sequencing on the NovaSeq 6000 instrument (150 bp paired-end reads). Raw reads were trimmed using Trim Galore and aligned to hg38 using STAR running the standard ENCODE parameters, and transcript quantification was performed in a strand-specific manner using RSEM with the GENCODE 29 annotation. The edgeR package in R was used to calculate TMM-normalized RPKM values for each gene, and the mean values across all replicates were used for all downstream analyses.

GO enrichment analysis

Protein coding and non-coding RNA genes participating in cell-type-specific XOR interactions were used for GO enrichment analysis. Only interactions with open chromatin peaks overlapping promoters (defined as the 1-kb region centred around the transcription start site of a gene) in their anchor bins and distal open chromatin peaks (defined as open chromatin peaks not overlapping promoters) in their target bins were used. A minimum RPKM of 0.5 was used to retain only genes that were expressed, and the resulting genes were input into DAVID 6.8 running functional annotation clustering using the 'GOTERM_BP_ALL' ontology. Group enrichment scores based on the geometric mean of EASE scores for terms in each group were reported. To report enriched biological processes for genes interacting with non-coding SNPs, we assigned non-coding SNPs for each disorder and trait to genes using interactions with the 5-kb bins containing their promoters. Next, we ran H-MAGMA using our annotations to generate ranked lists of gene-level association statistics which were used to perform functional enrichment analysis using the gprofiler2 package in R⁵³: `gost(ranked.list, organism = "hsapiens", ordered_query = T, significant = F, correction_method = "fdr", sources = "GO:BP")`.

Transcription factor motif enrichment analysis

We used 200-bp windows centred around open chromatin peaks participating in cell-type-specific XOR interactions for transcription-factor motif enrichment analysis using HOMER. We used the complete set of vertebrate motifs from the JASPAR database, specifying the '-float' option to adjust the degeneracy threshold, and the entire genome was used as the background. The binomial distribution was used to calculate *P* values. For the analysis of co-expression modules in the developing human cortex, we downloaded co-expression modules as previously published⁴. Specifically, we used the 'all' network set for all four cell types, as well as network sets matched to individual cell types as follows: 'page.rg' for RG, 'page.ipc' for IPCs, 'page.n' for excitatory neurons, and 'vage.in' for interneurons. This was to capture biological variation both between and within cell types, respectively. We used HOMER to perform transcription factor motif enrichment analysis for the set of open chromatin peaks interacting with promoters of genes assigned to each co-expression module. For ranking transcription factors according to the number of co-expression modules they were enriched for in each network set and cell type, an FDR threshold of 0.05 was applied.

SIPs

We used an approach similar to calling super-enhancers⁵⁴ to annotate SIPs in each cell type. For each anchor bin, we calculated the cumulative interaction score, defined as the sum of the $-\log_{10}$ FDR for interactions overlapping each anchor bin. We used this metric as it accounts for noise and is directly associated with the interaction strength in PLAC-seq data. Next, we prepared plots of ranked cumulative interaction scores for anchor bins in each cell type and defined SIPs to be anchor bins located past the point in each curve at which the slope is equal to 1.

Cell-type-specific versus shared genes

We classified each gene as cell-type-specific or shared according to its Shannon entropy score across all four cell types. Specifically, for each gene, we calculated its relative expression value in each cell type, defined as its RPKM in that cell type divided by the sum of its RPKMs across all four cell types. Next, we calculated the Shannon entropy score for each gene on the basis of its relative expression values across all four cell types. We classified a gene as specific for a cell type if met the following conditions: its Shannon entropy score was <0.01, its RPKM was >1 in that cell type, and its RPKM in that cell type was the highest across all four cell types. All other genes with RPKM >1 were classified as shared.

TE enrichment in SIPGs

TE enrichment in SIPGs was evaluated as follows. The foreground enrichment was defined as the number of TEs at the class, family, or subfamily levels overlapping SIPGs in each cell type. The background enrichment was defined as the number of TEs overlapping all interacting 5-kb bins (both SIPGs and non-SIPGs). At least 50% of a TE had to overlap a 5-kb bin for it to be considered overlapping. The overall enrichment was defined as the foreground enrichment divided by the background enrichment multiplied by the proportion of interacting 5-kb bins that were assigned to SIPGs.

For the enrichment of SIPGs for ERVL-MaLR TEs, the foreground enrichment for each SIPG was defined as the number of distal interacting regions containing one or more ERVL-MaLR TEs for that SIPG. The background enrichment for each SIPG was defined as the number of randomly shuffled distal interacting regions containing one or more ERVL-MaLR TEs for that SIPG. We computed the background enrichment over 100 permutations. The overall enrichment was defined as the foreground enrichment divided by the background enrichment.

The significance for each SIPG was calculated using the hypergeometric distribution as follows:

$$P = \frac{\binom{m}{q} \binom{n}{k-q}}{\binom{m+n}{k}}$$

in which q is the number of distal interacting regions containing one or more ERVL-MaLR TEs for that SIPG, m is the number of 5-kb bins containing one or more ERVL-MaLR TEs on the same chromosome, n is the number of 5-kb bins containing no ERVL-MaLR TEs on the same chromosome, and k is the size of the SIPG.

ZNF143 motif enrichment

For the enrichment of SIPGs for ERVL-MaLR TE-localized ZNF143 motifs, the foreground enrichment for each SIPG was defined as the number of ERVL-MaLR TE-localized ZNF143 motifs in its distal interacting regions. FIMO⁵⁵ was used to detect ZNF143 motifs within ERVL-MaLR TEs. The background enrichment was defined as the total number of ZNF143 motifs in the SIPG. The overall enrichment was defined as the foreground enrichment divided by the background enrichment multiplied by the proportion of the SIPG that is occupied by ERVL-MaLR TEs. The significance for each SIPG was calculated using a Poisson distribution where the number of events (k) is the foreground enrichment and the rate parameter (λ) is the background enrichment multiplied by the proportion of the SIPG that is occupied by ERVL-MaLR TEs.

For evaluating the genome-wide enrichment of ZNF143 motifs in ERVL-MaLR and THE1C TEs, we first used FIMO to scan all ERVL-MaLR and THE1C TEs for instances of ZNF143 motifs. As a background, we scanned 100 sets of chromosome- and length-matched, non-overlapping sequences randomly sampled, avoiding gaps and blacklisted regions in the human genome. We used a similar approach to evaluate the enrichment of ZNF143 motifs in ERVL-MaLR TEs in SIPGs. For evaluating the enrichment of ZNF143 motifs in SIPGs, we compared the mean numbers of ZNF143 motifs per 5-kb bin for distal interacting regions across all SIPGs to 100 sets of chromosome- and length-matched, non-overlapping sequences randomly sampled, avoiding gaps and blacklisted regions in the human genome. For comparing the distributions of the mean numbers of ZNF143 motifs per 5-kb bin for actual versus shuffled SIPGs, we sampled distal interacting regions for each SIPG 100 times on the same chromosome in a non-overlapping manner.

Target gene annotation for enhancers and GWAS SNPs

To determine whether a human-gained enhancer, Vista enhancer element, or GWAS SNP interacted with a gene, we determined whether any of its promoters participated in interactions with the element of interest on the other end. All human-gained enhancers and Vista enhancer elements were expanded to a minimum width of 5 kb, and all GWAS SNPs were expanded to a minimum width of 1 kb to account for potential functional sequences around each element. Furthermore, we determined the proportion of GWAS SNPs interacting with their nearest and more distal genes, except when all the promoters for the nearest gene fell within the same 5-kb bin as the GWAS SNP and could not be resolved for interactions.

Partitioning SNP heritability for complex disorders and traits

We leveraged LDSC to partition SNP heritability separately for each complex neuropsychiatric disorder and trait on the basis of joint models incorporating PLAC-seq anchor or target bins across all cell types. We also ran LDSC using a baseline model⁴³ consisting of coding, UTR, promoter, and intron regions, histone marks, DNase I hypersensitive sites, ChromHMM/Segway predictions, regions that are conserved in mammals, super-enhancers, FANTOM5 enhancers, and linkage disequilibrium-related annotations (recombination rate, nucleotide

diversity CpG content, and so on) that are not specific to any cell type. This informs us whether our epigenomic annotations for a given cell type are informative for SNP heritability enrichment compared to a comprehensive set of genomic features that has been widely adopted in the field. To compare different epigenomic annotations for each cell type, we used both distal open chromatin peaks and 100-kb windows around the transcription start and end sites of cell-type-specific genes according to their Shannon entropy scores and RPKM >1.

Validation of ERVL-MaLR-localized ZNF143 motifs

CRISPRi and quantitative PCR with reverse transcription (qRT-PCR) were used to validate ERVL-MaLR TE-localized ZNF143 motifs at distal interacting regions in the *ADRA2A* SIPG. Of the 12 distal interacting regions containing ERVL-MaLR TE-localized ZNF143 motifs, we were able to design sgRNAs to target ZNF143 motifs overlapping open chromatin peaks for 7 of the regions. ZNF143 motifs were extended by 100 bp in both directions for designing sgRNAs. To maximize CRISPRi efficiency, we designed two sgRNAs for each region and cloned them into the dual expression cassette in the CRISPRi vector as described for CREST-seq⁵⁶. sgRNA sequences were confirmed by Sanger sequencing and packaged into lentivirus. Primary cell cultures enriched for excitatory neurons on the basis of SATB2 staining were infected with lentivirus for 24 h, and mRNA was extracted on day 7. qRT-PCR was used to quantify *ADRA2A* expression using the following primers: TCGTC ATCATCGCCGTGTTTC (forward) and AAGCCTTGCCGAAGTACCAG (reverse). All sgRNA sequences used for validation can be found in Supplementary Table 13.

