Cell-type-specific effects of genetic variation on chromatin accessibility during human neuronal differentiation

Dan Liang1,2, Angela L. Elwell1,2, Nil Aygün1,2, Oleh Krupa1,2, Justin M. Wolter1,2, Felix A. Kyere1,2, Michael J. Lafferty1,2, Kerry E. Cheek1,2, Kenan P. Courtney1,2, Marianna Yusupova3,4,5, Melanie E. Garrett6, Allison Ashley-Koch6,7, Gregory E. Crawford8,9, Michael I. Love10, Luis de la Torre-Ubieta3,4,5,11, Daniel H. Geschwind3,4,5,11 and Jason L. Stein1,2

Common genetic risk for neuropsychiatric disorders is enriched in regulatory elements active during cortical neurogenesis. However, it remains poorly understood as to how these variants influence gene regulation. To model the functional impact of common genetic variation on the noncoding genome during human cortical development, we performed the assay for transposase-accessible chromatin using sequencing (ATAC-seq) and analyzed chromatin accessibility quantitative trait loci (caQTL) in cultured human neural progenitor cells and their differentiated neuronal progeny from 87 donors. We identified significant genetic effects on 988/1,839 neuron/progenitor regulatory elements, with highly cell-type and temporally specific effects. A subset (roughly 30%) of chromatin accessibility-QTL were also associated with changes in gene expression. Motif-disrupting alleles of transcriptional activators generally led to decreases in chromatin accessibility, whereas motif-disrupting alleles of repressors led to increases in chromatin accessibility. By integrating cell-type-specific chromatin accessibility-QTL and brain-relevant genome-wide association data, we were able to fine-map and identify regulatory mechanisms underlying noncoding neuropsychiatric disorder risk loci.

Genome-wide association studies (GWAS) have revealed hundreds of common single nucleotide polymorphisms (SNPs) that are associated with risk for neuropsychiatric disorders and interindividual differences in brain structure1–12. A crucial next step is to understand the molecular mechanisms underlying the effect of these variants. This is complicated by many factors, including unknown causal variant(s) at an associated locus due to linkage disequilibrium (LD), unknown cell-type(s), tissue-type(s) or developmental time period(s) in which a genetic risk variant exerts its effects and unknown regulatory function of noncoding risk variants. Nevertheless, a commonly assumed model to explain molecular mechanisms underlying risk loci is that noncoding risk alleles disrupt transcription factor (TF) binding within cell-type-specific regulatory elements (REs) leading to alterations in gene expression and downstream effects on risk13–15. Thus, understanding genetic effects on cell-type-specific regulatory activity is an essential aspect of moving from genetic association to a meaningful biological understanding of disorder risk.

With this in mind, several consortia including ENCODE, GTEx and PsychENCODE have taken major steps to build maps of noncoding genome function across the body14–15. These and other efforts have connected noncoding genetic variation to genes in developing and adult brain tissue by profiling three-dimensional chromatin interactions and by measuring the genetic effects on gene expression, called expression quantitative trait loci (eQTLs)13–15. Although these studies are an important first step in connecting noncoding risk loci to genes, they do not explain how the gene is regulated via genetic variation.

Risk variants for multiple neuropsychiatric disorders are enriched in REs active at mid-gestation in humans, during cortical neurogenesis15. Histone acetylation QTLs (haQTLs) and chromatin accessibility QTLs (caQTLs) are powerful tools to identify the effect of genetic variation on noncoding REs and provide further understanding of regulatory mechanisms at tissue or cell-type-specific levels13–16. However, the ability to connect human genetic variation to longitudinal changes in regulatory architecture during brain development is limited by the inaccessibility of brain tissue from the same individual over multiple time points. Here, we leveraged a well validated model of human brain development based on in vitro culture of primary human neural progenitors (pHNPCs)17 to study the functional effects of genetic variation on chromatin architecture during neurogenesis. We measured chromatin accessibility in cell culture in a cell-type-specific manner at two key stages of neural development, during progenitor proliferation ($n_{donor}=76$) and after

1Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. 2UNC Neuroscience Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. 3Neurogenetics Program, Department of Neurology, David Geffen School of Medicine University of California, Los Angeles, Los Angeles, CA, USA. 4Center for Autism Research and Treatment, Semel Institute, David Geffen School of Medicine University of California, Los Angeles, Los Angeles, CA, USA. 5Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA. 6Duke Molecular Physiology Institute, Duke University, Durham, NC, USA. 7Department of Medicine, Duke University, Durham, NC, USA. 8Center for Genomic and Computational Biology, Duke University, Durham, NC, USA. 9Department of Pediatrics, Division of Medical Genetics, Duke University, Durham, NC, USA. 10Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. 11Department of Psychiatry and Biobehavioral Sciences, Semel Institute, David Geffen School of Medicine University of California, Los Angeles, Los Angeles, CA, USA. 12E-mail: jason_stein@med.unc.edu
differentiation using their sorted neuronal progeny ($n_{\text{progenitor}} = 61$). We identified thousands of caQTLs and allele specific chromatin accessibility (ASCA) sites, most of which were highly cell-type-specific. We use the effects of these genetic variations to understand how disrupting TF binding motifs affect chromatin accessibility and gene expression, as well as to understand the cell-type-specific regulatory mechanisms underlying genetic risk for neuropsychiatric disorders.

**Results**

**Genome-wide chromatin accessibility profiling.** We generated pHNPC lines from 14–21 gestation-week genotyped fetal brains ($n = 87$) using a neural sphere isolation method that results in cultures with high fidelity to the in vivo developing brain (Fig. 1a and Methods). pHNPCs were cultured and isolated at two stages: progenitor cells and 8-week differentiated and sorted neurons (Extended Data Fig. 1a,b). Using immunofluorescence of neural cell markers, we found over 90% of the progenitor cells were positive for SOX2 and PAX6, indicating a highly homogenous population of radial glia cells (Fig. 1a and Extended Data Fig. 1c). After 8 weeks of differentiation, we sorted neurons labeled using a viral construct (AAV2-hSyn1-EGFP) by using FACS (fluorescence-activated cell sorting), which showed typical neuronal morphology (Fig. 1a, Extended Data Fig. 1b,d and Methods). We performed the assay for transposase-accessible chromatin using sequencing (ATAC-seq) on intact nuclei and found that libraries were high quality based on a comparison of quality metrics relative to previous in vivo developing brain data, as well as a sensitivity analysis and nucleosome periodicity (Extended Data Figs. 1e and 2a,b and Methods). We quantified accessibility as batch-effect-corrected reads within accessible peaks normalized for GC content, peak length and sequencing depth (Extended Data Fig. 2c,d,e,f). We found higher correlations of chromatin accessibility for libraries from the same donors cultured at different times as compared to correlations across donors (Extended Data Fig. 2d). To ensure independence for subsequent analyses, we randomly selected one library from each donor for each cell type (($n_{\text{progenitor}} = 76$ and $n_{\text{neuron}} = 61$) to identify accessible peaks ($n = 90,227$, average peak length of 409 bp; Methods).

To determine the in vivo relevance of these accessible peaks, we performed an overlap analysis using previously classified chromatin states from 93 in vivo human tissues and cell types (Fig. 1b and Methods). The accessible peaks from progenitors and neurons most strongly overlapped with enhancers and promoters in brain germinal matrix and fetal brain tissue, followed by other brain regions, indicating that these peaks were highly representative of the in vivo fetal brain. Principal component analysis (PCA) of chromatin accessibility revealed that progenitors and neurons clearly separate along the first principal component (Fig. 1c), indicating that cell-type was associated with the largest variability in chromatin accessibility profiles (64.91% of variance explained). These results demonstrate that chromatin accessibility measured from pHNPC cultures are representative of REs present in the developing human brain and that chromatin accessibility patterns are different between progenitors and neurons, consistent with previous data from fetal brain tissues (Fig. 1f).

**Identifying cell-type-specific REs.** To reveal cell-type-specific REs involved in neuronal differentiation, we performed an analysis to determine which peaks had significantly different chromatin accessibility between progenitors and neurons (Fig. 1d and Methods). We identified 35,379 peaks with greater accessibility in progenitors than neurons (progenitor peaks) and 44,729 peaks with greater accessibility in neurons than progenitors (neuron peaks, false discovery rate (FDR) < 0.05; Supplementary Table 1). At the promoter of SYNJ, which was used to label neurons for sorting after differentiation, we observed considerably higher accessibility in neurons, as expected (log(fold change) [LFC] = −2.88, $P = 2.83 \times 10^{-26}$; Fig. 1d). Among significant differentially accessible peaks, we found greater accessibility in progenitors at the promoters of genes highly or uniquely expressed in progenitors, such as the dorsal telencephalic marker $EMX2$ (Fig. 1e)\(^1\). Moreover, promoters of genes highly expressed in neurons, such as $DCX$, $BDNF$, $CAMK2B$ and $SYT13$ (ref. \(^2\)), showed greater chromatin accessibility in neurons (Fig. 1d,e).

We found an expected enrichment of Gene Ontology terms related to neurogenesis in genes with differentially accessible promoters (Extended Data Fig. 3a and Methods). We also found that differentially accessible peaks were significantly enriched in ATAC-seq peaks from the relevant in vivo fetal brain laminae (Fig. 1f)\(^1\). Specifically, progenitor peaks were more enriched in peaks with higher accessibility in the progenitor-enriched germinal zone. Conversely, neuron peaks were more enriched in peaks with higher accessibility in the neuron-enriched cortical plate. These results showed differentially accessible peaks represent cell-type-specific active REs and were in strong agreement with biological processes and gene-regulatory behavior present within in vivo fetal brain tissues.

To detect TFs involved in neuronal differentiation, we conducted a differential motif enrichment analysis to predict TF binding sites more active in either progenitors or neurons. We detected 62 TFs (FDR < 0.05) with binding sites present more often in progenitor peaks than neuron peaks (here called progenitorTFs), and 208 TF motifs present more often in neuron peaks than progenitor peaks (neuronTFs) (Methods, Extended Data Fig. 3b and Supplementary Table 2). Within progenitorTFs and neuronTFs, we found TFs previously characterized with key roles in neurogenesis, which provides further support that TF binding within accessible peaks from this in vitro system reflect the expected in vivo developmental processes (Extended Data Fig. 3c and Supplementary Table 2)\(^1\). We also identified several TFs that have not been previously associated with neuronal differentiation, such as $MEF2A$, $MIX-A$ and $HOX85$, which may be useful for directed differentiation of progenitors.

**caQTLs.** To identify genetic variants that influence chromatin accessibility within cell types representing longitudinal changes during cortical development, we performed caQTL analyses separately for progenitors and neurons using in total 90,227 peaks and 10 million genetic variants (Fig. 2a and Extended Data Fig. 4a,b). We stringently controlled for population stratification (Extended Data Fig. 4e) in association tests using a mixed linear model including a kinship matrix as a random effect, and ten genotype multidimensional scaling (MDS) components as fixed effects\(^12\). In addition, we included principal components across chromatin accessibility profiles and sorter in neurons as fixed effect covariates to reduce the impact of unmeasured technical variation\(^15\).

After the stringent multiple testing correction (Methods), we identified 1,839 progenitor caPeaks and 988 neuron caPeaks at FDR < 5%. The caPeaks were significantly enriched in active REs defined in the fetal brain (Extended Data Fig. 4d), consistent with their expected regulatory function. The most significant caSNPs of each caPeak are most often found near the peaks they are associated with (Fig. 2b). We found that the most significant caSNP or an LD-proxy was located in annotated functional regions for over 80% of caPeaks (Fig. 2c and Supplementary Table 3). These results indicate that most genetic variants affect chromatin accessibility by altering the sequence (and presumably TF binding sites) at the caPeak or disrupt chromatin accessibility at distal peaks that have secondary effects on the caPeak\(^1\).

To identify if genetic influences on chromatin accessibility also affected gene expression, we compared progenitor/neuron caQTLs with eQTLs derived from the same cell lines and eQTLs from the mid-gestation bulk cortical wall\(^12\). For the most significant caSNP for each caPeak, we estimated the posterior probability that the effect is shared with cell-type-specific or cortical wall eQTLs ($m$-value > 0.9, Supplementary Table 4). Thirty percent
of progenitor caQTLs and 34.9% of neuron caQTLs have shared effects with eQTLs in the same cell types, but a smaller proportion are shared with bulk cortical wall eQTLs (Fig. 2d). Those SNPs with shared effects between caQTLs and eQTLs showed strongly positive correlations in effect sizes \((r=0.85)\) in neurons and \((r=0.84)\) in progenitors, Fig. 2c), indicating that alleles associated with increased chromatin accessibility tend to be associated with increased gene expression.

