Title: Massively parallel characterization of psychiatric disorder-associated and cell-type-specific regulatory elements in the developing human cortex

Authors: Chengyu Deng[1,2†], Sean Whalen[3†], Marilyn Steyert[4,5,6], Ryan Ziffra[1], Pawel F. Przytycki[3], Fumitaka Inoue[7], Daniela A. Pereira[1,2,8], Davide Capauto[9], Scott Norton[9], Flora M. Vaccarino[9,10], Alex Pollen[11,12], Tomasz J. Nowakowski[4,5,6,12,13], Nadav Ahituv[1,2*], Katherine S. Pollard[2,3,13,14*]

Affiliations:
[1] Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco; San Francisco, CA, USA
[2] Institute for Human Genetics, University of California, San Francisco; San Francisco, CA, USA
[3] Gladstone Institutes; San Francisco, CA, USA
[4] Department of Anatomy, University of California, San Francisco; San Francisco, CA, USA
[5] Department of Psychiatry, University of California, San Francisco; San Francisco, CA, USA
[6] Department of Neurological Surgery, University of California, San Francisco; San Francisco, CA, USA
[7] Institute for the Advanced Study of Human Biology (WPI-ASHBi), Kyoto University; Kyoto, Japan
[8] Graduate Program of Genetics, Institute of Biological Sciences, Federal University of Minas Gerais; Belo Horizonte, Minas Gerais, Brazil
[9] Child Study Center, Yale University; New Haven, CT, USA
[10] Department of Neuroscience, Yale University; New Haven, CT, USA
[11] Department of Neurology, University of California, San Francisco; San Francisco, CA, USA
[12] Eli and Edythe Broad Center for Regeneration Medicine and Stem Cell Research, University of California, San Francisco; San Francisco, CA, USA
[13] Chan Zuckerberg Biohub, San Francisco; San Francisco, CA, USA
[14] Department of Epidemiology and Biostatistics, University of California, San Francisco; San Francisco, CA, USA
† These authors contributed equally to this work.
[*] Co-corresponding authors. Email: nadav.ahituv@ucsf.edu, katherine.pollard@gladstone.ucsf.edu

Abstract:

Nucleotide changes in gene regulatory elements are important determinants of neuronal development and disease. Using massively parallel reporter assays in primary human cells from mid-gestation cortex and cerebral organoids, we interrogated the *cis*-regulatory activity of 102,767 sequences, including differentially accessible cell-type specific regions in the developing cortex and single-nucleotide variants associated with psychiatric disorders. In primary cells, we identified 46,802 active enhancer sequences and 164 disorder-associated variants that significantly alter enhancer activity. Activity was comparable in organoids and primary cells, suggesting that organoids provide an adequate model for the developing cortex. Using deep learning, we decoded the sequence basis and upstream regulators of enhancer activity. This work establishes a comprehensive catalog of functional gene regulatory elements and variants in human neuronal development.

One Sentence Summary:

We identify 46,802 enhancers and 164 psychiatric disorder variants with regulatory effects in the developing cortex and organoids.

Main Text:

Introduction

Psychiatric disorders affect nearly one in five adolescents in the United States (*1*) and have a strong genetic etiology (*2*). Studies profiling gene expression across distinct anatomical regions have found an enrichment for psychiatric disorder-associated genes in stages of developmental neurogenesis in the marginal zone and deep cortical layer neurons (*2–4*). For example, most autism spectrum disorder (ASD) risk genes were found to regulate gene expression or be involved in neuronal communication during early brain development (*5*). Analysis of genes associated with schizophrenia via genome-wide association studies (GWAS) found several to be expressed in the prefrontal cortex at early developmental stages (*6*). Thus, decoding the genetic causes of psychiatric disorders requires deep knowledge of gene regulatory mechanisms in the developing brain.