Validation of distal interacting regions using CRISPRview

The CRISPRi vector was modified from the Mosaic-seq⁵⁷ and CROP-seq vectors⁵⁸. The hU6-sgRNA expression cassette from the CROPseq-Guide-Puro vector (Addgene 86708) was cloned and inserted downstream of the WPRE element in the Lenti-dCas9-KRAB-blast vector (Addgene 89567). The blasticidin resistance gene was replaced with either mCherry or EGFP. sgRNAs targeting open chromatin peaks in distal interacting regions were designed using CHOPCHOP⁵⁹. Single-stranded DNA was annealed and ligated into the CRISPRi vector at the BsmBI cutting locus. Single clones were picked following transformation, and the sgRNA sequences were confirmed by Sanger sequencing. For lentiviral packaging, the CRISPRi vector, pMD2.G (Addgene 12259), and psPAX (Addgene 12260) were transformed into 293T cells using PolyJet (SignaGen Laboratories SL100688) according to the manufacturer's instructions. Virus-containing medium was collected three times over 16 to 20 h and concentrated using Amicon 10K columns. All lentivirus was immediately stored at -80 °C. Primary cell cultures were infected with virus (MOI < 1) 24 h after plating, and cells were fixed with 4% PFA four days after infection for FISH and immunostaining. All sgRNA sequences used for validation can be found in Supplementary Table 13.

FISH experiments were performed using the RNAScope Multiplex Fluorescent V2 Assay kit (ACDBio 323100). Probes targeting intronic regions for *GPX3* (ACDBio 572341), *IDH1* (ACDBio 832031), *TNC* (ACDBio 572361), and *HES1* (ACDBio 560881) were custom-designed, synthesized, and labelled with TSA Cyanine 5 (Perkin Elmer NEL705A001KT, 1:1,000 dilution). Fixed cells were pre-treated with hydrogen peroxide for 10 min and protease III for 15 min, and probes were hybridized and amplified according to the manufacturer's instructions. Slides were washed with PBS before blocking with 5% donkey serum in PBS for 30 min at room temperature. Next, slides were incubated with primary antibodies against mCherry (Abcam ab205402, 1:200), GFP (Abcam ab1218, 1:500), and GFAP (Abcam ab7260, 1:400) for RG or SATB2 (Abcam ab92446, 1:300) for excitatory neurons overnight at 4 °C, followed by incubation with Alexa Fluor 488 donkey anti-mouse IgG (Thermo Fisher Scientific A21202, 1:800), Alexa-546 nm donkey anti-rabbit IgG (Thermo Fisher Scientific A10040, 1:500), and Alexa-594 nm goat anti-chicken IgG (Thermo Fisher Scientific A11042, 1:500) for 1 h

Article

at room temperature. Three-dimensional confocal microscopy images were captured using a Leica TCS SP8 with a 40× oil-immersion objective lens (NA 1.30). The z-step size was 0.4 μm. For five-colour multiplexed imaging, three sequential scans were performed to avoid overlapping spectra. The excitation lasers were 405 nm and 594 nm, 488 nm and 633 nm, and 561 nm. All images were obtained using the same acquisition settings. For FISH analysis, we developed a Python-based pipeline called Single-Molecule Automatic RNA Transcription Quantification (SMART-Q) for quantifying nascent RNA transcripts in single cells. In brief, the RNAscope channel was first filtered and fitted in three dimensions using a Gaussian model. Next, segmentation was performed in two dimensions on the DAPI channel to ascertain the location of each nucleus. Finally, segmentation was performed on the remaining channels to identify experimental and control sgRNA-infected RG or excitatory neurons for nascent RNA transcript quantification.

Reporting summary

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

Data availability

All datasets used in this study (PLAC-seq, ATAC-seq, RNA-seq) are available at the Neuroscience Multi-Omic Archive (NeMO Archive) under controlled access. Chromatin interactions, open chromatin peaks, and gene expression profiles for each cell type can be downloaded from the NeMO Archive using the following link: <https://assets.nemoarchive.org/dat-uoqy8b>. Cell-type-specific 3D epigenomes can be visualized on the WashU Epigenome Browser using the datahub at the following link: http://epigenomegateway.wustl.edu/browser/?genome=hg38&position=chr17:72918238-73349675&hub=https://shen-msong.s3-us-west-1.amazonaws.com/hfb_submission/hfb_datahub.json.

Code availability

All of the software used in this study are listed in the Reporting Summary along with their versions.

52. Corces, M. R. et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).
53. Reimand, J., Kull, M., Peterson, H., Hansen, J. & Vilo, J. g:Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Res.* **35**, W193–W200 (2007).
54. Hnisz, D. et al. Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).
55. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).
56. Diao, Y. et al. A tiling-deletion-based genetic screen for cis-regulatory element identification in mammalian cells. *Nat. Methods* **14**, 629–635 (2017).
57. Xie, S., Duan, J., Li, B., Zhou, P. & Hon, G. C. Multiplexed engineering and analysis of combinatorial enhancer activity in single cells. *Mol Cell* **66**, 285–299 (2017).
58. Datlinger, P. et al. Pooled CRISPR screening with single-cell transcriptome readout. *Nat. Methods* **14**, 297–301 (2017).
59. Labun, K. et al. CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Res.* **47**, W171–W174 (2019).

Acknowledgements This work was supported by the UCSF Weill Institute for Neuroscience Innovation Award (to Y.S. and A.R.K.), the National Institutes of Health (NIH) grants R01AG057497, R01EY027789, and UM1HG009402 (to Y.S.) and R35NS097305 (to A.R.K.), the Hillblom Foundation, and the American Federation for Aging Research New Investigator Award in Alzheimer's Disease (to Y.S.). This work was also supported by the NIH grants R01HL129132, U544HD079124, and R01MH106611 (to Y.L.), R01HG007175, U24ES026699, and U01HG009391 (to T.W.), and the American Cancer Society grant RSG-14-049-01-DMC (to T.W.). M.S. is supported by T32GM007175. M.P. is supported by the National Science Foundation Graduate Research Fellowship Program grant 1650113. U.C.E. is supported by 5T32GM007618-42. This work was made possible in part by the NIH grants P30EY002162 to the UCSF Core Grant for Vision Research, P30DK063720, and S101S10D021822-01 to the UCSF Parnassus Flow Cytometry Core.

Author contributions Y.S., A.A.P., and A.R.K. conceived the study. Y.S., M.H., A.A.P., and A.R.K. supervised the study. M.S., M.-P.P., X.Y., I.R.J., X.C., U.C.E., L.M., and J.L. performed experiments. M.S., A.A., S.B., J.D.R., B.L., I.R.J., and M.H. performed computational analysis. C.F. and M.N.K.C. performed TE analysis under the supervision of T.W. J.W., and W.L. performed SNP heritability enrichment analysis under the supervision of Y.L. M.S., M.-P.P., X.Y., and Y.S. analysed and interpreted the data. M.S., M.-P.P., X.Y., M.H., and Y.S. prepared the manuscript with input from all other authors.

Competing interests The authors declare no competing interests.

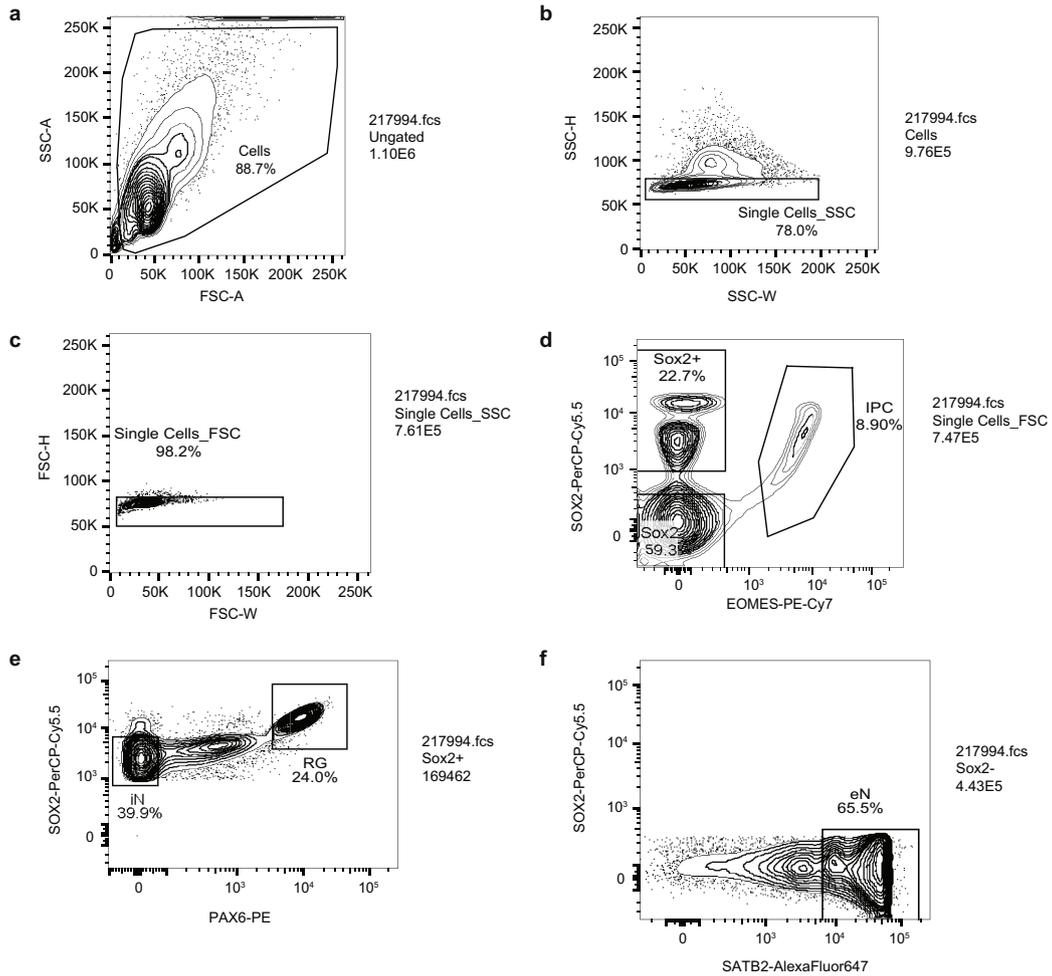
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-2825-4>.