We then compared the number and effect size differences between caQTLs and eQTLs. First, we subsampled the eQTL
Fig. 2 | caQTLs in progenitors and neurons. a, caQTL schematic. b, Number of the most significant caSNPs relative to the distance from the center of the caPeaks (left, neuron caQTLs; right, progenitor caQTLs). The most significant caQTLs for each caPeak can be found in Supplementary Table 3. c, Numbers of caQTLs in different functional categories. d, Schematic cartoon of fetal cortical and cell-type-specific eQTLs (left). Percentage of neuron/progenitor caQTLs with shared effects in fetal cortical and cell-type-specific eQTLs (all shared caQTLs and eQTLs can be found in Supplementary Table 4). e, For the most significant caSNP for each caPeak, correlation of effect sizes between shared caQTLs and eQTLs in neurons (left) and progenitors (right). f, Comparison of percentage variance explained ($r^2$) for shared caQTLs and eQTLs (subset to the same sample size) in neurons and progenitors. We found 500 (e) caSNP–caPeak–eGene combinations in neurons and 1,025 (e) caSNP–caPeak–eGene combinations in progenitors. We observed higher percentage variance explained for caQTLs than eQTLs in both neurons and progenitors. P values are estimated by the two-sided paired t-test. The center of the box is median of the data, the bounds of the box are 25th and 75th percentiles of the data and the whisker boundary is 1.5 times the interquartile range (IQR). Maximum and minimum are the maximum and minimum of the data.
dataset to ensure that caQTLs and eQTLs have the same sample sizes to avoid winner’s curse (Methods). The proportions of peaks (2.62/5.81% in neuron/progenitor) or genes (1.85/5.70%) influenced by genetic variation were comparable in caQTL or eQTLs. However, we observed that caQTLs generally explain more variance than eQTLs, indicating that caQTL studies have higher power than eQTL studies (Fig. 2f and Extended Data Fig. 4e).

Regulatory mechanisms of eQTLs. We next aimed to identify the cognate genes of cell-type-specific REs, fine-map causal variants at eQTL loci and predict regulatory mechanisms underlying eQTLs by overlapping cell-type-specific caQTLs with cell-type-specific or bulk fetal cortical eQTLs ($r^2 > 0.8$ between index ca/eSNPs, Supplementary Table 5). Using cell-type-specific eQTLs, we identified ca/eQTL overlaps in 152/373 RE–gene pairs in neurons/progenitors. Using the larger sample size of fetal cortical eQTLs, we identified 303/282 RE–gene pairs using neuron/progenitor caQTLs. Within these RE–gene pairs, we found many genes involved in neuronal differentiation such as FABP7, VAT1 and FGFI1 (refs. 22-25). We also identified RE–gene pairs where the caSNP-disrupted TF motifs have known function in neuronal differentiation. For example, the G allele of rs185220 was associated with increased chromatin accessibility of a caPeak (chr5:56,909,141–56,910,860) near the SETD9 transcription start site (TSS) and was associated with increased expression of SETD9 in neurons and progenitors (Fig. 3a–c). Several TF motifs were disrupted by this caSNP, but we prioritized REST based on its expression in progenitors22,26 (Fig. 3d; LFCC = −1.85, FDR = 2.75 $\times$ 10$^{-13}$) and evidence of binding at this site in ES cells and to a lesser degree in neurons differentiated from ES cells (Fig. 3a). The G allele of rs185220 led to disruption of the REST motif and increased chromatin accessibility, consistent with the function of REST as a repressor. Through integration of ca/eQTL data, we hypothesize a regulatory mechanism where the G allele of rs185220 disrupts REST binding, resulting in increased chromatin accessibility and increased expression of SETD9 (Fig. 3e).

As experimental validation of this caQTL, we found that the G allele of rs185220 increased the activity of this enhancer in progenitors relative to the A allele using a luciferase assay (Fig. 3f). In contrast with the previous example, we found the C allele of the caSNP, rs11544037, matched the motif of RAD21 and was associated with increased chromatin accessibility of the progenitor-specific enhancer (chr4:158,667,771–158,667,860 located roughly 5 kilobases (kb) upstream from the EFTDH TSS) and increased expression of EFTDH in fetal brain (Extended Data Fig. 5a–e). Experimental validation via a luciferase reporter assay showed a consistent result with the caQTL for this enhancer (Extended Data Fig. 5f). As a final example of regulatory mechanisms underlying a cell-type-specific eQTL, we found the C allele of rs11960262 associated with increased chromatin accessibility of a caPeak (chr5:142,684,441–142,686,700) located in the intron of the gene FGFI1 and also associated with increased FGFI1 expression specifically in progenitors (Extended Data Figs. 5g–i). The C allele of rs11960262 matched the motif of EGR1 (Extended Data Fig. 5j,k), which suggests that EGR1 binding at this caPeak was associated with increased chromatin accessibility and increased expression of FGFI1 in progenitors.

ASCA. We next tested for ASCA at heterozygous SNPs within accessible peaks. ASCA contrasts accessibility between two alleles within an individual at a given heterozygous SNP, so it is less susceptible to cross-individual confounding factors, such as population structure41. In total, we identified 1,602 significant progenitor ASCA and 3,288 significant neuron ASCA (FDR < 0.05; Supplementary Table 6). To determine whether caQTLs also show ASCA, we retained significant caQTLs (non-clumped, FDR < 0.05) using the same heterozygous donors and read level criteria for ASCA, observing that 90.1% of filtered neuron caQTLs were shared with neuron ASCA (Fisher’s test odds ratio [OR] = 51.48, P = 1.32 $\times$ 10$^{-22}$) and 86.9% of filtered progenitor caQTLs were shared with progenitor ASCA (Fisher’s test OR = 45.54, P = 1.37 $\times$ 10$^{-25}$). This demonstrates extremely high overlap between caQTLs and ASCA (Fig. 4a), which indicates minimal influence of cross-individual confounding effects on the caQTL results. Similarly, for all filtered caQTLs and significant ASCA in Fig. 4a, we found high correlations of effect sizes between caQTLs and ASCA ($r = 0.61$ for neurons, $r = 0.69$ for progenitors), indicating a shared direction and degree of effect (Fig. 4b). The alternative allele showed a slight bias, even after controlling reference mapping bias, for a higher correlation with increasing chromatin accessibility from the reference allele (neuron log$_{2}$(ALT/REF) are 53.4% positive, two-sided sign test $P = 0.034$; progenitor log$_{2}$(ALT/REF) are 52.5% positive, two-sided sign test $P = 0.121$).

However, we also detected significant ASCAs that were not significant caQTLs (Fig. 4b). These variants were found in larger peaks than those detected in both caQTL and ASCA (Extended Data Fig. 6a). These ASCA-but-not-caQTL variants likely have an effect on the accessibility of a subregion of the larger active RE. We posit that they are more detectable using ASCA because only reads containing the variant in the accessible region are tested, whereas they are not detectable in caQTLs, which integrate reads across the entire region (Extended Data Fig. 6b). Other ASCA-but-not-caQTL sites were presumably due to lower power for caQTL detection (Extended Data Fig. 6c).

We identified several loci that shared caQTLs, ASCA and eQTLs. For example, a previously described SETD9 locus also demonstrated strong ASCA at rs185220 in neurons and progenitors (Extended Data Fig. 6d). We were also interested in FABP7 (also known as BLBP), which is a marker for radial glia that plays an important role in establishment of radial glial fibers spanning the cortical anlage during cortical development28 (Fig. 4c). The C allele of rs144376334 was associated with increased chromatin accessibility of the caPeak (chr6:122,832,401–122,834,160) in both progenitors and neurons and increased expression of FABP7 (Fig. 4d). The C allele of rs144376334 also showed increased ASCA in both progenitors and neurons (Fig. 4e). rs144376334 disrupted several TF motifs that may drive the effect in both cell types, and we highlight JUN due to its higher expression in progenitors (Fig. 4e; LFCC = −1.22, FDR = 1.82 $\times$ 10$^{-11}$)228. The motif-disrupting allele was associated with decreased chromatin accessibility, consistent with activating REs (Fig. 4g). These results indicate the potential regulatory mechanism underlying this locus in progenitors is that the genetic variation disrupts JUN binding to a distal RE leading to decreased expression of FABP7.

Cell-type-specificity of caQTLs. To determine the cell-type-specificity of ca/eQTLs, we estimated the posterior probability that the allelic effect is shared between the two cell types ($m$-value > 0.9). caQTLs showed a lower proportion of effect sharing between neurons and progenitors (45.6 and 41.0% than eQTLs (78.0 and 56.7%) (Fig. 5a). We found the estimated proportion of true alternative hypotheses that the variant is associated with the trait ($\pi_{a}$) of the most significant neuron/progenitor caSNP–caPeak pairs in progenitors/neurons is 0.73/0.70; however, the $\pi_{a}$ of the most significant neuron/progenitor eSNP–eGene pairs in progenitors/neurons is 0.92/0.77, providing additional support that caQTLs have higher cell-type-specificity than eQTLs. We found 19/35% of progenitor/neuron caSNP–caPeak overlaps with neuron/progenitor caPeaks (Fig. 5b). For ASCA, we found 24/12% of progenitor/neuron ASCA are shared between cell types, which was in agreement with the cell-type-specificity observed in caQTLs. These results indicate that genetic variants often affect chromatin accessibility only within specific cell types.
We further characterized the cell-type-specificity of caQTLs by assessing differential accessibility of caPeaks (Fig. 5c). We found 71.0% of progenitor caPeaks were more accessible in progenitors (LFC > 0). Similarly, 69.8% neuron caPeaks were more accessible in neurons (LFC < 0). This implies that a RE must be accessible to or bound by DNA-binding proteins within a specific
cell-type to observe genetic effects on that RE. We next characterized the location of caPeaks relative to promoters (Fig. 5d). We found there was a higher percentage of cell-type-specific caPeaks that were distal to promoters than shared caPeaks. This result indicates cell-type-specific caQTLs are more likely to affect the chromatin accessibility of distal REs, which is consistent with...
observations that distal REs have higher cell-type-specificity than promoters.

To determine the direction and magnitude of the effect in caQTLs, we related the effect sizes of caQTLs or eQTLs between neurons and progenitors. We found that caQTLs showed a lower correlation between neurons and progenitors (\(r = 0.73\) and \(r = 0.70\)) as compared to eQTLs (\(r = 0.81\) and \(r = 0.81\)) (Fig. 5e,f, \(P < 2.2 \times 10^{-16}\) in neurons and progenitors), which is consistent with the observation that caQTLs showed a lower proportion of shared effects between neurons and progenitors than eQTLs (Fig. 5a). Together, these results indicate that caQTLs have higher cell-type-specificity than eQTLs, within the two cell types tested here.

Comparison to adult dorsolateral prefrontal cortex (DLPFC) caQTLs. Previous work identified genetic variants associated with chromatin accessibility in adult postmortem DLPFC using a sample of 272 individuals. We tested whether the caQTLs identified in our work, modeling a prenatal time period, were also present in the adult cortex. We found 56% of adult peaks are shared with neurons and progenitors (Extended Data Fig. 7a,b). We remapped caQTLs in neurons and progenitors using shared peaks and genetic variants
with the adult data. We did not find any significant neuron/progenitor caQTLs shared with significant caQTLs in adult cortex. For the 27 significant neuron caQTLs, we found the correlation ($r = 0.61$) of effect sizes with adult caQTLs are higher than the correlation ($r = 0.34$) in 35 significant progenitor caQTLs (Extended Data Fig. 7c), which may be expected given that progenitors are not present in the adult cortex. Together, these results indicate that caQTLs have high temporal specificity, as well as cell-type-specificity.

**Prediction of disrupted TF binding.** One favored model of how genetic variation influences chromatin accessibility is that SNPs disrupt TF motifs, decreasing the probability of TF binding to REs, altering chromatin accessibility36. To identify which TF motifs are disrupted by cell-type-specific caSNPs, we mapped TF motifs to the sequence surrounding the neuron-specific/progenitor-specific caSNPs and determined if an allele at the caSNP sufficiently decreases the relative entropy of TF motifs (Methods and Supplementary Table 7). We then performed an enrichment analysis to infer which TF's binding is often affected by caQTLs within each cell type. In progenitors, we found an enrichment of caSNP-disrupted REST and SOX11 motifs, which are known to contribute to neurogenesis32,37. In neurons, we found an enrichment of caSNP-disrupted RAR$\beta$ motifs, which is involved in prefrontal synaptogenesis and axon development38 (Fig. 6a). These results indicate that the motif disrupted TFs are involved in neuronal differentiation, indicating that the genetic variants that affect the activity of REs by disrupting the binding of TFs play functional roles during neurogenesis.