In the past decade, hundreds of psychiatric disorder-associated genetic risk loci have been identified, both by work of individual labs and large consortia, including the Psychiatric Genomics Consortium (*7*, *8*) and psychENCODE (*9*). A major portion of these loci reside in noncoding regions of the genome, likely within gene regulatory elements, and contain highly correlated variants due to linkage disequilibrium (LD), making them challenging to interpret and functionally characterize. Gene regulatory elements, such as enhancers and promoters, are known to regulate lineage- and region-specific transcription in the developing human cortex (*10*). While promoters are located adjacent to their target genes, enhancers can be located at distal locations from the gene(s) that they regulate. In addition, due to their cell-type specificity and spatiotemporal dynamic activity, enhancers are difficult to identify. Single-cell ATAC-seq (scATAC-seq) at different developmental stages of the human cortex enabled the identification of different cell populations and their candidate regulatory elements in the developing cortex (*11*, *12*). This work also showed significant correlations between these cell types and cerebral organoids. However, these studies are descriptive and do not provide a functional readout that can test the enhancer activity of these sequences and the effects of psychiatric disorder-associated variants on their function.

Massively parallel reporter assays (MPRAs) enable the simultaneous testing of thousands of sequences for their regulatory activity (*13*). The quantitative readout of an MPRA makes it possible to test different alleles of the same locus side-by-side in one experiment, enabling the detection of sequence variants that alter enhancer function (*14–16*). Because they generate data for large numbers of sequences, MPRAs have enabled the development of deep learning models that predict enhancers and their quantitative activity from sequence alone (*17*, *18*). Machine learning has been used to predict: 1) enhancers from epigenetic marks, sequence motifs, and evolutionary conservation (reviewed in (*19*)); 2) MPRA activity from sequences and epigenetic features (*20*); 3) gene expression from promoter-proximal sequences (*21–23*); and 4) epigenetic features from sequences (*24–30*). In addition, DNA sequence-based models (*17*, *18*) have the potential to be deployed at scale to screen variants without MPRA data and to design enhancers with desired properties, such as cell type specificity. These strategies have shed light on the sequence motifs and upstream regulators that are most important for regulating gene expression across different cell types and species.

Here, we used deep learning and a lentivirus-based MPRA (lentiMPRA) to test 102,767 sequences for their enhancer activity in primary human mid-gestation cortical cells and 10-week cerebral organoids. These sequences included cell-type-specific open chromatin regions and psychiatric disorder-associated quantitative trait loci (QTLs) predicted to be functional enhancers based on their epigenetic profiles. Combined, we discovered 46,802 sequences to be functional enhancers and 164 variants with significant allelic differences in enhancer activity regulating known disorder-associated genes such as *TBR1*, *MARK2* (autism spectrum disorder), and *NFKB2* (schizophrenia). We observed comparable activity levels between organoids and primary cortical cells, suggesting that organoids provide an adequate *in vitro* model to study the developing cortical regulatory landscape. Finally, we used our lentiMPRA data to train a deep learning model that predicts enhancer activity from sequence with state-of-the-art accuracy, enabling us to learn sequence determinants and upstream regulators of the human mid-gestation enhancer code. These findings provide a comprehensive catalog of functional cortical-cell-specific enhancers and psychiatric disorder-associated variants that alter their activity, improving our understanding of the molecular basis of neurodevelopment and the etiology of psychiatric disorders.

Results

LentiMPRA library generation and analysis

To comprehensively characterize human neurodevelopmental enhancers and their sequence variants across the mid-gestation cortex, we designed two lentiMPRA libraries (Methods) and tested them in primary human cortical cells (**Fig. 1A**). Due to the limited number of obtainable human primary cells and lentivirus integrations into these cells, each library was assayed independently.

The first library was designed to characterize cell-type specific enhancers. It consisted of 51,495 sequences obtained primarily from differentially accessible (DA) scATAC-seq peaks in the developing human brain (*11*). These DA peaks were further selected based on their: 1) overlap with histone H3 acetylation of lysine 27 (H3K27ac) peaks from bulk prefrontal cortex (PFC) tissue (*31*), microglia or non-microglia cells (*11*) (n = 24,611, 53%); or 2) overlap with H3K4me3 proximity ligation-assisted chromatin immunoprecipitation sequencing (PLAC-seq) peaks from intermediate progenitor cells (IPC), radial glia (RG), excitatory neurons (EN), or interneurons (IN) (*32*) (n = 12,412, 26.8%); or 3) overlap with promoter capture Hi-C (PCHi-C) from EN, hippocampal dentate gyrus (GE)-like neurons, lower motor neurons and astrocytes (*33*) (n = 13,712, 29.5%).