Correspondence and requests for materials should be addressed to M.H., A.R.K. or Y.S.

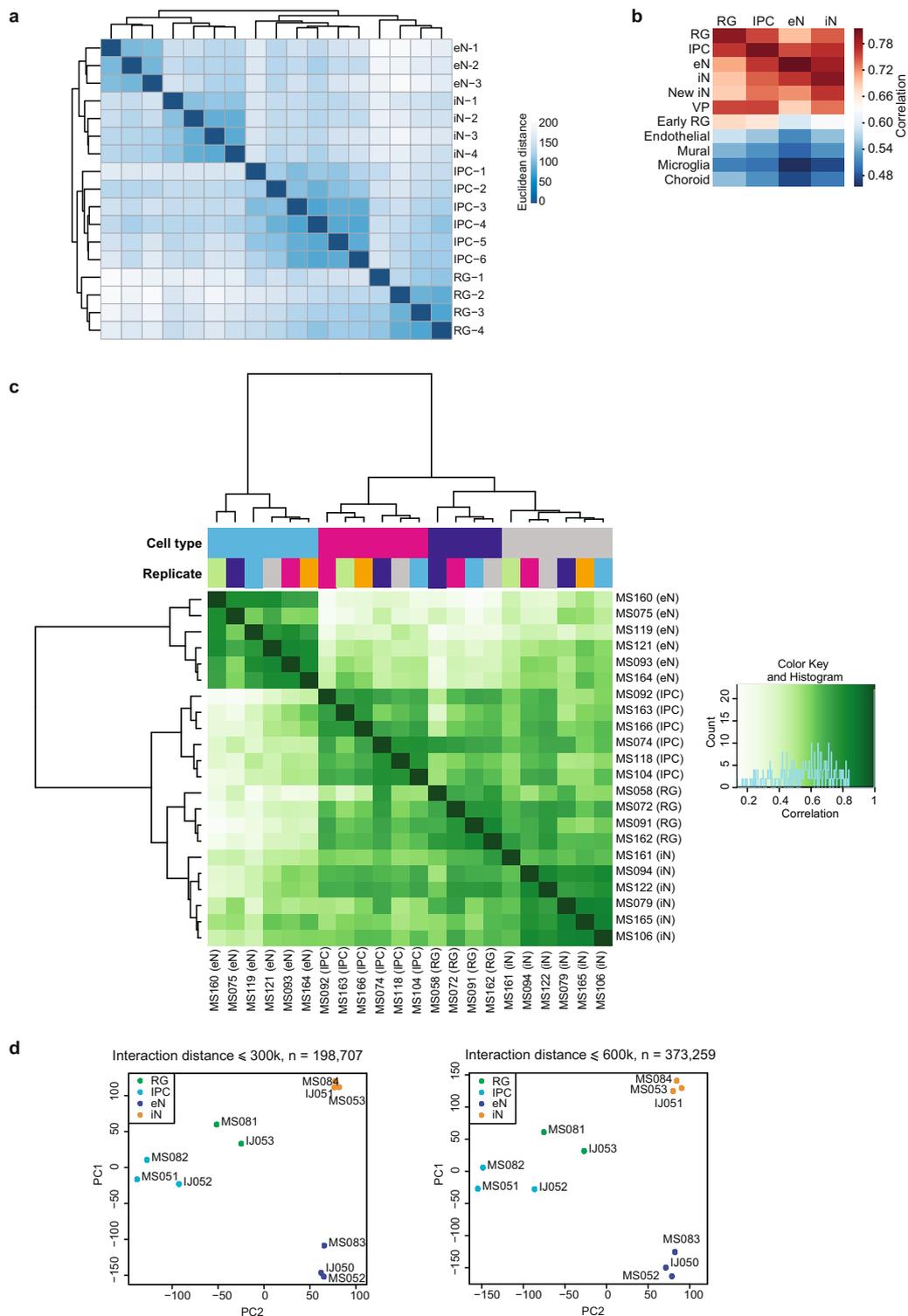
Peer review information *Nature* thanks Andrew Adey, Chongyuan Luo and Barbara Treutlein for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.



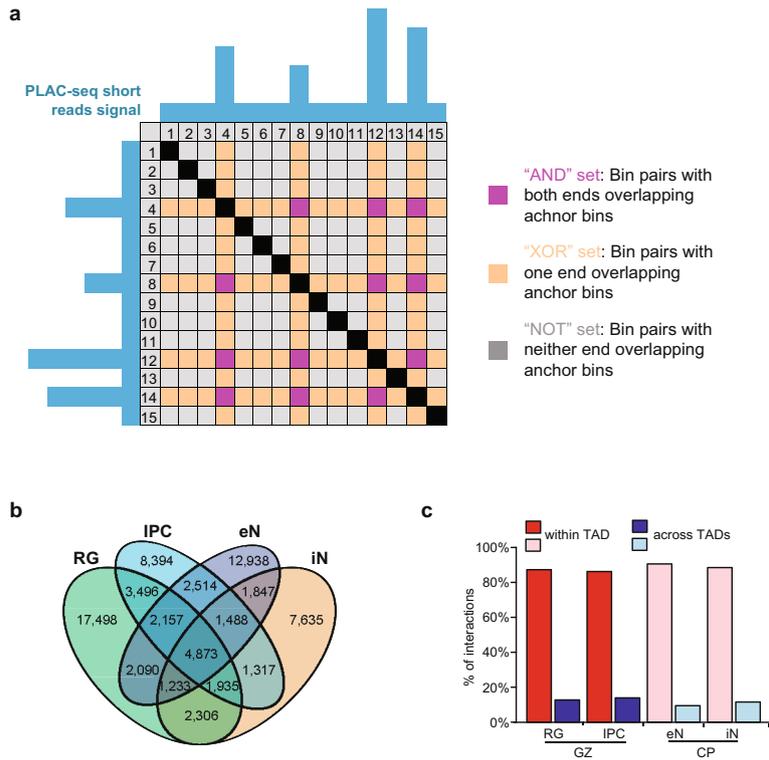
Extended Data Fig. 1 | Representative contour plots depicting FACS gating strategy. **a**, Cells were separated from debris of various sizes based on the forward scatter area (FSC-A) and side scatter area (SSC-A). **b, c**, Specifically, they were passed through two singlet gates using the width and height metrics of the side scatter (SSC-H versus SSC-W) (**b**) and forward scatter (FSC-H versus

FSC-W) (**c**). **d**, SOX2⁺ and SOX2⁻ and IPC populations were isolated by gating on EOMES-PE-Cy7 and SOX2-PerCP-Cy5.5 staining. **e**, RG and interneurons were isolated based on high PAX6/high SOX2 and medium SOX2/low PAX6 staining, respectively. **f**, excitatory neurons were isolated from the SOX2⁻ population by gating on SATB2-Alexa Fluor 647 staining.