We next tested the impact of the TF motif-disrupting alleles on chromatin accessibility. Among 532 tested motifs in progenitors and 514 tested motifs in neurons, we found that motif-disrupting alleles led to decreased accessibility for 40 (72.7% of significant TFs) TFs in progenitors and 44 (97.8% of significant TFs) TFs in neurons.
(Fig. 6b,c), such as the motif of POU3F2 (also known as BRN2) in progenitors and ASCL2 in neurons, which are both involved in neurogenesis29,40 (Fig. 6d). Conversely, we found the motif-disrupting allele was associated with increased chromatin accessibility at the motif of ZEB1, a known transcriptional repressor41 (Fig. 6e). These results indicate that binding of transcriptional activators is associated with increased chromatin accessibility. However, binding of transcriptional repressors is associated with decreased chromatin accessibility.

Regulatory mechanisms underlying GWAS loci. To investigate if genetic variants associated with brain-related traits are enriched in differentially accessible peaks during neurogenesis, we calculated partitioned heritability enrichment (Fig. 7a,b). We found cell-type-specific enrichments for neuropsychiatric disorders and associated behaviors in accessible regions. Genetic variants associated with several childhood or adult-onset neuropsychiatric disorders or traits, including autistic spectrum disorder, schizophrenia, major depressive disorder (MDD), neuroticism and depressive symptoms, showed significant partitioned heritability enrichment in progenitor peaks. With the exception of schizophrenia, these disorders did not show significant enrichment in neuron peaks. We observed partitioned heritability enrichment for both intelligence and educational attainment within neuron peaks and progenitor peaks. As a negative control, we did not observe enrichment of inflammatory bowel disease heritability in differentially accessible peaks. These results are consistent with the model that genetic variants alter the function of REs during cortical neurogenesis, which then leads to risk for neuropsychiatric disorders or related traits in childhood or adulthood41,42.

We found genetic variants associated with the cortical global surface area showed significant partitioned heritability enrichment in progenitor peaks, as well as the surface area of multiple cortical subregions including caudal anterior cingulate, entorhinal, lateral occipital, lingual and pericalcarine’ (Fig. 7b). Genetic variants associated with the thickness of the entorhinal cortex, but not average thickness across the entire cortex, also showed significant partitioned heritability enrichment in progenitor peaks. These results are consistent with the radial unit hypothesis, which posits that expansion of the neural progenitor pool in prenatal development leads to alterations in adult cortical surface area43.

To study the cell-type-specific gene-regulatory impact of genetic variants associated with neuropsychiatric disorders and brain structure traits, we performed a colocalization analysis of progenitor/neuron caQTLs with existing GWAS data (Supplementary Table 8). We identified overlapped signals (pairwise LD \( r^2 > 0.8 \) between the GWAS index and caQTL index) and then performed a conditional analysis to verify that the two variants mark the same locus (Methods). We found colocalized loci in neuropsychiatric disorders, including schizophrenia, MDD, neuroticism and bipolar disorder, as well as IQ and educational attainment (Fig. 7c). We also found colocalizations with global surface area and other brain structure associated loci. We found additional ASCA sites located within GWAS loci (Extended Data Fig. 8a and Supplementary Table 9). These results indicate that SNPs affect risk for brain-relevant traits and disorders by regulating the activity of REs in these two cell types during mid-fetal brain development, and provide a framework for exploring the mechanistic bases for these specific loci.

We next investigated regulatory mechanisms underlying colocalized loci using cell-type-specific caQTLs. Combining fetal cortical/cell-type-specific eQTL data, we found a colocalized locus across progenitor-specific caQTLs, fetal cortical eQTLs and MDD GWAS (Fig. 7d). We found more than 30 SNPs in high LD with an MDD GWAS index SNP (rs1950826). Eight of these variants were located in a caPeak (chr14:41,604,471–41,610,540). We prioritized one putatively causal SNP by testing for ASCA, finding that the A allele of the caSNP rs1950834 (protective allele for MDD), was associated with decreased accessibility of this caPeak in progenitors (Fig. 7e,f,h), which is consistent with luciferase reporter assay in previous work44. We also found the A allele of rs1950834 was associated with decreased expression of long-noncoding RNA (lncRNA) AL121821.1 (ENSG00000258636) in fetal cortex and progenitors (Fig. 7g). After conditioning on the MDD index SNP, rs1950826, the caQTL was no longer significant, indicative of colocalization (Fig. 7d). We found evidence to support that this SNP disrupts the binding of ETV1 (Fig. 7i and Methods)45. This suggests that the protective mechanism of this locus for MDD is via the protective allele at the caSNP disrupting ETV1 binding at an RE in progenitors, decreasing chromatin accessibility of this caPeak, and resulting in decreased expression of lncRNA AL121821.1.

As an additional example, we detected a colocalized locus between a neuron-specific caQTL and schizophrenia GWAS (Extended Data Fig. 8b). We found the C allele (schizophrenia protective allele) of rs9930307 was associated with decreased chromatin accessibility of a neuron caPeak (chr16:9,805,221–9,805,420) within an intron of GRIN2A (Extended Data Fig. 8c). This caSNP was also a neuron-specific ASCA site, providing further evidence of this allele’s impact on chromatin accessibility (Extended Data Fig. 8c). After conditioning on the schizophrenia index SNP (rs191183) in the caQTL analysis, the caSNP was no longer significant (Extended Data Fig. 8b). The motif of TP53 was disrupted by this caSNP (Extended Data Fig. 8d,e). Using a luciferase reporter assay, we found the C allele decreased the activity of this enhancer, which is consistent with the caQTL result (Extended Data Fig. 8f and Supplementary Table 10).

Discussion

Our caQTL analysis identified regulatory mechanisms underlying risk variants for neuropsychiatric disorders and brain-relevant traits. Currently, the function of individual noncoding brain-relevant risk loci is commonly understood through colocalization with eQTLs in adult postmortem brain tissue or chromatin interaction46,47. Our work is able to complement previous studies in several ways: (1) caQTLs allow fine mapping of causal variants

Fig. 7 | Cell-type-specific caQTLs lead to regulatory mechanisms underlying GWAS loci. a. Partitioned heritability enrichment. \( P \) values are estimated from LD score regression (two-sided test) and corrected by FDR (Methods). Data are presented as mean values ± s.e.m. IQ, intelligence quotient; ADHD, attention deficit hyperactivity disorder; IBID, inflammatory bowel disease. b. Partitioned heritability enrichment demonstrated a significant (FDR < 0.05) enrichment of heritability for surface area of the full cortex and other subregions within progenitor peaks. c. Numbers of colocalizations between caQTLs and GWAS loci. d. A colocalized locus between progenitor-specific caQTL and MDD GWAS. e. Association between rs1950834 and chromatin accessibility of the labeled peak in progenitors (\( n = 76 \)). \( P \) values are estimated by the mixed effects linear model using a two-sided test. f. ASCA of rs1950834 in progenitors. \( P \) values are estimated by the negative binomial generalized linear models from DESeq2 using a two-sided test (Methods). g. Association between rs1950834 and expression of IncRNA AL121821.1 in fetal brain (left, \( n = 235 \)) and progenitors (right, \( n = 85 \)). \( P \) values are estimated by the linear mixed effects model with a two-sided test. h. Zoomed in plot of caPeaks colored by genotype at rs1950834. i. The expression of TFs in which motifs are disrupted by rs1950834. j. The motif logo of ETV1; the boxed region is disrupted by rs1950834. For box plots in e and g, the center of the box is median of the data, the bounds of the box are 25th percentile and 75th percentile of the data, and the whisker boundary is 1.5 times the IQR. Maximum and minimum are the maximum and minimum of the data.

950
within LD-blocks by identifying putatively causal variants within peaks; (2) cell-type-specific caQTLs can prioritize cell types mediating the risk for neuropsychiatric illness because genetic effects on REs are highly cell-type-specific; (3) most previous eQTL studies have been performed in postmortem adult brain cortex\(^1\,\text{6,14}\), but cell types contributing to the heritability for multiple disorders and traits are not present at this time period suggesting that temporal specificity matters for understanding risk for these
disorders and (4) integration of caQTL, eQTL and brain-trait GWAS allows a more complete understanding of regulatory mechanisms leading to risk for neuropsychiatric disorders, where noncoding genetic variants disrupt TF binding to REs, affecting chromatin accessibility, influencing expression of genes, leading to downstream risk for neuropsychiatric disorders. While we prioritized TFs driving the regulatory effects based on motif disruption and expression in the cell types of interest, further experimental validation using complementary techniques (for example, chromatin immunoprecipitation–sequencing, ChIP-seq) is necessary to determine whether the caSNP does disrupt binding of the prioritized TF. Our study provides a resource to understand the impact of genetic variation on gene regulation during human cortical neurogenesis and provides an additional layer of information to explain the function of common variants associated with risk for neuropsychiatric illness and brain-related traits.

We found schizophrenia risk variants are enriched in progenitor REs (Fig. 7a), consistent with previous postmortem human studies, but contrary to a mouse gene expression study that found enrichments in neuronal but not progenitor cell types. Given that the mouse study may miss human specific REs and assigns variants to genes by proximity, we believe that our findings in combination with previous literature suggest that genetic alterations in both human progenitor and adult neuronal REs contribute to risk for schizophrenia.

We provide evidence to support that caQTLs have higher effect sizes and more cell-type-specificity than eQTLs in the two cell types we measured. This suggests that there are a limited number of mechanisms whereby genetic variation affects accessibility, including TF binding to DNA, whereas there are considerably more mechanisms by which variation can affect transcript levels, such as altering TF binding, affecting methylation or altering microRNA binding sites. This also suggests that caQTL analyses will identify more genetic variants involved in gene regulation than eQTLs given a limited sample size. However, because our comparison was conducted between only two cell types, other cell-type-specific caQTL and eQTL studies will be necessary to confirm the higher power and cell-type-specificity of caQTLs more broadly.

caQTL analysis is able to prioritize causal variants associated with REs, but cannot be used directly to predict the genes regulated by these elements. Most caQTLs did not result in changes in gene expression in either cell types or in bulk fetal cortical tissue. Previous work has suggested that multiple transcription factors, including those that translocate to the nucleus after response to an external stimulus, are needed to change gene expression levels at certain loci. caQTLs may therefore be more likely to colocalize with risk loci even in the absence of external stimuli (context independence), whereas eQTLs would require additional stimuli (context dependence). This also suggests that future work identifying ca/eQTLs in response to environmental stimuli relevant to neural proliferation, differentiation or function will be especially useful for interpreting GWAS risk loci.

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/s41593-021-00858-w.