The second library compared the reference and the alternative alleles for 17,069 variants. These were designed from brain QTLs (*34–36*) overlapping pseudo-bulked ATAC-seq peaks (*11*) that: 1) are within 100 kilobases (kb) of differentially expressed genes in schizophrenia, autism or bipolar disorder (expression QTL (eQTL) n = 14,021; chromatin QTL (caQTL) n = 149) (*9*); or 2) share a linkage disequilibrium (LD) block with GWAS SNPs for various psychiatric disorders (*8*, *37–44*) (eQTL n = 2,882, cQTL n = 17). Assaying both QTL and GWAS SNPs allowed us to overcome the systematic bias toward different types of variants in each type of study (*45*), thus enabling a comprehensive screening for functional regulatory variants. This library also included ~15,000 non-QTL sequences with a range of expected activity levels predicted from their epigenetic profiles.

To prioritize distal enhancers, promoter-overlapping peaks were excluded from both libraries. Each library contained 143 positive control sequences nominated from ATAC-seq and ChIP-seq data in brain organoid models (*46*) and used to define active sequences, plus ~2,000 additional controls used for quality control. We synthesized 270 base pair (bp) oligos, each centered on either the DA peak summit (library 1) or variant (library 2) followed by 15bp primers on either side to amplify the library. A 31bp minimal promoter (minP) and 15bp random barcode were placed downstream of each synthesized oligo via PCR and cloned into a lentiMPRA vector (**Fig. 1A**).

Both libraries were packaged into lentivirus and used to infect human primary cortical cells at mid-gestational week (GW) 18. Primary cortical cells were cultured for two days before infection, exhibiting complex morphology and smooth neurites. The presence of major cortical cell types was confirmed by immunocytochemistry before and after infection, including deep-layer excitatory neurons (dlENs), upper-layer excitatory neurons (ulENs), newborn excitatory neurons (earlyENs), RG, IPC, astrocyte/oligodendrocyte precursors (Astro/oligo), endothelial

and mural cells (EndoMural), medial ganglionic eminence derived interneurons (IN-MGE), caudal ganglionic eminence derived interneurons (IN-CGE), and microglia (MG) (**fig. S1A**). We performed three replicates for the DA library and five replicates for the variant library. Three days post-infection, when the majority of the non-integrated virus was gone, DNA and RNA were harvested and prepared for sequencing. DNA sequencing revealed that both libraries contained more than 96% of the designed oligos (DA library: 50,394 oligos; variant library: 51,319 oligos), and each oligo had on average over 50 unique barcode associations (median DA: 56, variant: 64). Overall, 97,762 sequences (95%) passed stringent quality control (Methods).

To measure enhancer activity, we quantified depth-normalized barcode abundance in DNA and RNA for each oligo and then calculated its batch-corrected RNA/DNA ratio. A high correlation of RNA/DNA ratios between replicates was observed (average Pearson correlation, DA: 0.93, variant: 0.91; **Fig. 1B**), confirming sufficient reproducibility. We next compared the activity distributions of positive and negative controls (**fig. S2A**). As expected, positive controls had significantly higher ratios than negative controls in both libraries (DA: p = 1e-3, variant: p = 8e-5, Wilcoxon test). Moreover, the distribution of ratios for randomly scrambled negative controls was highly comparable between libraries (median DA: 0.997, median variant: 0.994), indicating that activity was mostly driven by the minimal promoter. Altogether, these quality assessments suggest that our lentiMPRA robustly distinguishes between sequences with high, medium, and low regulatory activity.