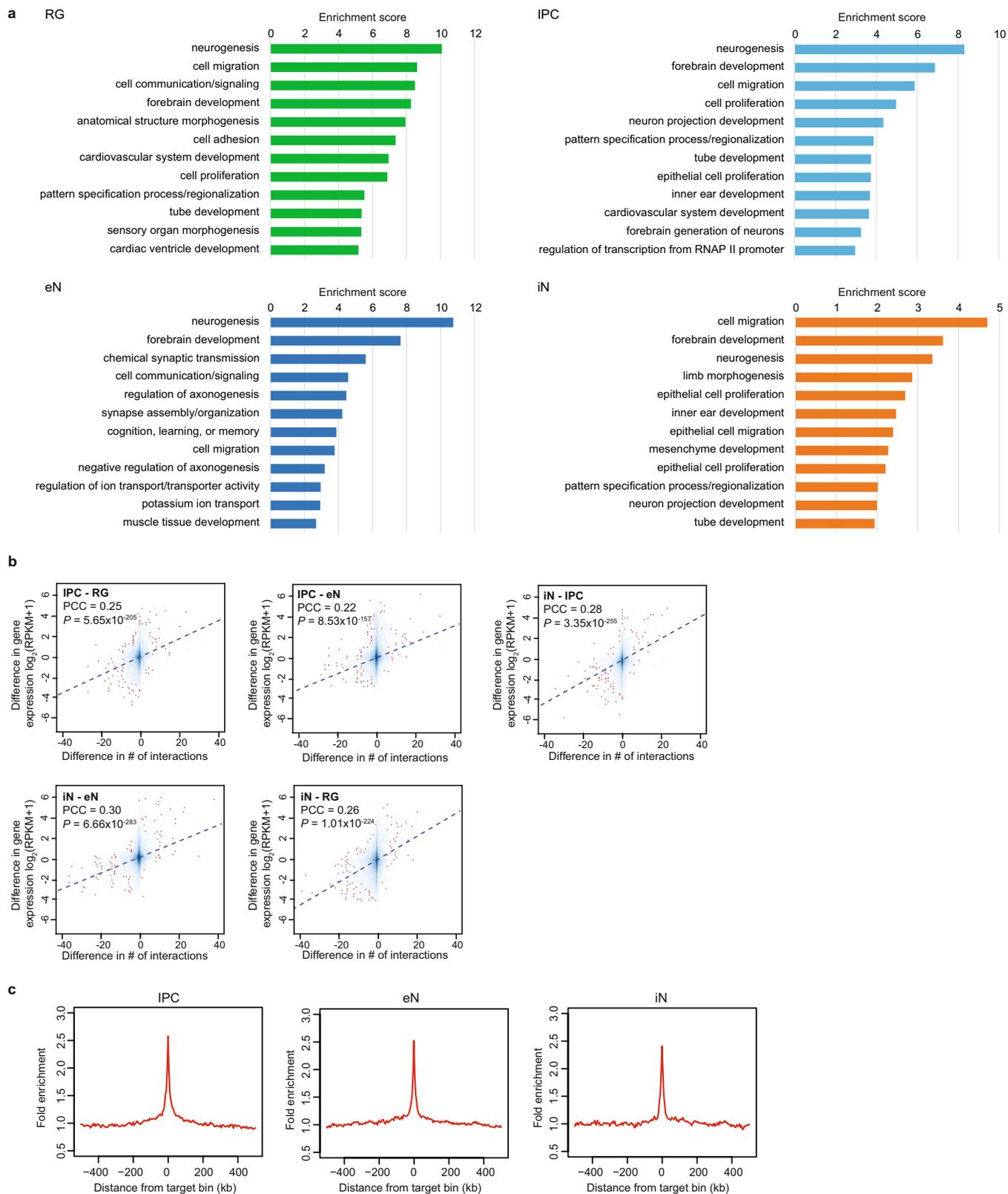
Extended Data Fig. 2 | Reproducibility between RNA-seq, ATAC-seq and PLAC-seq replicates. **a**, RNA-seq replicates were hierarchically clustered according to gene expression sample distances using DESeq2. **b**, Heat map showing correlations between gene expression profiles for the sorted cell populations and single-cell RNA sequencing (scRNA-seq) data in the developing human cortex. The sorted cell populations exhibit the highest correlation with their corresponding subtypes while exhibiting reduced

correlation with the endothelial, mural, microglial, and choroid plexus lineages. **c**, Heat map showing correlations and hierarchical clustering for read densities at open chromatin peaks across all ATAC-seq replicates. **d**, Principal component analysis (PCA) was performed based on normalized contact frequencies across all PLAC-seq replicates (Methods). PCA was performed using interacting 5-kb bins in both 300- and 600-kb windows.



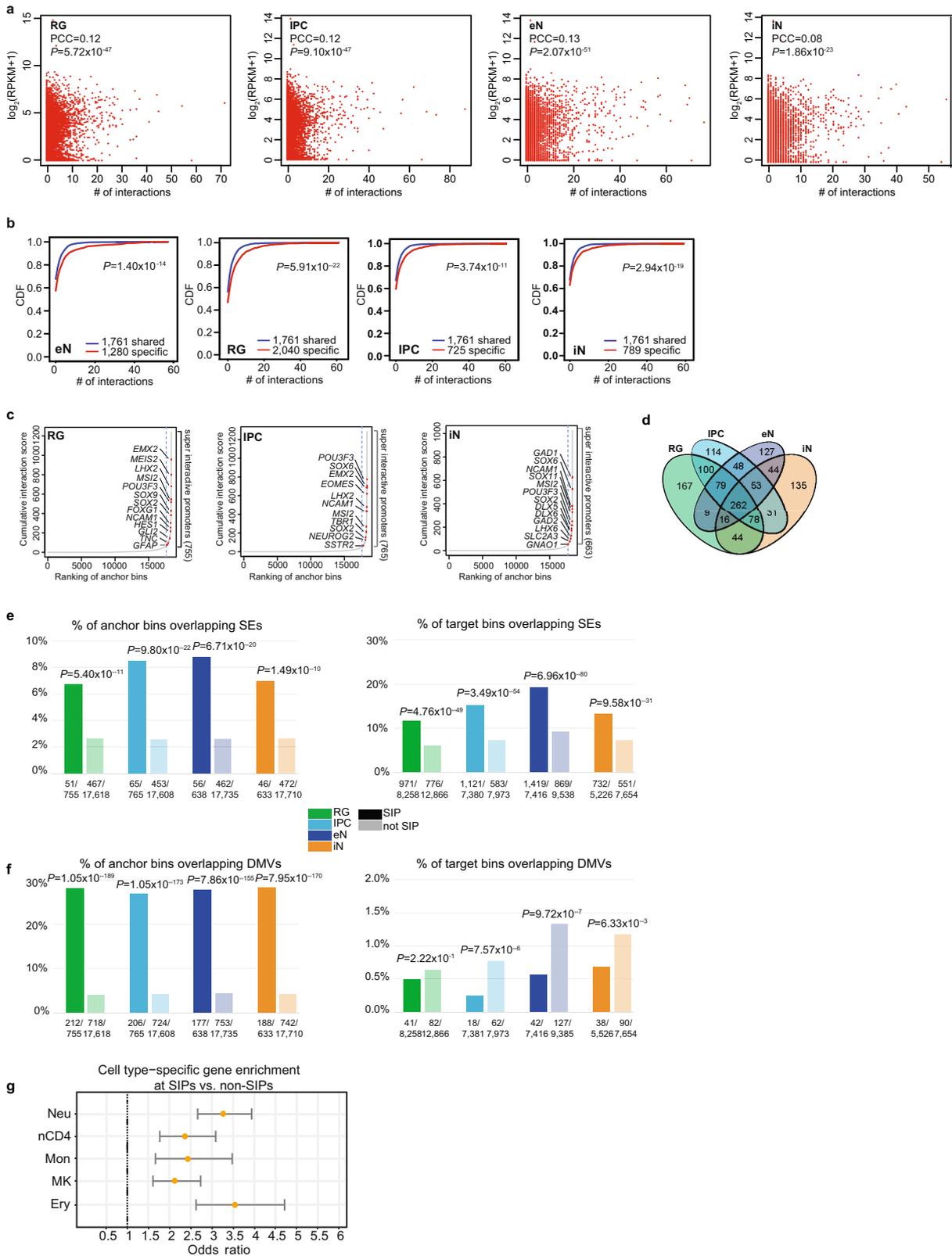
Extended Data Fig. 3 | Identification of significant H3K4me3-mediated chromatin interactions. a, Illustration of XOR and AND interactions in a representative PLAC-seq contact matrix. The blue tracks represent H3K4me3 peaks at anchor bins. Purple cells represent AND interactions where both of the interacting bins are anchor bins. Orange cells represent XOR interactions where only one of the interacting bins is an anchor bin. Grey cells represent

NOT interactions where neither of the interacting bins are anchor bins. **b,** Venn diagram displaying cell-type-specificity for interactions in each cell type. **c,** Proportions of interactions occurring within and across topologically associated domains in the germinal zones and cortical plate for matching cell types.



Extended Data Fig. 4 | Chromatin interactions influence cell-type-specific transcription. **a**, GO enrichment analysis for genes participating in cell-type-specific interactions. The top annotation clusters from DAVID are reported along with their group enrichment scores for each cell type (Methods). **b**, Scatter plots showing the correlation between the difference in the number of interactions for each promoter and the difference in the

expression of the corresponding genes across all cell types (Pearson product-moment correlation coefficient, two-tailed, $n=13,996$ anchor bins with promoters). The trend line from linear regression is shown. **c**, Fold enrichment of open chromatin peaks over distance-matched background regions in 1-Mb windows around distal interacting regions in IPCs, excitatory neurons and interneurons.