Received: 4 February 2020; Accepted: 15 April 2021; Published online: 20 May 2021

References
1. Grasby, K. L. et al. The genetic architecture of the human cerebral cortex. Science 367, eaay6690 (2020).
2. Sullivan, P. F. & Geschwind, D. H. Defining the genetic, genomic, cellular, and diagnostic architectures of psychiatric disorders. Cell 177, 162–183 (2019).
3. Barei, A., Nash, A. J., Dahoun, T., Howes, O. & Lenhard, B. Understanding the genetics of neuropsychiatric disorders: the potential role of genomic regulatory blocks. Mol. Psychiatry 25, 6–18 (2019).
4. Gamaizón, E. R. et al. Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. Nat. Genet. 50, 956–967 (2018).
5. Lee, P. H. et al. Principles and methods of in-silico prioritization of non-coding regulatory variants. Hum. Genet. 137, 15–30 (2018).
6. Albert, F. W. & Kruglyak, L. The role of regulatory variation in complex traits and disease. Nat. Rev. Genet. 16, 197–212 (2015).
7. Kumasaka, N., Knights, A. J. & Gaffney, D. J. High-resolution genetic mapping of putative causal interactions between regions of open chromatin. Nat. Genet. 51, 128–137 (2019).
8. Davis, C. A. et al. The encyclopedia of DNA elements (ENCODE): data portal update. Nucleic Acids Res. 46, D794–D801 (2018).
9. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74 (2012).
10. GTEx Consortium. et al. Genetic effects on gene expression across human tissues. Nature 550, 204–213 (2017).
11. Wang, D. et al. Comprehensive functional genomic resource and integrative model for the human brain. Science 362, eaa464 (2018).
12. eChENCIEC Consortia, et al. The PsychENCODE project. Nat. Neurosci. 18, 1707–1712 (2015).
13. Won, H. et al. Chromosome conformation elucidates regulatory relationships in developing human brain. Nature 538, 523–527 (2016).
14. Walker, R. L. et al. Genetic control of expression and splicing in developing human brain informs disease mechanisms. Cell 179, 750–771.e22 (2019).
15. de la Torre-Ubieta, L. et al. The dynamic landscape of open chromatin during human cortical neurogenesis. Cell 172, 289–304.e18 (2018).
16. Broyios, J. et al. Evaluation of chromatin accessibility in prefrontal cortex of individuals with schizophrenia. Nat. Commun. 9, 3121 (2018).
17. Schwartzentzruber, J. et al. Molecular and functional variation in iPSC-derived sensory neurons. Nat. Genet. 50, 54–61 (2018).
18. Stein, J. L. et al. A quantitative framework to evaluate modeling of cortical development by neural stem cells. Neuron 83, 69–86 (2014).
19. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods 10, 1213–1218 (2013).
20. Hansen, D. V., Lui, J. H., Parker, P. R. L. & Kriegstein, A. R. Neurogenic radial glia in the outer subventricular zone of human neocortex. Nature 464, 554–561 (2010).
21. Pollen, A. A. et al. Molecular identity of human outer radial glial cell of the developing neocortex. Cell 163, 55–67 (2015).
22. Polioudakis, D. et al. A single-cell transcriptomic atlas of human neocortical development during mid-gestation. Neuron 103, 785–801.e8 (2019).
23. Kang, H. M. et al. Variance component model to account for sample structure in genome-wide association studies. Nat. Genet. 42, 348–354 (2010).
24. Yang, J., Zaitlen, N. A., Goddard, M. E., Vischer, P. M. & Price, A. L. Advantages and pitfalls in the application of mixed-model association methods. Nat. Genet. 46, 100–106 (2014).
25. Pickrell, J. K. et al. Understanding mechanisms underlying human gene expression variation with RNA sequencing. Nature 504, 768–772 (2010).
26. Avigan, N. et al. Genetic influences on cell type specific gene expression and splicing during neurogenesis elucidate regulatory mechanisms of brain traits. Preprint at bioRxiv https://doi.org/10.1101/2020.10.21.349019 (2020).
27. Huang, Q. Q., Ritchie, S. C., Brozynska, M. & Inouye, M. Power, false discovery rate and Winner's Curse in eQTL studies. Nucleic Acids Res. 46, e133 (2018).
28. Gate, R. E. et al. Genetic determinants of co-accessible chromatin regions in activated T cells across humans. Nat. Genet. 50, 1140–1150 (2018).
29. Loeb-Hennard, C., Cousin, X., Prengel, I. & Kremmer, E. Cloning and expression of the rat homolog of the human neurotrophin-3 receptor VNT1. Nucleic Acids Res. 28, 1707–1712 (2000).
30. Feng, L., Hatten, M. E. & Heintz, N. Brain lipid-binding protein (BLBP): a novel signaling system in the developing mammalian CNS. Neuron 12, 895–908 (1994).
31. Hsu, Y.-C. et al. Brain-specific 1B promoter of FGF1 gene facilitates the isolation of neural stem/progenitor cells with self-renewal and multipotent capacities. Dev. Dyn. 238, 302–314 (2009).
32. Ballas, N., Grunseich, C., Lu, D. D., Speh, J. C. & Mandel, G. REST and its corepressors mediate plasticity of neuronal gene chromatin throughout neurogenesis. Cell 121, 645–657 (2005).
33. Pastinen, T. Genome-wide allele-specific analysis: insights into regulatory variation. Nat. Rev. Genet. 11, 533–538 (2010).
34. Heinz, S., Romanoski, C. E., Benner, C. & Glass, C. K. The selection and function of cell type-specific enhancers. Nat. Rev. Mol. Cell Biol. 16, 144–154 (2015).
35. Diedenhofen, B. & Musch, J. cocor: a comprehensive solution for the statistical comparison of correlations. PLoS ONE 10, e0121945 (2015).
36. Behera, V. et al. Exploiting genetic variation to uncover rules of transcription factor binding and chromatin accessibility. Nat. Commun. 9, 782 (2018).
37. Bergsland, M., Werme, M., Malewicz, M., Perlmann, T. & Muhr, J. The establishment of neuronal properties is controlled by Sox4 and Sox11. Genes Dev. 20, 3475–3486 (2006).
38. Pattabiraman, K., Shibata, M., Lorente Galdos, B., Andrijevic, D. & Sestan, N. Regulation of prefrontal patterning, connectivity and synaptogenesis by retinoic acid. Biol. Psychiatry 87, S132 (2020).
39. Tsunemoto, R. et al. Diverse reprogramming codes for neuronal identity. Nature 557, 375–380 (2018).
40. He, X. et al. Expression of a large family of POU-domain regulatory genes in mammalian brain development. Nature 340, 35–41 (1989).
41. Wang, H. et al. ZEB1 represses neural differentiation and cooperates with CTBP2 to dynamically regulate cell migration during neocortex development. Cell Rep. 27, 2335–2353.e6 (2019).
42. Rakic, P. Specification of cerebral cortical areas. Science 241, 170–176 (1988).
43. Li, S. et al. Regulatory mechanisms of major depressive disorder risk variants. Mol. Psychiatry 25, 1926–1945 (2020).
44. Dobbyn, A. et al. Landscape of conditional eQTL in dorsolateral prefrontal cortex and co-localization with schizophrenia GWAS. Am. J. Hum. Genet. 102, 1169–1184 (2018).
45. Li, M. et al. Integrative functional genomic analysis of human brain development and neuropsychiatric risks. Science 362, eaat7615 (2018).
46. Skene, N. G. et al. Genetic identification of brain cell types underlying schizophrenia. Nat. Genet. 50, 825–833 (2018).
47. Gaffney, D. J. et al. Dissecting the regulatory architecture of gene expression QTLs. Genomic Biol. 13, R7 (2012).
48. Bell, J. T. et al. DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. Genome Biol. 12, R10 (2011).
49. Chen, K. & Rajewsky, N. Natural selection on human microRNA binding sites inferred from SNP data. Nat. Genet. 38, 1452–1456 (2006).
50. Alasoo, K. et al. Shared genetic effects on chromatin and gene expression indicate a role for enhancer priming in immune response. Nat. Genet. 50, 424–431 (2018).

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 America, Inc. 2021
Methods

Tissue acquisition and culture of PHNPCs. Human fetal brain tissue was obtained from the UCLA Gene and Cell Therapy Core following institutional review board regulations. The tissue is often fragmented during acquisition from the surgical specimen. In a following laboratory of D. Gerstner, flat, thin pieces of tissue that have the morphology of developing cortex were selected, and in some cases the tissue was sufficiently intact to be certain of cortical identity. Presumed cortical tissue from 14–21 gestation weeks was dissociated into a single cell suspension, cultured as neurospheres, plated for a low number of passages (2.5 ± 1.8 s.d.) on laminin/fibronectin and polyornithine coated plates, and then cryopreserved as HNPs following our previous work.15

Cryopreserved HNPs were shipped to UNC Chapel Hill after a signed material transfer agreement by both institutions. All proliferation, differentiation, sorting, library preparation, and analysis were performed at UNC Chapel Hill (following institutional review board regulations under the Office of Human Research Ethics). In total, HNPs from 2 donors were cultured (34% are female and 66% are male) and after quality control, we retained 87 donors for analysis. Donors were thawed in ‘rounds’ of approximately ten donors, so as to create a manageable workload of cell culture (Extended Data Fig. 1a). Donors were randomly assigned into groups and thawed 3 weeks apart. We performed specific experimental events on the same day of the week and had the same interval of time between events for each round. Experimental events included thawing cells, feeding cells, splitting cells, counting and plating cells, washing cells before differentiation, coating plates with attachment factors, adding virus, lifting cells for sorting, sorting and ATAC-seq library preparation. As much as possible, the same cells, feeding cells, splitting cells, counting and plating cells, washing cells before time between events for each round. Experimental events included thawing 9500, 1% glutamax (100–9500), PenStrom (Thermo Fisher, catalog no. A-11039, lot no. 1759025), Goat anti-RT AF647 (Thermo Fisher, catalog no. A-11011, lot no. 1820020), goat anti-RB AF568 (Thermo Fisher, catalog no. A-11039, lot no. 1759025), Goat anti-RT AF647 (Thermo Fisher, catalog no. A-11014, lot no. 1821166), goat anti-RA AF568 (Thermo Fisher, catalog no. A-11011, lot no. 1820020), goat anti-CH AF488 (Thermo Fisher, catalog no. A-11039, lot no. 1759025), anti-RB AF488 (Thermo Fisher, catalog no. A-11014, lot no. 1821166), goat anti-RA AF568 (Thermo Fisher, catalog no. A-11011, lot no. 1939600) and goat anti-MA AF568 (Thermo Fisher, catalog no. A-11031, lot no. 2026148).

Library preparation for HNPs and neurons. Library preparation was conducted using the published ATAC-seq protocol52. ATAC-seq libraries were prepared immediately following cellular dissociation. Progenitor nuclei were counted using a hemocytometer while neuron nuclei were counted during sorting. Then, 50,000 nuclei were aliquoted into the first step of the ATAC-seq published protocol. Libraries were performed following the protocols and instructions supplied with the ATAC-seq kit. For library prep step was modified to use KAPA pure beads (AmplureXP beads at a 1:1 ratio to remove DNTPs, salts, primers or primer dimers) instead of Qiagen Minelute clean-up kit. All libraries were sequenced to a minimum depth of 13.6 million read pairs and an average depth of 25.5 million read pairs using 50bp PE sequencing on an Illumina HiSeq2500 or MiSeq (Extended Data Fig. 1b). In total, we acquired 98 ATAC-seq libraries from progenitors (n_atac = 85, n_atac_repl = 13) and 70 ATAC-seq libraries from neurons (n_atac = 66, n_atac_repl = 4).

Recording technical variables and randomization. To reduce the impact of batch effects on interpretation of our results, we attempted to either have no batches when possible (for example, perform all experiments using the same lot of a reagent) or when this was not possible, randomize technical variables (the round a donor was thawed in, sequencing pool) such that they had minimal correlation with variables of interest. To extensively document the impact of technical variables on our results, we performed a relational MySQL database to keep track of many technical and biological variables throughout each experimental event. Each downstream ATAC-seq library preparation therefore is able to be traced back to all technical and biological variables associated with its cell culture. The variables recorded were as follows:

Media: basal media lots, growth factor lots, supplement lots and antibiotic lots. Virus: lot number. Donor: sex inferred from genotype and gestation week. Culture: passage, round, thaw date, each split date, split ratio, trypsin lot, PBS lot, polyornithine lot, fibronectin lot, plate and well position, cells per well, date of virus addition, differentiation time, date of differentiation media addition, person plating for differentiation, virus used, person performing splits, person performing sampling, Fisher test, fixed cells used, number of live cells, number of times pipetting up and down during transposase reaction, transposase reaction volume, barcode indices used for multiplexing of each sample, number of PCR cycles added in the ATAC-seq protocol and final DNA concentration after library preparation complete. Sequencing: sequencing date, sequencing company, type of sequencer and read length.