To identify sequences capable of driving gene expression, we defined active sequences as those with RNA/DNA ratios higher than the median of positive controls in their respective libraries (DA: 1.047, variant:1.068), conservatively treating the remaining sequences as inactive. We further defined the highest activity sequences as those above the 75th percentile of the positive controls and the lowest activity sequences as those below the 25th percentile. Combining both libraries, we identified a total of 46,802 active sequences (48% of 97,762) and 25,557 with the highest activity. We next evaluated the various properties of active lentiMPRA sequences. Compared to inactive sequences, active sequences are significantly more conserved (p=5.8e-28, Wilcoxon test), and their target genes are expressed at higher levels during mid-gestation (p=6.4e-6, Wilcoxon test; 23% of all sequences mapped to target genes using PLAC-seq data (*32*)). Comparing the highest versus lowest activity sequences, we found gene ontology (GO) enrichment for neurodevelopmental terms, such as 'nervous system development' and 'neuron differentiation' (**Fig. 1D**). Next, we analyzed transcription factor binding sites (TFBSs) and observed enrichment for known neurodevelopmental TFBSs including the *DLX*, *LHX* and *SOX* gene families. We also found enrichment for universal stripe factors (USFs) including *EGR1*, *MAZ* and members of the *KLF/SP* family (**Fig. 1E**). USFs colocalize at most promoters and enhancers, increasing chromatin accessibility and residence time for cofactors (*47*), and our finding suggests that they play a similar role with lentiMPRA reporter constructs integrated into the genome. Together, these results indicate that our active sequences have biological functions in brain development.