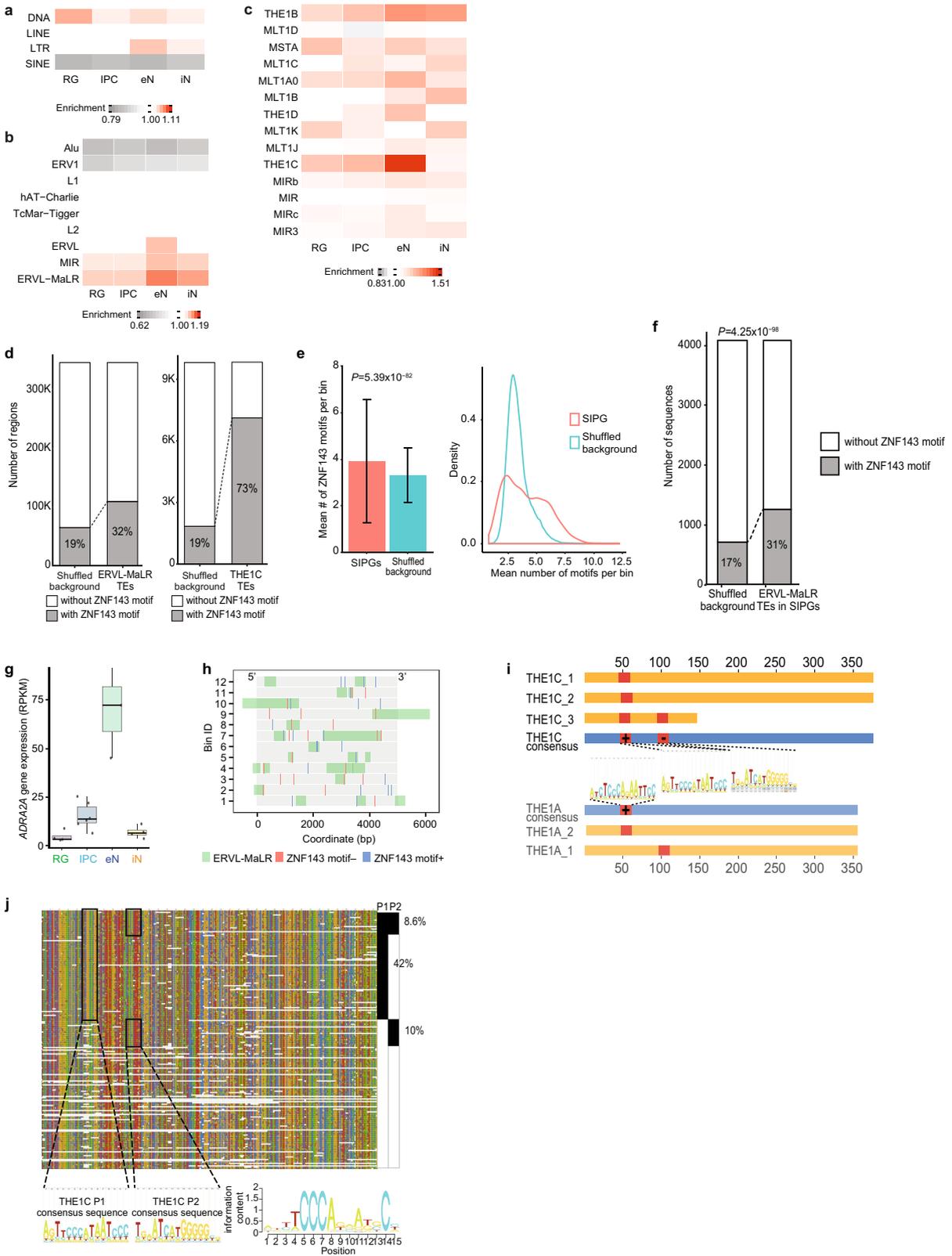


Extended Data Fig. 5 | See next page for caption.

Article

Extended Data Fig. 5 | SIPs are enriched for lineage-specific genes. a, Scatter plots showing the correlation between interaction counts and gene expression at promoters for each cell type (Pearson product-moment correlation coefficient, two-tailed, $n = 13,996$ anchor bins with promoters). **b,** CDF plots of the numbers of interactions for shared versus cell-type-specific genes for each cell type. P values determined by two-tailed two-sample t -test, two-tailed. **c,** Anchor bins were ranked according to their cumulative interaction scores in RG, IPCs and interneurons. SIPs are located past the point in each curve where the slope is equal to 1. **d,** Venn diagram displaying cell-type-specificity for SIPs in each cell type. **e, f,** Enrichment of super-enhancers and

DNA methylation valleys at SIPs versus non-SIPs (left) and distal interacting regions for SIPs versus non-SIPs (right), P values determined by two-tailed Fisher's exact test. Super-enhancers were based on data in the fetal brain and adult cortex, and DNA methylation valleys were based on data in 40- and 60-day cerebral organoids with closely matched gene expression profiles to mid-fetal cortex samples. **g,** Forrest plot showing that SIPs identified in haematopoietic cells are analogously enriched for cell-type-specific over shared genes. Odds ratios and 95% confidence intervals are shown. We identified 554, 709, 460, 712 and 401 SIPs in neutrophils, naive CD4⁺ T cells, monocytes, megakaryocytes, and erythroblasts, respectively.



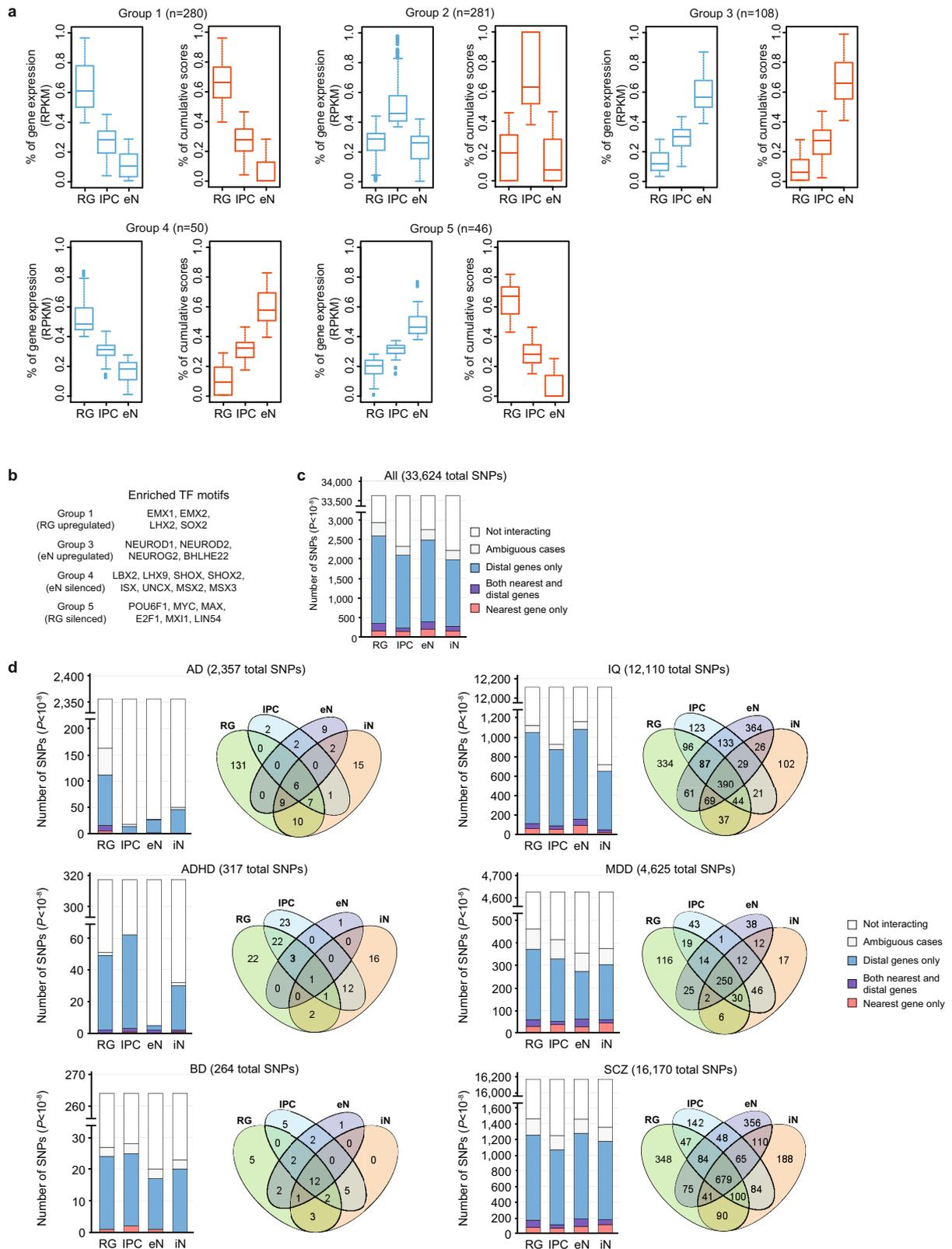
Extended Data Fig. 6 | See next page for caption.