Randomization was performed multiple times. First, randomization was performed to assign each donor to a thawing ‘round’. Randomization was performed at this stage by randomly ordering all donors and selecting those to go in each round (generally about ten donors per round). After culture and library

Immunocytochemistry. HNPs were plated onto polyornithine/laminin coated German glass coverslips at roughly 80,000 cells per cm². Proliferation and 8-week differentiated cultures were fixed with 4% paraformaldehyde in parallel with the ATAC-seq protocol and final DNA concentration after library preparation complete. Coverslips were treated with the following primary antibodies in 0.02% Tween-20/PBS overnight at 4°C: Pax6 (1:300, BioLegend, catalog no. 901501, lot no. B2771601), Sox2 (1:500, Cell Sciences, catalog no. RA0979, lot no. 5050), Sox1 (1:500, EMD Millipore, catalog no. AB5603, lot no. 3187396), GFP (1:500, Millipore, catalog no. AB16901, lot no. 27221295), HOPX (1:1,000, Sigma-Aldrich, catalog no. HPA030180, lot no. C0105752), Nkx2.1 (1:500, Millipore, catalog no. MAB5460, lot no. 3074948), Tbr2 (1:300, eBioscience, catalog no. 14-4877-82, lot no. 2042007), Gad67 (1:500, EMD Millipore, catalog no. MAB5460, lot no. 3015328), TUBB3 (1:1,000, BioLegend, catalog no. 901329, lot no. B284869), SAG2 (1:500, EMD Millipore, catalog no. sc-81376, lot no. 132317) and Ctip2 (1:500, Abcam, catalog no. ab18465, lot no. GR3242845-3). Coverslips were then treated with the following secondary antibodies for 1 h at room temperature at 1:1,000 dilution: goat anti-RB AF488 (Thermo Fisher, catalog no. A-11034, lot no. 1812166), goat anti-BAF568 (Thermo Fisher, catalog no. A-11011, lot no. 1820020), goat anti-CH AF488 (Thermo Fisher, catalog no. A-11039, lot no. 1759025), goat anti-RT AF647 (Thermo Fisher, catalog no. A-21247, lot no. 2119156), goat anti-MS AF568 (Thermo Fisher, catalog no. A-11001, lot no. 1939600) and goat anti-MA AF568 (Thermo Fisher, catalog no. A-11031, lot no. 2026148).

Library preparation was conducted using the published ATAC-seq protocol52. ATAC-seq libraries were prepared immediately following cellular dissociation. Progenitor nuclei were counted using a hemocytometer while neuron nuclei were counted during sorting. Then, 50,000 nuclei were aliquoted into the first step of the ATAC-seq published protocol. Libraries were performed following the protocols and instructions supplied with the ATAC-seq kit. For library prep step was modified to use KAPA pure beads (AmplureXP beads at a 1:1 ratio to remove DNTPs, salts, primers or primer dimers) instead of Qiagen Minelute clean-up kit. All libraries were sequenced to a minimum depth of 13.6 million read pairs and an average depth of 25.5 million read pairs using 50bp PE sequencing on an Illumina HiSeq2500 or MiSeq (Extended Data Fig. 1b). In total, we acquired 98 ATAC-seq libraries from progenitors (n_atac = 85, n_atac_repl = 13) and 70 ATAC-seq libraries from neurons (n_atac = 66, n_atac_repl = 4).

All libraries were sequenced to an average depth of 25.3 ± 21.3 s.d.) million read pairs (Extended Data Fig. 1b), which resulted in an average depth of 14 (± 4.8 s.d.) million reads per sample after filtering for mitochondrial contamination and duplicates. We performed a sensitivity analysis for read depth versus peak calling that showed greater than 15 million filtered read pairs per library led to a fewer number of new peaks called, indicating a reasonable balance between read depth and peaks called on the libraries generated here (Extended Data Fig. 2a).
ATAC-seq data preprocessing. Sequencing reads were first quality controlled via fastq (v0.10.17) to check for sequence quality in each library. We observed high-quality sequencing for all libraries (PHRED > 20, average duplication rate of 43.07%), which is almost entirely mitochondrial DNA contamination (Extended Data Fig. 1e), which is in agreement with previous studies using the same ATAC-seq method1, and average GC content of 45% (http://www.biosinformatics. babraham.ac.uk/projects/fastq/). Sequencing adapters were removed using BBRMAP/BBDDUK (https://cgi.doc.gov/data-and-tools/bbttools/bb-tools-user-guide/ bbduk-guide/). We also calculated the number of total reads, the number of unique nonmitochondrial reads, duplicate rate, mitochondrial duplicate rate, TSS enrichment score and FRIP score (the fraction of reads in called peak regions) for neuron samples and progenitor samples to check the library quality using atacqv53. Then, sequencing reads were mapped to the human genome including decoy sequences (GRCh38/hg38) using bwa mem (v0.7.17) and WASP9 to remove mapping bias at any bi-allelic SNP using imputed genotype data from each sample. Duplicate reads were then removed using WASP. Only uniquely mapped reads mapping to chr 1–22, X and Y were kept (mitochondrial genome and unmapped contaminants were removed) using samtools (v1.3.1) (gnu/bio/encodeDCC/wgEncodeMapability/wgEncodeDacMapabilityConsensusExcludable. bed.gz, converted to hg38 using UCSCtools/liftOver (v320)), then were removed by bedtools (v2.26.3).

We did a sensitivity analysis for peak calling using preprocessed BAM files. It showed acquiring 9 × n times for a given cell type were excluded. Peaks were called for these selected samples. Significant correlations with technical variables were removed after batch correction. The batch corrected number of reads within CSAW peaks and the linear regression between the (no. of bases in state AND overlap feature)/no. of bases in genome) and the ((no. of bases overlap feature)/no. of bases in genome) × (no. of bases in state)/no. of bases in genome) as described previously by the Roadmap Epigenomics Consortium16. The significance of this enrichment was calculated using a binomial test as in the GREAT algorithm17.

Differential TF binding analysis. We performed an analysis to identify motifs with differential prevalence in differentially accessible peaks (Supplementary Table 2, Extended Data Fig. 3b). To avoid bias caused by different numbers of progenitor peaks and neuron peaks, here we only used the top 2,000 progenitor peaks with the highest LFC and the top 2,000 neuron peaks with the lowest LFC.

Potential TF binding sites were called in the human genome using TFBSTools (v1.4.0) with a minimum score threshold of 80% based on position weight matrices from the JASPAR201618 dataset. Only the most recent version of the position weight matrix for a given TF was used. To select regions of the genome that are highly conserved among vertebrates, and likely to be functional, 100-wp-threshold was applied to regions >200 bp were saved (downloaded from UCSC genome browser). Only transcription factor binding sites (TFBSs) within conserved regions were retained for further analyses. We carried out differential motif enrichment analysis using a logistic regression model to identify motifs present more often in progenitor peaks as compared to neuron peaks, or vice versa. Logistic regression explicitly controlled for differences in peak width and peak conservation between progenitor and neuron differentially accessible peaks. The analysis was implemented in R as: glm(TFBS ~ ProgenitorNeuron + peak width + conservedbppercent, family = binomial). The dependent variable (TFBS) was a binary representation of whether each differentially accessible peak contained a motif of a TF or not. The independent variable of interest marked whether a peak was progenitor (ProgenitorNeuron = 1) or neuron (ProgenitorNeuron = 0). Other covariates included peak width (peak width) and the percentage of the peak with conservation (conservedbppercent) as defined above. Significant differential motif enrichment was determined by FDR adjusted P < 0.05 threshold of the ProgenitorNeuron covariate. progenitor1 TFs were defined as significantly differentially enriched motifs present more often in progenitor peaks as compared to neuron peaks, whereas neuron1 TFs were defined as significantly differentially enriched motifs present more often in neuron peaks as compared to progenitor peaks.

Genotype preprocessing. Genotyping was performed using the Illumina Human Omni2.5+ HumanOmni2.5+ or HumanOmniExpress platform. SNP markers were exported into PLINK format. SNP marker names were converted from Illumina KGP IDs to rsIDs using the conversion file provided by Illumina. Quality control was performed in PLINK v1.90b3 (ref. 20) (Extended Data Fig. 4a). SNPs were filtered based on Hardy–Weinberg equilibrium (–hwe 1e-6), minor allele frequency (–maf 0.01), individual missing genotype rate (–mind 0.10), variant minor allele frequency rate (–pvalue 0.05) resulting in 1,790,704 directly genotyped SNPs. Multidimensional scaling (MDS) analysis of genotypes from all individuals used in the study was completed in PLINK v1.90b3.
a strong effect of genotyping batch on genotype data based on MDS1 and MDS2 from different genotyping batches. We used PLINK v.1.90b3 to call sex from genotype data. For the samples with unknown sex from genotype data, we plotted PCA (PC1 versus PC2) of ATAC-seq and genotyping data (Extended Data Fig. 4b). We removed any BAM files with [FREEMIX] > 0.02 or [CHPMIX] > 0.02 (nmiss = 5), and corrected sample swaps (nmiss = 7). After this filtering step, our sample size was 76 unique donors for progenitor samples and 61 unique donors for neuron samples for the caQTL studies.

Imputation. Filtered genotype data were prephased by SHAPEIT\(^1\) \(\times 2.837\). Minimac4 (ref. \(^2\)) (v1.0.0) was used to impute the filtered genotyped markers using reference haplotype panels from the 1,000 Genomes Project (The 1,000 Genomes Project Consortium Phase 3) that contain a total of 37.9 million SNPs in 2,504 individuals from any ancestry, including those from West Africa, East Asia and Europe. We separated chrX into pseudoautosomal regions and nonpseudoautosomal regions, then prephased and imputed them separately.

After genotype imputation, we extracted the genotypes for all individuals assayed for chromatin accessibility. Imputed genotype data were filtered for variant missing genotype rate \(< 0.05\), Hardy–Weinberg equilibrium \(P\) \(< 1 \times 10^{-10}\) and minor allele frequency \(\leq 1\%\). We retained variants with imputation quality R\^{2}quared \(> 0.3\) by Minimac4, resulting in roughly 13.6 million SNPs.

caQTL mapping. We calculated MDS for genotype data of our samples and genotype data from HapMap3 (https://www.sanger.ac.uk/resources/downloads/human/hapmap3.html) following the protocol from ENIGMA consortium (http://enigma.ini.usc.edu/wp-content/uploads/2012/07/ENIGMA2_1KGP_cookbook_v3.pdf). We identified multiple ancestries of donors of our samples in the MDS plot (MDS1 versus MDS2) (Extended Data Fig. 4c).

To control for population stratification and cryptic relatedness of our samples when mapping caQTLs, we ran caQTL analysis with EMMAX\(^3\), which accounts for population structure using a genetic relatedness or kinship matrix. We used the emmax-kin function \((v \cdot h \cdot s \cdot d \leq 10)\) to create the IBS kinship matrix for each tested genetic variant from nonimputed genotype data excluding all genetic variants on the same chromosome with the tested genetic variant.\(^4\)

We performed proximal caQTL mapping using a window of 100 kb up- and downstream of each of the caPeaks using VST normalized read counts of each peak for each donor (Extended Data Fig. 4a). We performed caQTL analysis separately in neurons and progenitors using imputed genotype data. To prevent results driven by only one minor allele homozygous donor, we retained the variants where the number of minor allele homozygous donors is not one and there are at least two heterozygous donors. In addition to the kinship matrix, for the progenitor caQTLs, we include sorted locations, the first ten genotype MDSs and eight principal components across VST counts of the chromatin accessibility data. For neurons, we include the first ten genotype MDSs and seven principal components of VST counts of the chromatin accessibility data. These principal component numbers were chosen to maximize the number of caQTLs for each cell type. Nominal EMMAX \(P\) values were corrected for multiple testing using the software eigenMT\(^5\) and Benjamini–Hochberg FDR correction \((\leq 0.05)\) within neuron caQTLs and within progenitor caQTLs separately (eigenMT–FDR < 0.05). We retained the most significant caSNP for each significant caPeak that survives the eigenMT–FDR threshold and defined the caSNP–caPeak pair as the most significant caQTL for a given caPeak. The percentage variance explained was calculated using the method from a previous study.\(^6\)

Identify correlated caPeaks. To identify correlated caPeaks, we defined primary caPeaks as the caPeaks harboring caSNP(s). We then defined secondary caPeaks as peaks that are associated with the caSNP of a primary peak. We calculated Pearson correlation values between all the primary caPeaks within a +2 Mbp from the center of its secondary caPeak (including the secondary caPeak), then corrected the Pearson’s correlation value using the Benjamini–Hochberg FDR correction.\(^1\) If the secondary caPeak was significantly (FDR < 0.05) correlated with the primary caPeak, this caSNP–caPeak pair was classified as ‘caSNP in correlated caPeak’.

ASCA. We used tools from the Genetic Analysis Toolkit to extract allele specific read counts for every SNP. We first filtered for SNPs within each donor that had sufficient read depth by retaining SNPs with total counts greater than or equal to 10 for neuron and progenitor samples, separately. Then to calculate ASCA, we retained those SNPs with average read counts for all heterozygous donors greater than or equal to 15. Finally, we retained only those SNPs that meet these previous thresholds for at least five heterozygous donors. DEmseq was used to calculate the LFC (alternative read counts/reference read counts) for filtered SNPs across all heterozygous donors. The non-heterozygous donors were excluded from the differential analysis for each SNP using sample-specific weights, and maximum likelihood estimation was used for dispersion estimation followed by Wald tests of the estimated LFC. FDR < 0.05 was used as the threshold for significance.