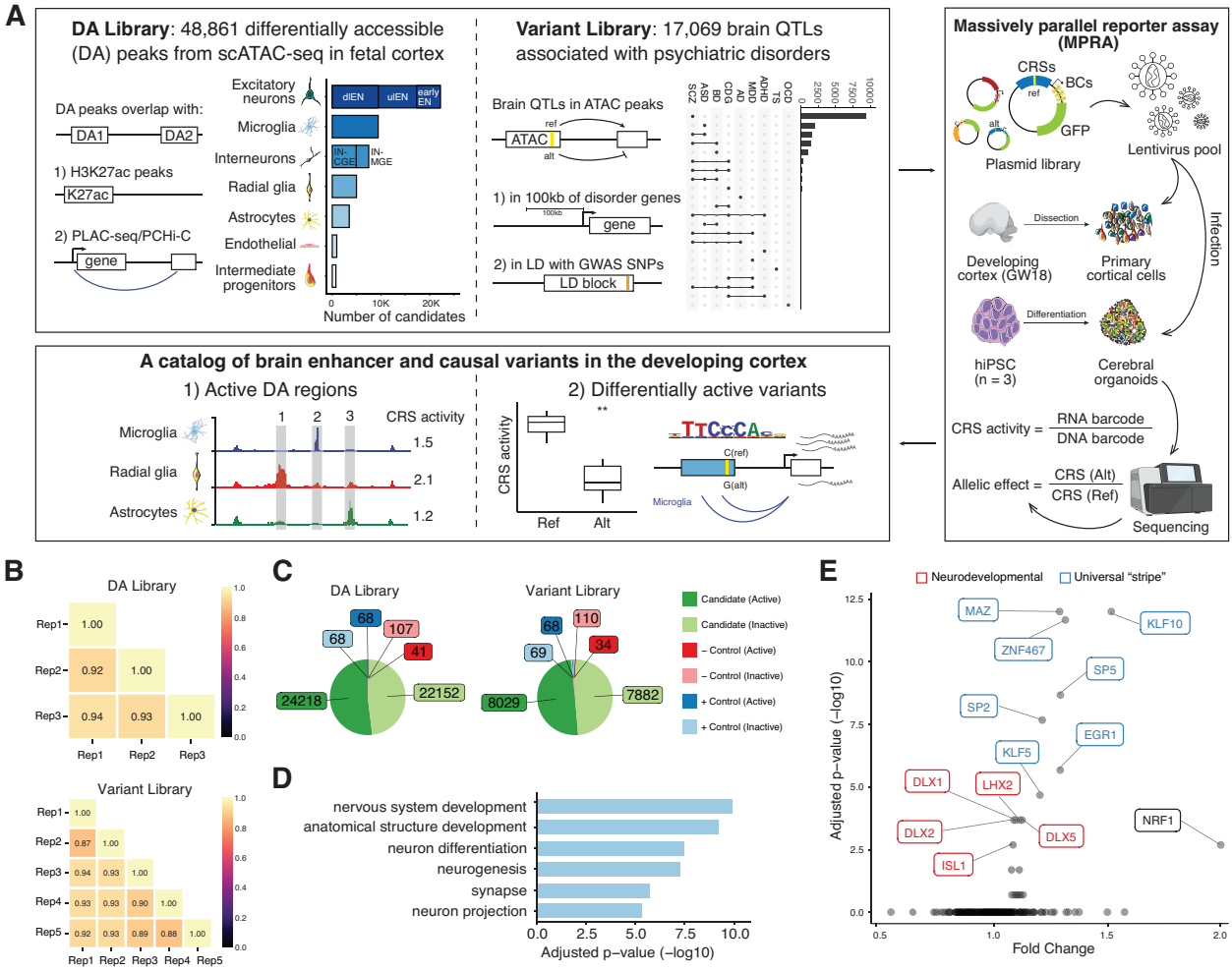

Fig. 1. Design and overall lentiMPRA results. (**A**) Experimental overview of the two lentiMPRA libraries. Library 1 contains 48,861 differentially accessible (DA) regions from scATAC-seq in the developing human cortex that either overlap H3K27ac peaks or PLAC-seq/PCHiC loop. The number of DA candidates for each cell type was portrayed in the bar plot. dlEN, deep layer excitatory neuron; ulEN, upper layer excitatory neuron; IN-CGE, CGE derived interneuron; IN-MGE, MGE derived interneuron. Library 2 includes 17,069 brain QTLs that are 100kb from differentially expressed cross-disorder neurodevelopmental genes or in linkage disequilibrium (LD) with psychiatric disorder GWAS SNPs. The number of variants associated with each disorder is shown in the upset plot (top 22 intersects shown). SCZ, schizophrenia; ASD, autism spectrum disorder; BD, bipolar disorder; CDG, congenital disorders of glycosylation; AD, Alzheimer's disease; MDD, major depressive disorder; ADHA, attention-deficit/hyperactivity disorder; TS, Tourette syndrome; OCD, obsessive-compulsive disorder. Both libraries were cloned into a lentiMPRA vector and packaged into lentivirus and used to infect primary cortical cells dissociated from GW18 tissues and human induced pluripotent cell (hiPSC)-derived cerebral organoids. Following infection, DNA and RNA were extracted and sequenced and an RNA/DNA barcode count ratio was calculated for each candidate regulatory sequence (CRS) allowing the identification of active DA regions and differentially active variants. (**B**) Correlation of log2(RNA/DNA) between technical replicates in primary cortical cells for library 1 and 2, respectively. (**C**) Pie charts showing the number of active and inactive sequences for candidates, positive (+) and negative (-) controls in both libraries. (**D**) Top enriched GO terms from the 'Biological Process', 'Cellular Component' and 'Molecular Function' ontologies for nearest genes of the highest activity sequences (both libraries combined). Closest genes of the lowest activity sequences were used as the background set. The complete list of GO terms is available in fig. S2B. (**E**) TF motif enrichment analysis for highest activity sequences (both libraries). Red: neurodevelopmental TFs, Blue: USFs.

Identification of thousands of active cortical-cell-type specific enhancers

Of 46,370 DA sequences passing quality control, 24,218 (52%) were active enhancers in primary cortical cells (**Data S1**). Among these active DA, 4,656 (19%) were transiently accessible regions across pseudotime of excitatory neurogenesis in a previous scATAC-seq study (*11*). We then separated DA sequences based on their cell-type specificity and found that for each cell type 43-62% of the sequences were active (**Fig. 2A**). We observed that abundant cell types, including neurons and radial glia, had a slightly higher percentage of active sequences compared to less abundant cell types, such as microglia and endothelial/mural cells. Compared to the lowest activity DA sequences, the highest activity DA sequences are enriched for TFBS of neurodevelopmental transcription factors, and many of these show positional enrichment within the ATAC-seq peak (**Fig. 2B**). For example, active sequences tend to have motifs for ATOH1, NEUROD2, and TCF4 upstream of the peak summit, whereas ASCL1 and SPI1 motifs are enriched downstream of the summit.