Article

Extended Data Fig. 6 | Transposable elements in SIP formation. a–c,

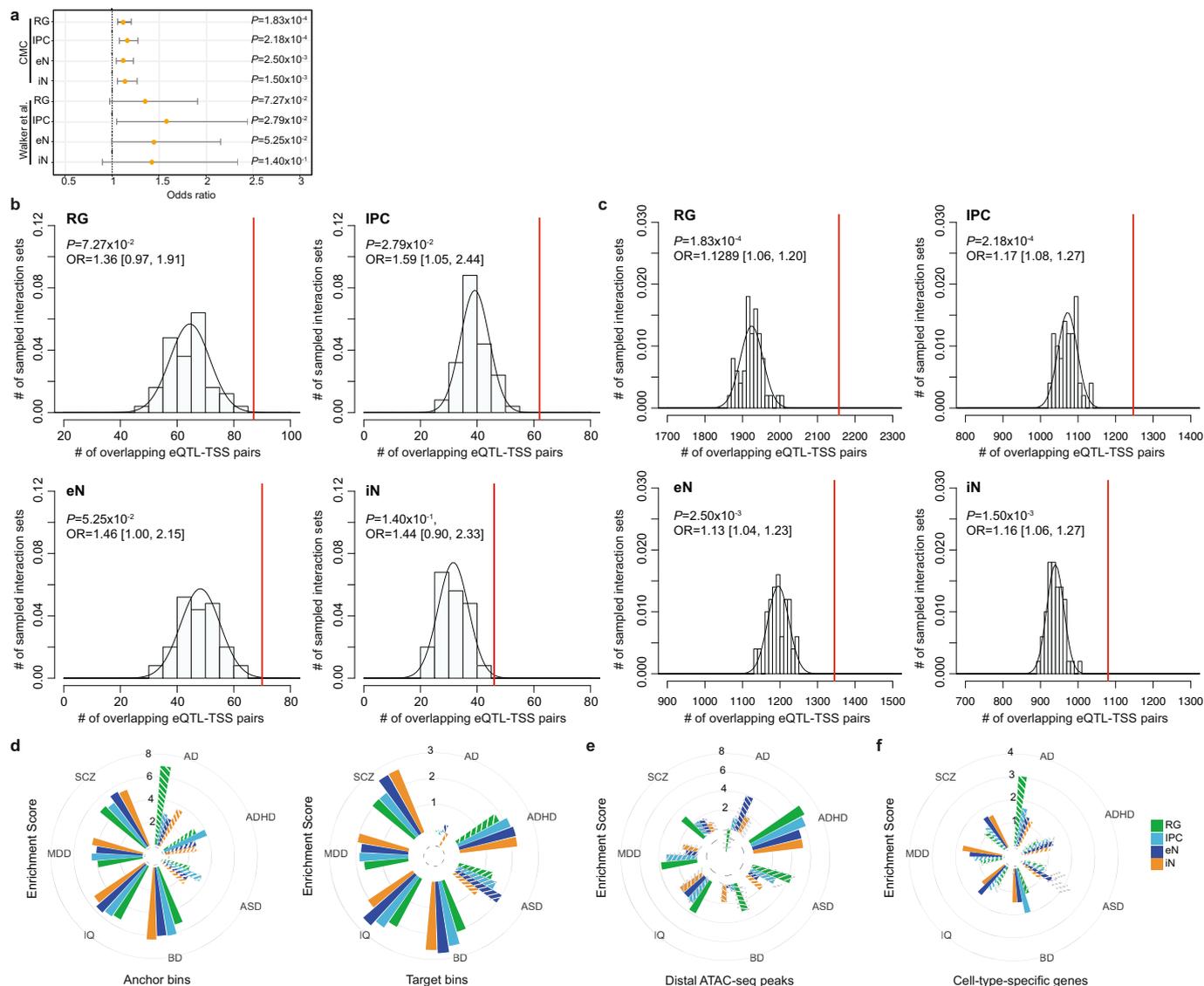
Enrichment of TEs at the class (a), family (b), and subfamily (c) levels in SIPGs for each cell type. Only TE families occupying more than 1% of the genome are shown in b. Only TE subfamilies from the MIR and ERLV-MaLR TE families occupying more than 0.1% of the genome are shown in c. d, Both ERLV-MaLR TEs (left, 32% versus 19% of sequences, $P < 2.2 \times 10^{-16}$, binomial test, two-tailed) and THE1C TEs (right, 73% versus 19% of sequences, $P < 2.2 \times 10^{-16}$, binomial test, two-tailed) are enriched over background sequences for ZNF143 motifs in excitatory neurons. e, ZNF143 motifs are enriched at SIPGs in excitatory neurons (left, $P = 5.39 \times 10^{-82}$, two-sample *t*-test, two-tailed, $n = 8,894$ distal interacting regions). Data are mean and s.e.m. Distributions comparing the number of ZNF143 motifs per bin for actual versus shuffled SIPGs are shown (right, $P < 2.2 \times 10^{-16}$, Kolmogorov–Smirnov test, two-tailed, $n = 638$ SIPGs).

f, ERLV-MaLR TEs in SIPGs are enriched over background sequences for ZNF143 motifs in excitatory neurons (31% versus 17% of sequences, $P = 4.3 \times 10^{-98}$, binomial test, two-tailed). g, Box plots showing elevated *ADRA2A* gene expression in excitatory neurons. The median, upper and lower quartiles, minimum and maximum are indicated. h, Illustration of the 12 distal interacting regions containing ERLV-MaLR TE-localized ZNF143 motifs in the *ADRA2A* SIPG. ZNF143 motifs are coloured by strand. The bin numbers correspond to Fig. 3g. i, Conservation of ERLV-MaLR TEs in the *ADRA2A* SIPG. Blue bars indicate consensus sequences, yellow bars indicate ERLV-MaLR TEs, and red bars indicate ZNF143 motifs. j, Alignment of THE1C TEs in the human genome to their consensus sequence. The THE1C subfamily contains two ZNF143 motifs, one at positions 47–61 (P1), and another at positions 96–110 (P2).



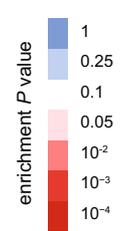
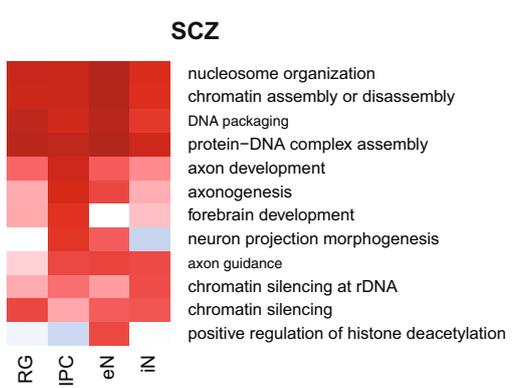
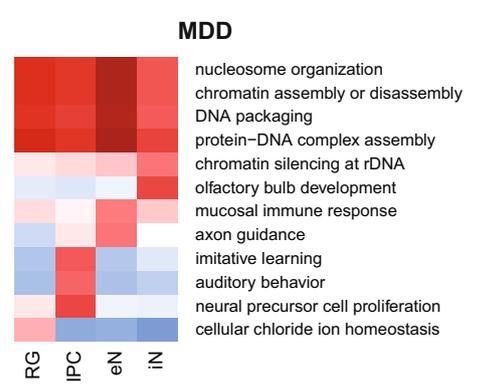
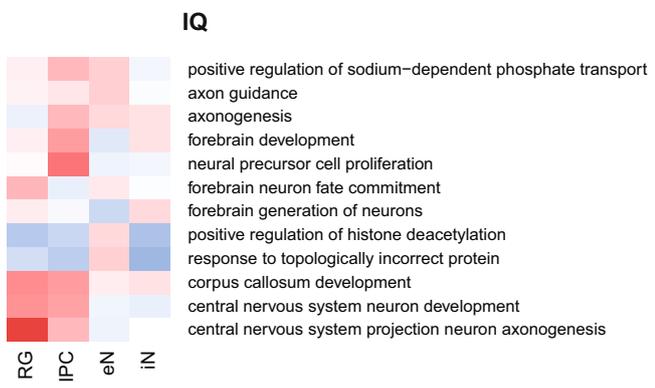
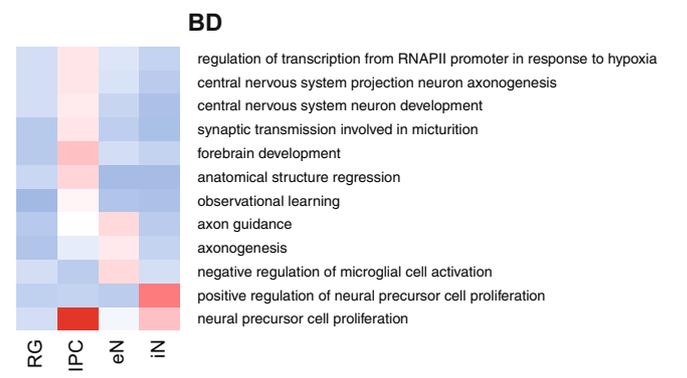
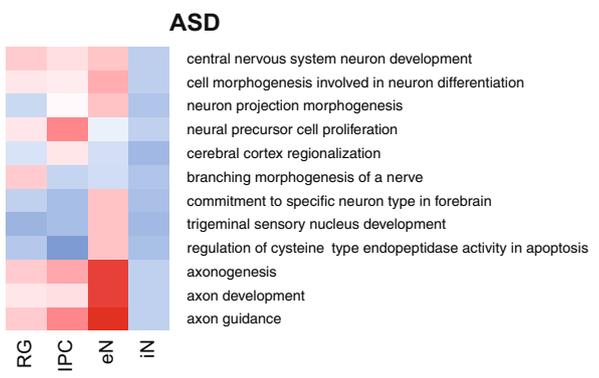
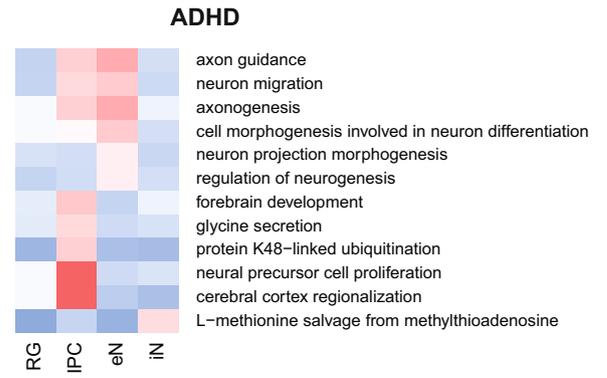
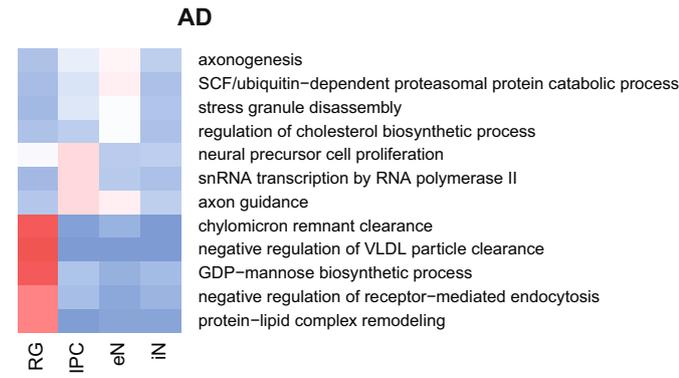
Extended Data Fig. 7 | Developmental trajectories and mapping complex disorder- and trait-associated variants to their target genes. **a**, Box plots showing the distributions of gene expression and cumulative interaction scores for the groups identified in Fig. 4a. The median, upper and lower quartiles, minimum, and maximum are indicated. **b**, Groups 4 and 5 are