Bulk fetal brain eQTL mapping. Bulk fetal cortical wall eQTL data described in a previous publication,\(^8\), was reanalyzed in this study with the following modifications: (1) we used a linear mixed model implemented in EMMAX to more stringently control for population stratification, and (2) we add 23 more donors to the analysis because these donors were genotyped after the publication of the genotype data. RNA-seq data (Ribo-nome RNA-seq and RNA-seq) data from flash frozen human fetal brain cortical wall tissues derived from 235 donors at 14–21 gestation weeks were used for eQTL analysis. In the progenitor eQTL analysis, 41% samples were overlapped with the samples in fetal brain eQTL analysis, and 36% samples in the neuron eQTL analysis were overlapped with the samples in fetal brain eQTL analysis. Gene–based annotations of the genome were derived from H. sapiens gene ensemble v92 (GRCh38) for eQTLs. Only genes that are expressed in more than 5% of donors with at least ten counts were included in the analysis. VST normalized expression values were used as phenotypes for eQTL analysis. Genomic DNA from human fetal brain cortical wall tissues derived from 235 donors at 14–21 PCW was extracted. Each donor tissue was genotyped on a dense array (Illumina Omni Methylation) and imputed using a reference panel (1,000 Genomes, described above). Variants were retained in the analysis if there were at least two heterozygous donors and no homozygous minor allele donors, or if there were at least two minor allele homozygous donors. For the effect size comparison analysis of fetal brain eQTL versus caQTLs (Fig. 2e,f), we subsampled fetal brain eQTLs to the same sample size as the caQTL while maintaining the population composition similar to the larger donor pool.

We conducted cis-eQTL analysis by evaluating association between each gene’s expression and variants within ±1 MB window of transcription start site of each gene by implementing linear mixed model association software, EMMAX.\(^\) All markers on the chromosome of this candidate marker were excluded from the IBS kinship matrix generated with emmax-kin-function \((v \cdot h \cdot s \cdot d \leq 10)\), and added as a random variable into linear mixed model for association test. In addition to kinship matrix, ten MDS components of genotype and first ten principal components of gene expression were included into the covariate matrix. After association, nominal \(P\) values were corrected for multiple testing using the eigenMT and Benjamini–Hochberg FDR correction, and associations with lower than a 5% eigenMT–FDR threshold value were accepted as significant. We retained only the most significant eSNP for each significant eGene in this study.

\(M\) value calculation. Using MetaSoft (v2.0.1),\(^\) we calculated \(m\)-values between caQTLs and eQTLs. First, we selected the most significant caSNP-a given caPeak in either neurons or progenitors. Then, we found the SNPs–gene corresponding to that caSNP in bulk fetal brain or in the cell-type-specific eQTL (for sharing in Fig. 2d). The caSNP may or may not be an eSNP, and the eSNP may be associated with multiple genes. Then, we selected the most significant eSNP for an eGene and found the corresponding SNPs–peak pairs in neuron/ progenitor caQTL analysis (for sharing in Fig. 2f). We subsampled fetal brain eQTLs to the same sample size as the caQTL while maintaining the population composition similar to the larger donor pool.

Overlap of caQTLs with eQTLs. To identify RE–gene pairs in fetal brains, we listed SNPs with pairwise \(D^2 > 0.8\) with the caSNPs in the caPeak using genotype data from neuron/progenitor samples, separately, then we listed SNPs with pairwise \(D^2 > 0.8\) with index eSNP using the LD matrix from neuron samples/progenitor samples for cell-type-specific eQTL analysis. We labeled the caPeak and the eGene as an RE–gene pair if any SNP from the above two categories is overlapped in neurons/progenitors (Supplementary Table 5).

Estimation of sharing via \(\pi\), The R package ‘qvalue’ (v2.20.0)\(^9\) was used to estimate the \(\pi\) of the input nominal \(P\) values of the cell-type-specific eQTL and caQTL data, then we used one minus the estimated \(\pi\) to get \(\pi\). We found all the nominal SNPs–peak pairs that corresponded to the most significant neuron–caPeak–caSNP–caPeak pairs, then used the nominal \(P\) values of the neuron–SNP–peak pairs to estimate the proportion of true neuron caQTLs in the SNPs–peak pairs \((\pi)\). In the same way, we estimated the proportion of true progenitor caQTLs using the most significant neuron caSNP–caPeak pairs. For neuron eQTLs, we listed all the neuron SNPs–gene pairs using the most significant progenitor eSNP–eGene pairs, then used
the nominal P values of the neuron SNP–gene pairs to estimate the proportion of true neuron eQTLs in the SNP–gene pairs (\(\pi\)). Similarly, we estimated the proportion of true progenitor eQTLs using the most significant neuron eSNP–eGene pairs.

Comparison to adult DLPFC caQTls. We acquired adult DLPFC ATAC-seq data from Sage Bionetworks-Synapse website via the psychENCODE Knowledge Portal under the accession number syn5321694, ref. \(^[41]\). To calculate the overlap of caQTls between cultured neural cells and adult DLPFC, we first extracted read counts within adult DLPFC peaks in ATAC-seq data from neurons and progenitors. We found 65,573 DLPFC peaks have an average read counts greater than five across all neurons and progenitor samples, and these peaks demonstrate cell-type/tissue-type specificity in chromatin accessibility as visualized in a PCA plot (Extended Data Fig. 7a,b). Then, using the shared peaks and the same SNPs with DLPFC caQTls, we remapped caQTls in neurons and progenitors using the same models as previously described. We found 27 significant neuron caQTls and 35 significant progenitor caQTls using the same eigenMT-BHFDR threshold as previously used in caQTL mapping in neurons and progenitors. Using the same SNP–peak pairs from DLPFC caQTls, we found the \(\pi\) is 0.001 in neuron caQTls and 0.04 in progenitor caQTls, which indicates a high temporal specificity of caQTls. We also found low correlations of effect sizes in significant neuron/progenitor caQTls and SNPs with pairwise LD \(r^2 > 0.8\) with index GWAS SNP (\(P < 5 \times 10^{-8}\) and exhibited the strongest association in upstream/downstream 100 kb from the center of this caPeak) using the LD matrix from European genotype data from 1,000 Genome Project phase 3 with population code EUR. Second, we labeled the caPeak as a potentially colocalized locus if any SNP was shared between the above two categories. Third, we performed a conditional caQTL analysis for significant (eigenMT-BH FDR < 0.05) caSNPs conditioning on the index GWAS SNP\(^[42]\). If the caQTL is no longer significant (eigenMT-BH FDR > 0.05), then we called the caQTL a colocalized locus with this GWAS trait.

 Luciferase reporter assay. DNA fragments of differentially accessible chromatin peaks containing SNPs for functional validation were synthesized using Thermo Fisher Scientific’s Gene String service. Fragments were amplified by PCR with primers containing KpnI and HindIII restriction sites. Digested fragments were then cloned into the multiple cloning region of the pGL3.23 vector (Promega), containing a minimal promoter upstream of the luc2 luciferase reporter gene. To generate corresponding alternate alleles, we performed site-directed mutagenesis on the cloned, insert-containing luciferase plasmids using the Q5 site-directed mutagenesis kit (NEB). All cloned sequences were verified by Sanger sequencing for the correct mutations and analyzed by NanoDrop to ensure high concentration and transfection-grade quality. Oligonucleotide sequences used for cloning are listed in Supplementary Table 10.

HNPCs grown in 96-well plates were cotransfected with 120 ng per well of luciferase reporter plasmid and 30 ng per well of renilla control plasmid (pRL-SV40 Promega). Two days posttransfection, luciferase activity was measured using a CLARIOStar Plus Plate Reader (BMG Labtech). Each luciferase reading was then normalized by its corresponding renilla reading to control for transfection efficiency and to calculate RLU. A total of eight unique donors with at least three well replicates per plasmid per donor were used for analysis.

REST ChIP–seq data in H1 cells and neurons differentiated from H1 cells. We acquired the alignments of REST ChIP–seq data in H1 embryonic stem cells and neurons differentiated from H1 cells from the ENCODE portal\(^[43]\) (https://www.encodeproject.org/) with the following identifiers: ENCSR000BTV and ENCSR000BHM. We normalized the read counts by library sizes then plotted the coverage using Gviz\(^[44]\).

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

**Data availability.** Data generated in this paper (including metadata) can be accessed via dbGaP at accession number phs001958.v1.p1. The RNA-seq and genotype datasets used for fetal cortical eQTL analysis are available at dbGaP with accession number phs001900. REST ChIP–seq data in H1 embryonic stem cells and neurons differentiated from H1 cells are available via the ENCODE portal (https://www.encodeproject.org/) with the following identifiers: ENCSR000BTV and ENCSR000BHM.

**Code availability.** All code used in this paper is deposited on bitbucket at https://bitbucket.org/steinlabins/celltypespecificcaqtls_wasp/src/master/.

**References.**

51. Thiel, G., Greengard, P. & Südhof, T. C. Characterization of tissue-specific transcription by the human synapsin I gene promoter. Proc. Natl Acad. Sci. USA 88, 3431–3435 (1991).
52. Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: a method for assaying chromatin accessibility genome-wide. Curr. Protoc. Mol. Biol. 109, 21.9.1–21.9.29 (2015).
53. Orchard, P., Kyono, Y., Hensley, J., Kitzman, J. O. & Parker, S. C. J. Quantification, dynamic visualization, and validation of bias in ATAC-seq data with ataqv. Cell Syst. 10, 298–306.e4 (2020).
54. Li, H. Alignment sequence. close sequences and assembly contigs with BWA-MEM. Preprint at arXiv https://arxiv.org/abs/1303.3997 (2013).
55. van de Geijn, B., McVicker, G., Gilad, Y. & Pritchard, J. K. WASP: allele-specific software for robust molecular quantitative trait locus discovery. Nat. Methods 12, 1061–1063 (2015).
56. L. H. et al. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079 (2009).
57. Quinlan, A. R. & Hall, I. M. BEDtools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842 (2010).
58. Feng, I., Liu, T. & Zhang, Y. Using MACS to identify peaks from ChIP–seq data. Curr. Protoc. Bioinformatics. 34, 21.14.1–21.14.14 (2011).
59. Lu, S. T. L. & Smyth, G. K. CSAW: a Bioconductor package for differential binding analysis of ChIP–seq data using sliding windows. Nucl. Acids Res. 44, e45 (2016).
60. Hansen, K. D., Irizarry, R. A. & Wu, Z. Removing technical variability in RNA-seq data using conditional quantile normalization. *Biosciences* 13, 204–216 (2012).

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

62. Ritchie, M. E. et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43, e47 (2015).

63. Roadmap Epigenomics Consortium, et al. Integrative analysis of 111 reference human epigenomes. *Nature* 518, 317–330 (2015).

64. McLean, C. Y. et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* 28, 495–501 (2010).

65. Ernst, J. & Kellis, M. Large-scale imputation of epigenomic datasets for systematic annotation of diverse human tissues. *Nat. Biotechnol.* 33, 364–376 (2015).

66. Alexa, A. & Rahnenfuehrer, J. topGO: Enrichment analysis for gene ontology. R Package Version 2 (2010).

67. Matheiler, A. et al. JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* 44, D110–D115 (2016).

68. Pollard, K. S., Hubisz, M. J., Rosenberg, K. M. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* 20, 110–120 (2010).

69. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Methods Mol. Biol.* 811, 335–351 (2012).

70. Jun, G. et al. Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. *Am. J. Hum. Genet.* 91, 839–848 (2012).

71. Delaneau, O., Marchini, J. & Zagury, J.-F. A linear complexity phasing method for thousands of genomes. *Nat. Methods* 9, 371–376 (2011).

72. Das, S. et al. Next-generation genotype imputation service and methods. *Genome Biol.* 15, 16–32 (2014).

73. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a new approach. *Biometrics* 44, 1268–1270 (1988).

74. Shim, H. et al. A multivariate genome-wide association analysis of 10 LDL subfractions, and their response to statin treatment, in 1868 Caucasians. *PLoS One* 6, e17909 (2011).