enriched for interactions with transcription factors containing domains associated with transcriptional repression. **c**, **d**, Counts of the numbers of GWAS SNPs ($P < 10^{-8}$) interacting with their nearest gene only, with both their nearest and more distal genes, and with more distal genes only across all diseases (**c**) and specific disorders and traits (**d**).



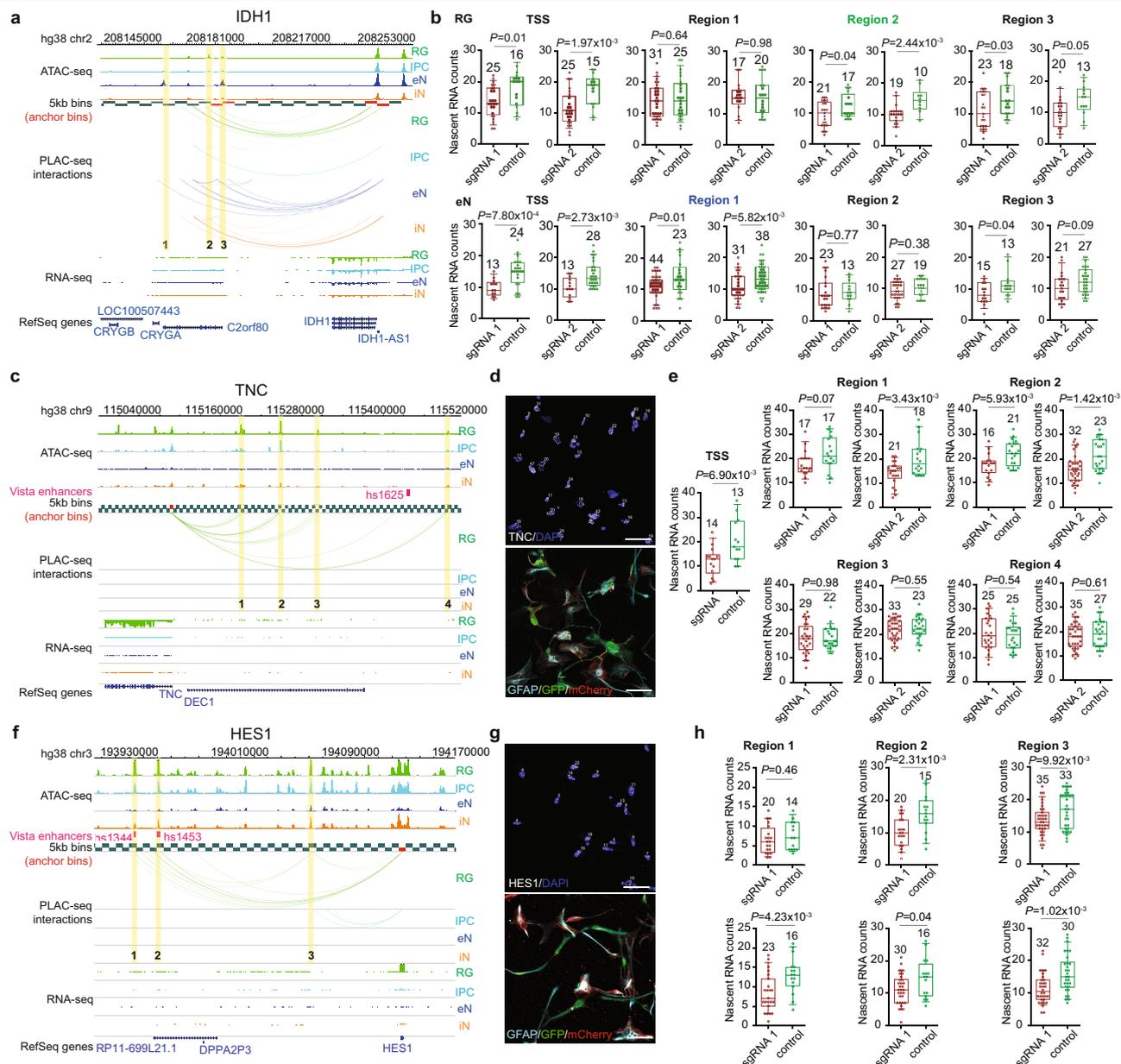
Extended Data Fig. 8 | Partitioning SNP heritability for complex disorders and traits using alternative epigenomic annotations. **a**, Forrest plot showing the enrichment of fetal and adult brain eQTL-TSS pairs in our interactions compared to $n = 50$ sets of distance-matched control interactions. P values determined by two-tailed Fisher's exact test. Odds ratios and 95% confidence intervals are shown. The increased significance of adult brain eQTLs can be attributed to the larger sample size of the CommonMind Consortium (CMC) study ($n = 1,332,863$), while larger odds ratios were observed for the more closely matched fetal brain eQTLs ($n = 6,446$). **b, c**, Histograms displaying the numbers of adult and fetal brain eQTL-TSS pairs recapitulated by $n = 50$ sets of

distance-matched control interactions in each cell type. The numbers of eQTL-TSS pairs recapitulated by our interactions are indicated by red lines. P values determined by two-tailed Fisher's exact test. **d**, LDSC enrichment scores for each disease and cell type, conditioned on the baseline model previously described⁴³ and stratified by PLAC-seq anchor and target bins. Non-significant enrichment scores are shown as striped bars. **e, f**, LDSC enrichment scores for each disease and cell type, conditioned on the baseline model previously described⁴² and using either distal open chromatin peaks (**e**) or cell-type-specific genes (**f**). Non-significant enrichment scores are shown as striped bars.



Extended Data Fig. 9 | Enriched biological processes for genes interacting with non-coding variants for each disease and cell type. Gene Ontology enrichment analysis for genes interacting with non-coding variants for each

disease and cell type using H-MAGMA and gProfileR. *P* values determined by two-tailed Fisher's exact test, BH method. The full results can be found in Supplementary Table 12.



Extended Data Fig. 10 | Characterization of RG- and excitatory neuron-specific loci using CRISPRview. **a, b**, Validation of distal interacting regions at the *IDH1* locus in RG and excitatory neurons. Silencing region 1, which interacts with the *IDH1* promoter only in excitatory neurons, results in the significant downregulation of *IDH1* expression in excitatory neurons but not in RG. Silencing region 2, which interacts with the *IDH1* promoter only in RG, results in the significant downregulation of *IDH1* expression in RG but not in excitatory neurons. Silencing region 3, which interacts with the *IDH1* promoter in both RG and excitatory neurons, results in the significant downregulation of *IDH1* expression in both cell types. Interactions between the promoter of *IDH1* and distal interacting regions containing open chromatin peaks that were targeted

for silencing are highlighted. Box plots show results for experimental (red) and control (green) sgRNA-treated cells for each region. *P* values determined by two-tailed two-sample *t*-test. The median, upper and lower quartiles and 10% to 90% range are indicated. Open circles represent single cells. Sample sizes are indicated above each box plot. **c–h**, Validation of distal interacting regions at the *TNC* and *HES1* loci in RG. Interactions between the promoters of *TNC* and *HES1* and distal interacting regions containing open chromatin peaks that were targeted for silencing are highlighted. Representative images show staining for intronic RNAscope probes (white), DAPI (blue), GFAP (light blue), GFP (green), and mCherry (red). Scale bar, 50 μ m.

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 [Authors & Referees](#) 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

- | | | |
|-------------------------------------|-------------------------------------|--|
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | A description of all covariates tested |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | 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) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | 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

Sequencing data was obtained from the HiSeq X Ten and NovaSeq 6000 instruments (Illumina).

Data analysis

We used the following software in our study: MAPS (Juric et al. 2019), bedtools 2.27.1, MEME 5.0.3, HOMER 4.10, RStudio v1.1.456, R 3.5.1 and 3.6.1, bowtie 1.2.1.1, STAR 2.7.0f, RSEM 1.3.1, TrimGalore 0.4.5, edgeR 3.20.9, Picard Tools 1.141, MACS2 2.1.1, DiffBind 2.6.6, DESeq2 1.18.1, LDSC 1.0.0, DAVID 6.8, MAGMA 1.07b, gprofiler2 0.1.8, and SMART-Q 1.1.

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/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

All datasets used in this study (PLAC-seq, ATAC-seq, RNA-seq) are available at the Neuroscience Multi-Omic Archive (NeMO Archive) under controlled access. Chromatin interactions, open chromatin peaks, and gene expression profiles for each cell type can be downloaded from the NeMO Archive using the following link: <https://assets.nemoarchive.org/dat-ui0qy8b>

Cell type-specific 3D epigenomes can be visualized on the WashU Epigenome Browser using the datahub at the following link: http://epigenomegateway.wustl.edu/browser/?genome=hg38&position=chr17:72918238-73349675&hub=https://shen-msong.s3-us-west-1.amazonaws.com/hfb_submission/hfb_datahub.json

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](https://www.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	Our study design ranges from between two to two and six to six pairwise comparisons across our assays and has a power of approximately 80% for detecting a mean difference of 1.80 to 5.66 standard deviations at a nominal significance threshold of 0.05. We chose this study design because effect sizes of this magnitude are compatible with our research goals.
Data exclusions	No data were excluded from the analyses.
Replication	All attempts at replication were successful and are described in the text and in Extended Data Fig. 2.
Randomization	Randomization is not relevant to our study because we do not apply any differential treatment, intervention, or perturbation to our samples. Instead, we compare 3D epigenomes for different cell types.
Blinding	Blinding is not relevant to our study because we do not apply any differential treatment, intervention, or perturbation to our samples. Instead, we compare 3D epigenomes for different cell types.

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
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used	EOMES-PE-Cy7 (Invitrogen, Cat 25-4877-42, Clone WD1928, Lot 1923396, 1/10 dilution), PAX6-PE (BD Biosciences, Cat 561552, Clone O18-1330, Lot 8187686, 1/10 dilution), SOX2-PerCP-Cy5.5 (BD Biosciences, Cat 561506, Clone O38-678, Lot 8165744, 1/10 dilution), SATB2-Alexafluor647 (Abcam, Cat ab196536, Clone EPNCIR130A, Lot GR3208103-I and GR228747-2, 1/100 dilution), mCherry (Abcam, Cat ab205402, Lot GR3271744-2, 1/200 dilution), GFP (Abcam, Cat ab1218, Clone 9F9.F9, Lot GR213436-38, 1/400 dilution), GFAP (Abcam, Cat ab7260, Lot GR3240356-1, 1/500 dilution), SATB2 (Abcam, Cat ab92446, Clone EPNCIR130A, Lot GR3252015-2, 1/300 dilution), Alexa Fluor 488 donkey anti-mouse IgG secondary antibody (Thermo Fisher Scientific, Cat A21202, Lot 2018296, 1/800 dilution), Alexa Fluor 546 donkey anti-rabbit IgG secondary antibody (Thermo Fisher Scientific, Cat A10040, Lot 2020130, 1/500 dilution), Alexa Fluor 594 Goat anti-chicken IgY secondary antibody (Thermo Fisher Scientific, Cat A11042, Lot 1977707, 1/500 dilution).
Validation	EOMES: From manufacturer: "Applications Tested: This WD1928 antibody has been pre-titrated and tested by intracellular staining and flow cytometric analysis of normal human peripheral blood cells using the Foxp3/Transcription Factor Staining Buffer Set (cat. 00-5523) and protocol. Please refer to Best Protocols: Protocol B: One step protocol for (nuclear) intracellular proteins located under the Resources Tab online. This can be used at 5 μ L (0.06 μ g) per test. A test is defined as the amount (μ g) of antibody that will stain a cell sample in a final volume of 100 μ L. Cell number should be determined empirically but can range from 10^5 to 10^8 cells/test." PAX6: See Thomsen et al, Nature Methods, 2016 SOX2: Manufacturer validated for Western Blot, Flow Cytometry, Bioimaging, & Immunofluorescence. Also, see Thomsen et al, Nature Methods, 2016

SATB2: From Manufacturer: "The cells were fixed with 80% methanol (5 min) and then permeabilized with 0.1% PBS-Triton X-100 for 15 min. The cells were then incubated in 1x PBS / 10% normal goat serum to block non-specific protein-protein interactions followed by the antibody (ab196536, 1/50 dilution) for 30 min at 22°C. Isotype control antibody was Rabbit IgG (monoclonal) Alexa Fluor® 647 (ab199093) used at the same concentration and conditions as the primary antibody. Unlabelled sample (blue line) was also used as a control. Acquisition of >5,000 events were collected using a 25mW Red Solid State Diode laser (635nm) and 675/30 bandpass filter. This antibody gave a positive signal in A431 cells fixed with 4% formaldehyde (10 min)/permeabilized with 0.1% PBS-Triton X-100 for 15 min used under the same conditions."

mCherry:

Chicken polyclonal to mCherry. Immunogen: Recombinant full-length protein (His-tag) corresponding to mCherry. The immunogen sequence is from the following reference(Shaner NC et al. Nature Biotechnology 22:1567-1572, 2004.). Tested applications: suitable for WB, ICC/IF. ab205402 has been referenced in at least 7 publications.

GFP:

Mouse monoclonal [9F9.F9] to GFP. Immunogen: Fusion protein corresponding to GFP aa 1-246. Full-Length Fusion Protein. Derived from the jellyfish *Aequorea victoria*. This antibody shows no reactivity against red fluorescence protein (RFP) but known to cross-react with the wild type (wt), recombinant (rGFP) and enhanced (eGFP) forms. It is tested to suitable for WB, IHC-Fr, Sandwich ELISA, ICC/IF, IP. ab1218 has been referenced in at least 236 publications.

GFAP:

Rabbit polyclonal to GFAP. Immunogen: Recombinant full-length protein corresponding to Human GFAP. Isotype 1 expressed in and purified from *E. coli*. It specifically recognizes mammalian GFAP on western blots and immunocytochemically, and reacts to Mouse, Rat, Cat, Dog, Human, Common marmoset. This antibody is suitable for IHC-FoFr, IHC-Fr, IHC-FrFl, ICC/IF, WB, IHC-P, IHC - Wholemount, ICC. ab7260 has been referenced in at least 401 publications.

SATB2 (used in CRISPRview)

Rabbit monoclonal [EPNCIR130A] to SATB2. Tested applications including WB, IHC-P, ICC/IF, Flow Cyt, IHC-Fr, IHC-FoFr. Reacting with samples from mouse, rat, human. Immunogen: Synthetic peptide within Human SATB2. This product is a recombinant monoclonal antibody, which offers several advantages including high batch-to-batch consistency and reproducibility, improved sensitivity and specificity, long-term security of supply and animal-free production. It has been referenced in at least 20 publications.

Alexa Fluor 488 donkey anti-mouse IgG secondary antibody:

Donkey anti-Mouse IgG (H+L) Highly Cross-Adsorbed Secondary Polyclonal Antibody. To minimize cross-reactivity, these donkey anti-mouse IgG whole antibodies have been affinity-purified and show minimum cross-reactivity to bovine, chicken, goat, guinea pig, hamster, horse, human, mouse, rat, and sheep serum proteins. This product has been referenced in at least 68 publications.

Alexa Fluor 546 donkey anti-rabbit IgG secondary antibody:

Donkey anti-Rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Polyclonal Antibody. These donkey anti-rabbit IgG (H+L) whole secondary antibodies have been affinity-purified and show minimum cross-reactivity to bovine, chicken, goat, guinea pig, hamster, horse, human, mouse, rat, and sheep serum proteins. This product has been referenced in at least 31 publications.

Alexa Fluor 594 Goat anti-chicken IgY secondary antibody:

Goat anti-Chicken IgY (H+L) Secondary Polyclonal Antibody. These goat anti-chicken IgY (H+L) whole secondary antibodies have been affinity-purified and show minimum cross-reactivity. Cross-adsorption or pre-adsorption is a purification step to increase the specificity of the antibody resulting in higher sensitivity and less background staining. This product has been referenced in at least 26 publications.

Flow Cytometry

Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

Donated developing human dorsal cortical samples were cut into pieces, placed into papain dissociation media, triturated, filtered, washed, and fixed using 2% PFA.

Instrument

Samples were sorted on the FACSaria II, FACSaria Ilu, or FACSaria Fusion instruments using the following laser lines: 488 nm, 561 nm, and 633 nm.

Software	Data collection was performed using FACSDiva, and representative images were created using FlowJo.
Cell population abundance	Radial glia represented 1-5% of singlets. Intermediate progenitor cells represented 4-9% of singlets. Excitatory neurons represented 30-40% of singlets. Interneurons were the most variable, representing up to 9% of singlets. Sample purity was assessed by running the sorted cell populations back on the same machine and ranged from 85-95%. Events that appeared outside the original gates were largely non-fluorescent.
Gating strategy	FSC area and SSC area was used to separate cells from debris. FSC area gates were from 10k to 250K, and SSC area gates were from 20K to 250K. This gated population then went through two additional gates to filter out doublets using SSC-H versus SSC-W and FSC-H versus FSC-W. Based on the contour plots, gates were drawn to focus on the main population and eliminate the small doublet population above it. For antibody staining, a signal between 10^3 and 10^4 was considered moderately positive (low), and a signal above 10^4 was considered high.

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.