75. Davis, J. R. et al. An efficient multiple-testing adjustment for eQTL studies that accounts for linkage disequilibrium between variants. *Am. J. Hum. Genet.* 98, 216–224 (2016).

76. Price, A. L., Zhao, K., Reich, D. & Patterson, N. New approaches to population stratification in genome-wide association studies. *Nat. Rev. Genet.* 11, 459–463 (2010).

77. Davis, J. R. et al. An efficient multiple-testing adjustment for eQTL studies that accounts for linkage disequilibrium between variants. *Am. J. Hum. Genet.* 98, 216–224 (2016).

78. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B Stat. Methodol.* 57, 289–300 (1995).

79. Shim, H. et al. A multivariate genome-wide association analysis of 10 LDL subfractions, and their response to statin treatment, in 1868 Caucasians. *PLoS ONE* 10, e0120758 (2015).

80. Kang, H. M. et al. Efficient control of population structure in model organism association mapping. *Genetics* 178, 1709–1723 (2008).

81. Han, B. & Eskin, E. Interpreting meta-analyses of genome-wide association studies. *PLoS Biol.* 8, e1000255 (2010).

82. Dabney, A., Storey, J. D. & Warnes, G. R. qvalue: Q-value estimation for high-throughput genomic testing. *Bioinformatics* 28, 1353–1354 (2012).

83. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a new approach. *Biometrics* 44, 1268–1270 (1988).

84. Shim, H. et al. A multivariate genome-wide association analysis of 10 LDL subfractions, and their response to statin treatment, in 1868 Caucasians. *PLoS ONE* 10, e0120758 (2015).

85. Kang, H. M. et al. Efficient control of population structure in model organism association mapping. *Genetics* 178, 1709–1723 (2008).

86. Han, B. & Eskin, E. Interpreting meta-analyses of genome-wide association studies. *PLoS Biol.* 8, e1000255 (2010).

87. Dabney, A., Storey, J. D. & Warnes, G. R. qvalue: Q-value estimation for high-throughput genomic testing. *Bioinformatics* 28, 1353–1354 (2012).

88. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a new approach. *Biometrics* 44, 1268–1270 (1988).
Brain Development), S. Weissman (Yale University), K. White (University of Chicago), J. Willsey (University of California, San Francisco) and P. Zandi (Johns Hopkins University. We acknowledge the ENCODE Consortium and the ENCODE production laboratory(ies) generating the particular dataset(s).

Author contributions
J.L.S., D.H.G. and L.T.U. conceived the study. J.L.S. directed and supervised the study. J.L.S. along with D.H.G. provided funding. A.L.E., K.E.C., K.P.C., M.Y., L.T.U. and J.L.S. cultured HNP cells. A.L.E. performed library preparation. M.J.L. preprocessed the RNA-seq data for eQTL. N.A. performed eQTL analysis. O.K. performed immunocytochemistry. O.K., J.M.W., F.A.K. and D.L. performed the functional validation assays. M.E.G., A.A.-K. and G.E.C. provided access to adult dIPFC caQTL data. M.I.L. aided in ASCA methodology. D.L. performed preprocessing, differential accessibility, caQTL, ASCA, colocalization and motif analyses. J.L.S. and D.L. wrote the paper. All authors commented on and approved the final version of the paper.

Competing interests
The authors declare no competing interests.

Additional information
Extended data are available for this paper at https://doi.org/10.1038/s41593-021-00858-w.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41593-021-00858-w.

Correspondence and requests for materials should be addressed to J.L.S.

Peer review information Nature Neuroscience thanks Andrew Jaffe and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.
Extended Data Fig. 1 | Flowchart for cell culture and preprocessing of ATAC-seq data. a, Flowchart of cell culture for 17 rounds. b, The FACS gates for sorting EGFP + neurons. c, Images of immunofluorescence for cell markers in progenitor cultures. Immunolabeling experiments were repeated in at least 10 unique donor cell lines with similar results. The scale bar presents 100 μm. d, Images of immunofluorescence for cell markers in 8-week differentiated cultures. Immunolabeling experiments were repeated in at least 10 unique donor cell lines with similar results. The scale bar presents 100 μm. e, Box plot for total sequence depth (forward reads and reverse reads), unique read number (forward reads and reverse reads), duplicate rate, mitochondrial duplicate rate, TSS enrichment and the fraction of reads in called peak regions (FRiP score) in neurons (N = 61) and progenitors (N = 76) compared to previously published data (N_GZ = 3 biologically independent samples with 3-4 replicates for each sample, N_CP = 3 biologically independent samples with 3 replicates for each sample)15. The center of the box is median of the data, the bounds of the box are 25th percentile and 75th percentile of the data, and the whisker boundary is 1.5 times the IQR. Maximum and minimum are the maximum and minimum of the data.
Extended Data Fig. 2 | See next page for caption.
Extended Data Fig. 2 | ATAC-seq data QC. a, Peak calling versus library sequencing depth. We observed a slower rise in the number of new peaks called after 15 millions filtered read pairs. This indicates a reasonable balance between read depth and number of peaks called using an average of 14 million read pairs after filtering in our samples. b, Insert size histograms for 3 randomly selected neuron and progenitor samples. c, PCA plot for ATAC-seq data (N = 137) before batch correction (left) and after batch correction (right), colored by sorter. We corrected normalized reads within ATAC-seq peaks in neurons by sorter locations. Then, we corrected normalized reads within ATAC-seq peaks in neurons and progenitors by cell culture round. d, Correlations of batch corrected normalized reads across donors and within donors. Correlations within donors was significantly higher than correlations across donors in progenitor (n = 15). Correlations within donors were higher than correlations across donors in neurons (n = 4), but not significant (p = 0.07). P values are estimated by two-sided wilcoxon tests. The center of the box is median of the data, the bounds of the box are 25th percentile and 75th percentile of the data, and the whisker boundary is 1.5 times the IQR. Maximum and minimum are the maximum and minimum of the data. e, Correlations between PC1 to PC10 from normalized reads in neurons with known technical and biological factors. f, Correlations between PC1 to PC10 from batch correction normalized reads in progenitors with known technical and biological factors.
Extended Data Fig. 3 | Annotating differentially accessible peaks during neuronal differentiation. a, Gene ontology (GO) enrichment of differentially accessible peaks at the TSS. Progenitor peaks (left) and neuron peaks (right) showed enrichment for GO terms related to proliferation and differentiation, as expected. b, TFs with significantly differentially enriched conserved binding sites in differentially accessible peaks. The statistical test identifies TFs likely involved in neural progenitor proliferation and maintenance (progenitorTFs; top) or neurogenesis and maturation (neuronTFs; bottom). The top 30 significantly enriched TFs were shown in this figure, and the full list can be found in Supplementary Table 2. Within progenitorTFs, we found TFs previously characterized to have key roles for neural stem cell renewal and reprogramming, such as SOX2\(^{101,102}\), and those known to be required for the maintenance of stem cells in cortex, such as NR2F1, ETV5, and SP2\(^{103-105}\). Within neuronTFs, NEUROG2 and LMX1A were identified, which are known to drive neuronal differentiation\(^{106}\), as well as TFs shown to induce neuronal identity from fibroblasts, including ASCL2 and the POU family\(^{39}\). NeuronTFs also included CUX1/2, a marker for layer II-III neurons\(^{106}\) and other laminar markers such as TBR1 and FOXP1. c, Schematic of known functions for selected progenitorTFs and neuronTFs.
Extended Data Fig. 4 | Features of caQTLs. a, Flowchart for caQTL data analysis. b, PCA plot for ATAC-seq data on sex chromosomes (chrX and chrY), colored by sex from genotype data, showing sex could be called using ATAC-seq data. c, MDS plot for genotype data of HapMap3 and donors in this study, colored by populations from HapMap3 data. ASW: African ancestry in Southwest USA; CEU: Utah residents with Northern and Western European ancestry from the CEPH collection; CHB: Han Chinese in Beijing, China; CHD: Chinese in Metropolitan Denver, Colorado; GIH: Gujarati Indians in Houston, Texas; JPT: Japanese in Tokyo, Japan; LWK: Luhya in Webuye, Kenya; MEX: Mexican ancestry in Los Angeles, California; MKK: Maasai in Kinyawa, Kenya; TSI: Toscan in Italy; YRI: Yoruba in Ibadan, Nigeria. d, Neuron and progenitor caPeaks enrichment at epigenetically annotated regulatory elements from fetal brain (Epigenetics Roadmap ID = E081). e, Comparison of percent variance explained ($r^2$) for shared neuron/progenitor caQTLs and fetal brain eQTLs (subset to the same sample size). P values are estimated by two-sided paired student-t tests. The center of the box is median of the data, the bounds of the box are 25th percentile and 75th percentile of the data, and the whisker boundary is 1.5 times the IQR. Maximum and minimum are the maximum and minimum of the data.
Extended Data Fig. 5 | Examples of fine-mapping and regulatory mechanisms underlying eQTLs. a, Colocalization of a progenitor-specific caQTL and fetal cortical eQTL for ETFDH. b, caQTL for rs11544037 and the labeled peak in progenitor (N = 76). P-values are estimated by a mixed linear effects model using a two-sided test (Methods). c, eQTL of ETFDH in bulk fetal cortex (N = 235). P-values are estimated by a mixed linear effects model using a two-sided test (Methods). d, The expression of TFs whose motifs are disrupted by rs11544037 (LFC = -0.32, FDR = 7.55e-18). e, The motif Logo of RAD21, where the red box shows the position disrupted by rs11544037. Schematic cartoon of mechanisms for rs11544037 regulating chromatin accessibility and gene expression. f, Luciferase signals for alleles of rs11544037 in progenitors (N = 8). P value is from two-sided paired t-tests. g, Co-localization of a progenitor-specific caQTL and eQTL for FGF1. h, CaQTL for rs11960262 and the labeled peak in progenitor (N = 76). P-values are estimated by a mixed linear effects model using a two-sided test (Methods). i, eQTL of ETFDH in progenitors (N = 85). P-values are estimated by a mixed linear effects model using a two-sided test (Methods). j, The expression of TFs in which motifs are disrupted by rs11960262. k, The motif Logo of EGR1, where the red box shows the position disrupted by rs11960262. Schematic cartoon of mechanisms for rs11960262 regulating chromatin accessibility and gene expression. (For box plots in (b-c), (f) and (h-i), the center of the box is the median, the bounds of the box are 25th percentile and 75th percentile of the data, and the whisker boundary is 1.5 times the IQR. Maximum and minimum are the maximum and minimum of the data.)
Extended Data Fig. 6 | Features of ASCA. a, Density plot for caPeak length from shared caQTLS and ASCA, and from peaks only significant in ASCA in neurons (top) and progenitors (bottom). P values are estimated by two-sided Student’s t-tests. b, The neuron ASCA (caSNP: rs62332390; caPeak: chr4:148,441,611-148,446,300; P values are estimated by the negative binomial generalized linear models from DESeq2 using a two-sided test) is not a significant caQTL (N = 61; P values are estimated by the mixed linear model with a two-sided test) in neurons because the caPeak was very wide (4,689 bp) and only the region near the ASCA SNP shows an association with genotype. c, The neuron ASCA (caSNP:rs77191441; caPeak: chr5:116,571,961-116,576,710; P values are estimated by the negative binomial generalized linear models from DESeq2 using a two-sided test) is not a significant caQTL (N = 6; P values are estimated by the mixed linear effects model with a two-sided test) in neurons due to low minor allele frequency leading to less power to detect a caQTL. d, ASCA between rs185220 (see Fig. 3) and chromatin accessibility in progenitors (left) and neurons (right). P-values are estimated by the negative binomial generalized linear models from DESeq2 using a two-sided test. (For box plots in (b) and (c), the center of the box is the median, the bounds of the box are 25th percentile and 75th percentile of the data, and the whisker boundary is 1.5 times the IQR. Maximum and minimum are the maximum and minimum of the data.).
Extended Data Fig. 7 | Comparison to adult dorsolateral prefrontal cortex (DLPFC) caQTLs. a, Shared accessible peaks overlap at epigenetically annotated regulatory elements from different tissues. Accessible peak bp percentage overlapped with epigenetically annotated regulatory elements. From left to right, tissues ordered by bp percentage overlap with enhancers and promoters. Shared peaks overlap with both adult and fetal regulatory elements. b, PCA plot for read counts from shared peaks in adult DLPFC, neurons and progenitors. c, Correlations of effect sizes for significant neuron caQTLs and the same SNP-Peak pairs in adult DLPFC (left). Correlations of effect sizes for significant progenitor caQTLs and the same SNP-Peak pairs in adult DLPFC (right).
Extended Data Fig. 8 | An example of a neuron-specific caQTL leading to regulatory mechanisms underlying GWAS loci. a, Numbers of colocalizations between ASCA and GWAS loci. b, The neuron-specific significant caQTL (caSNP: rs9930307; caPeak: chr16: 9,805,221-9,805,420) co-localized with schizophrenia GWAS locus (index SNP: rs7191183). c, Box plot for the caQTL (left, N = 61; P values are estimated by the mixed linear effects model using a two-sided test) and ASCA (right) (caSNP: rs9930307; caPeak: chr16: 9,805,221-9,805,420; P values are estimated by the negative binomial generalized linear models from DESeq2 using a two-sided test). d, The expression of TFs in which motifs are disrupted by rs9930307. e, The motif logo of TP53 and the position disrupted by rs9930307. f, The box plot for luciferase signal for alleles of rs9930307 in progenitors (N = 8). P value is from two-sided paired student-t tests. (For box plots in (c) and (f), the center of the box is median of the data, the bounds of the box are 25th percentile and 75th percentile of the data, and the whisker boundary is 1.5 times the IQR. Maximum and minimum are the maximum and minimum of the data.).
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.

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

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

Software and code

Policy information about availability of computer code

Data collection

No software was used for data collection

Data analysis

- BD FACS Diva software (v9.0);
- SH800S software;
- fastqc (v0.11.7);
- atacqv (v1.0.0);
- BBMAP/BBUDK (v38.82);
- BWA mem (v0.7.17);
- Samtools (v1.9);
- USCS tools/liftOver (v320);
- Bedtools (v2.26);
- MACS2 (v2.0.8);
- PicardTools (v2.18.22);
- CSAW (v1.16.1);
- cqn (v1.28.1);
- DESeq2 (v1.22.2);
- limma (v3.38.3);
- TIFFTools (v1.4.0);
- JASPAR2016 (v1.18.0);
- PLINK (v1.90b3);
- VerifyBamID (v1.1.3);
- SHAPEIT (v2.837);
- Minimap2 (v1.0.0);
- motifBreakR (v1.14.0);
- LD Score Regression (v1.0.0);
MotifDb (1.26.0); EMMAX (Beta); GREAT algorithm; Metasoft (v2.0.1); qvalue (v2.20.0); eigenMT; Gviz (v2.18.0); All code is available on Bitbucket: https://bitbucket.org/steinlabunc/celltypespecificcaqtls_wasp/src/master/

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

Data generated in this manuscript (including metadata) can be accessed via dbGaP (phs001958.v1.p1; https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001958.v1.p1).

The RNaseq and genotype dataset used for fetal cortical eQTL analysis are available at dbGaP with accession number: phs001900. REST ChIP-seq data in H1 embryonic stem cells and neurons differentiated from H1 cells are available via the ENCODE portal (https://www.encodeproject.org/) with the following identifiers: ENCSR000BTV and ENCSR000BHM.

Field-specific reporting

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

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

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

Life sciences study design

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

Sample size
We acquired 98 ATAC-seq libraries from progenitors (N=85 cell lines and 13 replicated libraries) and 70 ATAC-seq libraries from neurons (N=66 cell lines and 4 replicated libraries) from in total 92 genotyped donors. Previous caQTL studies with sample sizes from 40 to 100 donors and eQTLs with 70 to 80 donors were powered to detect significant associations (PMID: 30478436; PMID: 29379200; PMID: 29022597; PMID: 29988122; PMID: 29208628). Here, we used 92 donors and after QC, we retained 61 donors for neurons and 76 donors for progenitors for mapping QTLs.

Data exclusions
For differential accessibility analysis and chromatin accessibility QTL studies, we removed mixed ATAC-seq libraries or those that did not match with genotype data (5 donors with FREEMIX > 0.02 or CHIPMIX > 0.02 based on verifybamID results). We also corrected sample swaps (N_donor=7). Then we randomly selected one of the replicated libraries for each cell line to maintain independence for neurons (N_donor=61) and progenitors (N_donor=76). These criteria were established prior to running the study following previous work (PMID: 32732423).

Replication
We cultured cells from the same cell lines in different rounds as technical replicates, for progenitors (N= 11 cell lines and 13 replicated libraries in round 1-7,12 and 13) and neurons (N=4 cell lines and 4 replicated libraries, in round 2,6,12 and 13). We correlated the batch corrected VST counts in CSAW peaks for neuron and progenitor replicates either within donors or calculated correlations across donors. The correlations of replicates within donors are higher than that of samples across donors in both neuron and progenitor samples.

Randomization
To reduce the impact of batch effects on interpretation of our results, we attempted to either have no batches when possible (e.g., perform all experiments using the same lot of a reagent) or when this was not possible, randomize technical variables (round a donor was thawed in, sequencing pool) such that they had minimal correlation with variables of interest. In order to extensively document the impact of technical variables on outcome measures, we maintained a relational MySQL database which allowed us to keep track of many technical and biological variables throughout each experimental event. Each downstream ATAC-seq library preparation therefore is able to be tracked back to all technical and biological variables associated with its cell culture.

Blinding
The investigators were not explicitly blinded to the donor during cell culture or library preparation. However, the investigators did not have knowledge of donor genotype when performing cell culture or library preparation.

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.
We used previously validated antibodies for immunofluorescence. We did not conduct any antibody validation as part of this study.

The Pax6 antibody is validated for multiple commonly used applications such as IHC (Immunohistochemistry) and WB (Western Blot) and shows reactivity across multiple species (https://www.biологен.д.м/en-us/products/purified-anti-pax-6-antibody-11511). In our hands, it shows expected nuclear localization in neural progenitor cells and does not co-localize in differentiated neurons expressing hSyn-EGFP. This antibody was shown to label the progenitors in mouse ventricular zone of the developing cortex (PMID: 32238932).

Both Sox2 antibodies are validated for multiple commonly used applications such as WB and IHC and are reactive to the human isoform (https://www.cellsignal.com/products/primary-antibodies/sox2-d6d9-xp-rabbit-mab/3579 and https://www.emdmillipore.com/US/en/product/Anti-GAD67-Antibody-clone-1G10.2,MM_NF-MAB5406). In our hands, they also show high nuclear co-localization with Pax6+ neural progenitors and do not co-localize in differentiated neurons expressing hSyn-EGFP. They were used to label the neural progenitors in ventricular zone in human brain organoids (PMID: 3161540 and PMID: 23995685).

The GFP antibody is validated in multiple applications such as IHC, WB, and Immunoprecipitation (IP) and is reactive across multiple species (https://www.emdmillipore.com/US/en/product/Anti-Green-Fluorescent-Protein-Antibody,MM_NF-AB16901). In our hands, it co-localized strongly in differentiated neurons infected with AAV-hSyn-EGFP with almost no background signal in cells not transduced with GFP virus.

The TUBB3 (Tuj1) antibody has been previously validated by IHC (https://www.biolegend.com/en-us/products/purified-anti-pax-6-antibody-11511). In our hands, it shows expected nuclear localization in neural progenitor cells and does not co-localize in differentiated neurons expressing hSyn-EGFP. This antibody was shown to specifically label cells in the mouse developing cortical plate in previous work (PMID: 29503187).

The Tbr2 antibody has also been previously validated in human brain tissue by IHC (https://www.thermofisher.com/antibody/product/EDMES-Antibody-clone-WD1928-Monoclonal/14-4877-82) and shows expected nuclear localization in Tuj1+/hSyn-EGFP+ intermediate progenitor cells in differentiated cultures with no detectable expression in proliferative cultures. This antibody was shown previously to label cells within the human inner and outer subventricular zone of the developing cortical wall where intermediate progenitors are known to reside (PMID: 26324239).

The HOPX antibody has been previously used to detect outer radial glia in primary human brain tissue by IHC and shows expected cytoplasmic localization in Sox2+ neural progenitors (https://www.sigmaaldrich.com/catalog/product/sigma/hpa030180). This antibody was used to label outer radial glia in human brain organoids (PMID: 31168097).

All secondary antibodies were used at 1:1000 dilution. AF488-antiRb (Thermo Fisher, Catalog#: A-11011, Lot#: 1832035), AF488-antiCH (Thermo Fisher, Catalog#: A-11039, Lot#: 1759025), AF647-antiRb (Thermo Fisher, Catalog#: A-21247, Lot#: 2119156), AF488-antiMs (Thermo Fisher, Catalog#: A-11001, Lot#: 1939600), AF658-antiMS (Thermo Fisher, Catalog#: A-11031, Lot#: 2026148)
were derived from dorsal telencephalon and contained predominantly Pax6+ progenitors, we did not detect the presence of these cell-types. However, both of these antibodies have been previously validated in iPSC-derived cultures differentiated towards a ventral forebrain fate and have shown expected nuclear localization in Nkx2.1+ progenitors and localization in the soma for Gad67+ GABAergic interneurons (PMID: 28279351).

All secondary antibodies are broadly used in multiple IF applications and are highly cross adsorbed to minimize across species reactivity.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)  Human fetal brain tissue was obtained from the UCLA Gene and Cell Therapy Core following IRB regulations. The tissue is often fragmented during acquisition from the surgical procedure. Flat, thin pieces of tissue that have the morphology of developing cortex were selected, and in some cases the tissue was sufficiently intact to be certain of cortical identity. Presumed cortical tissue from 14-21 gestation weeks was dissociated into a single cell suspension, cultured as neurospheres, plated for a low number of passages (2.5 ± 1.8 s.d.) on laminin/fibronectin and polyornithine coated plates, and then cryopreserved as human neural progenitors (HNPs) following our previous work (Stein et al., 2014). Cryopreserved HNPs were shipped to UNC Chapel Hill after signed material transfer agreement by both institutions.

Authentication  Quality controlled genotype data and ATAC-seq BAM files from the same cell lines were used to identify any sample swaps between the ATAC-seq and genotype data using VerifyBamID.

Mycoplasma contamination  To check if there was any contamination from mycoplasma while in culture, we downloaded 98 mycoplasma genomes (from NCBI) and then mapped all ATAC-seq data to every mycoplasma genome. Less than 0.01% of each ATAC-seq sample mapped to any mycoplasma genome, which demonstrated that our cultures were not contaminated with mycoplasma.

Commonly misidentified lines  (See ICLAC register)  N/A

Human research participants

Policy information about studies involving human research participants

Population characteristics  Human fetal brain tissues were obtained from the UCLA Gene and Cell Therapy Core following IRB regulations, then differentiated into cell lines. Presumed cortical tissues are during 14-21 gestation weeks, and the genotype data can be accessed via dbGAP (phs001958.v1.p1). Sex was determined using genotype data and ATAC-seq data, and we found 34% of donors are female and 66% are male.

Recruitment  The mother consented to use of the tissue in research. We are blinded from any personal information for the participants. Any donors with known aneuploidy were excluded from the population.

Ethics oversight  Prenatal tissue was obtained from the UCLA Gene and Cell Therapy core according to IRB guidelines following voluntary termination of pregnancy. This study was performed under the auspices of both the UCLA Office of Human Research Protection and the Office of Human Research Ethics at UNC, which both determined that it was exempt because samples were derived from anonymous pathological specimens.

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

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  Human neural progenitor cells were proliferated for 3 weeks then differentiated for 5 weeks. Virus for labeling live neurons (AAV2-hsyn1-eGFP) was added at 20,000 MOI. After another 3 weeks differentiation, cells were lifted using Papain with DNase, resuspended in FACS Buffer (PBS + 7.5% BSA), filtered through a cell-strainer, and live/dead stains were added (AnnexinV (Biolegend) and Sytox (Life Tech)), and sent to cell sorter (BD FACSAria II and Sony SH800S) to sort for live neurons labeled with GFP.

Instrument  SH800Z and FACSAria II

Software  The native SH800S software for SH800S and BD FACSDiva Software (v9.0) for FACSAria II.
| Cell population abundance | About 20% of singlet cells were GFP positive. |
|---------------------------|---------------------------------------------|
| Gating strategy           | We first used SSC-A and FSC-A, then SSC-W and SSC-H to remove debris. FSC-W and FSC-H were used to select singlets. We used a non-transduced control (no virus added) for each round of sorting to define the gates of a live GFP+ neuron. A post-sorting procedure was used to verify that the GFP positive cell population fell within the gates defining a GFP+ neuron. |

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