Next, we assessed the characteristics of active DA sequences across cell types (**Fig. 2C and table S1**). In almost all cell types, active DA sequences are more conserved than inactive sequences, consistent with prior knowledge that neurodevelopmental enhancers tend to exhibit strong conservation across vertebrate evolution (*48*). Inhibitory neurons derived from the ganglionic eminence exhibited the largest differences in conservation scores between active and inactive sequences, fitting with the general transitory role of the ganglionic eminence in guiding neuronal migration (*49*). Next, to test whether active DAs had regulatory activities endogenously, we predicted the target genes of each DA sequence using PLAC-seq data (*32*) and calculated cell-type-matched expression using scRNA-seq data in the developing human cortex (*11*). For many neuronal subtypes, genes interacting with active DAs showed higher expression compared to genes interacting with inactive DAs. For example, putative genes regulated by dIEN-specific active DAs were transcribed at a higher level in dIEN, compared to genes regulated by dlEN-specific inactive DAs ($p = 0.00016$, Wilcoxon test). We also found a significant higher number of strong TF motifs in active DAs specific to AstroOligo, EndoMural, Microglia and RG, while USF motifs were enriched in DAs specific to IN-CGE plus these same four glial and vascular cell types. These results indicate that the activity of DA sequences in lentiMPRA is associated with motif content and target gene expression in the matched cell type. Lastly, we observed unique sets of enriched motifs in active DAs of each cell type, and found many of them formed functional protein-protein interaction (PPI) networks (**fig. S4**).

To further verify the cell-type-specific activity of active DAs in our lentiMPRA, we selected eleven DA peaks with high MPRA activity from six different cell types (**table S2**) and tested them individually for their enhancer function in cortical cells followed by co-staining with antibodies for cell-specific markers (**Fig. 2D**). Each sequence was individually cloned into our lentiMPRA vector in front of a GFP reporter construct. As tissue samples are difficult to obtain at exactly GW18, we used cortical sections covering the mid-gestation period more generally (GW16-20) for infection. We first validated that the cortical sections can be uniformly infected by transducing them with a constitutively active enhancer vector, finding that our virus was able to infect with even spatial distribution across cell types (**fig. S3**). We next infected tissues with the individual enhancer viruses and found that all candidates showed GFP expression (**fig. S3**), in agreement with our lentiMPRA results. Cell-type specificity was inferred from GFP spatial location, counterstaining with cell markers, and morphology. While some candidates did not

display strong cell-type-specific signals, we found three excitatory neuron-specific DA sequences (EN-1, ulEN-2, and dlEN-2) showing enhancer activity in the expected cell type. For example, a ulEN-specific DA region (chr5:89274678-89274948, hg38, **Fig. 2E**) drove GFP expression predominantly in the upper areas of the cortical plate and largely co-localized with *SATB2*, an upper layer excitatory neuron marker. Using PLAC-seq data (*32*), we found that this region has an EN-specific interaction with the promoter of *MEF2C* and *MEF2C-AS1*, known ASD and SCZ genes with EN-specific expression in the developing cortex. Another example is a pan-excitatory neuron specific accessible region (chr2:165141999-165142269, hg38, **Fig. 2F**) that showed notably higher GFP signal in the cortical plate (CP) and subplate (SP) compared to the ventricular zone (VZ) and outer subventricular zone (OSVZ), with most of the cells positive for GFP and *SATB2* located in the top layer of CP.

Not all sequences showed enhancer activity within or unique to their predicted cell type. Two regions (ulEN1 and dlEN1) showed GFP expression outside the regions where the expected cell type is enriched and did not overlap with cell marker staining. Candidate sequences specific to Astro/Oligo or RG showed GFP signal around the VZ but also near the CP. In these cases, GFP+ cells showed complex morphologies: some matched with the expected cell type(s) while others did not (**fig. S3**). To conclude, we independently validated the enhancer activities of eleven sequences with high lentiMPRA activity, finding all of them to drive GFP expression in cortical cells, with three exhibiting cell-type-specificity consistent with predicted activity from scATAC-seq.

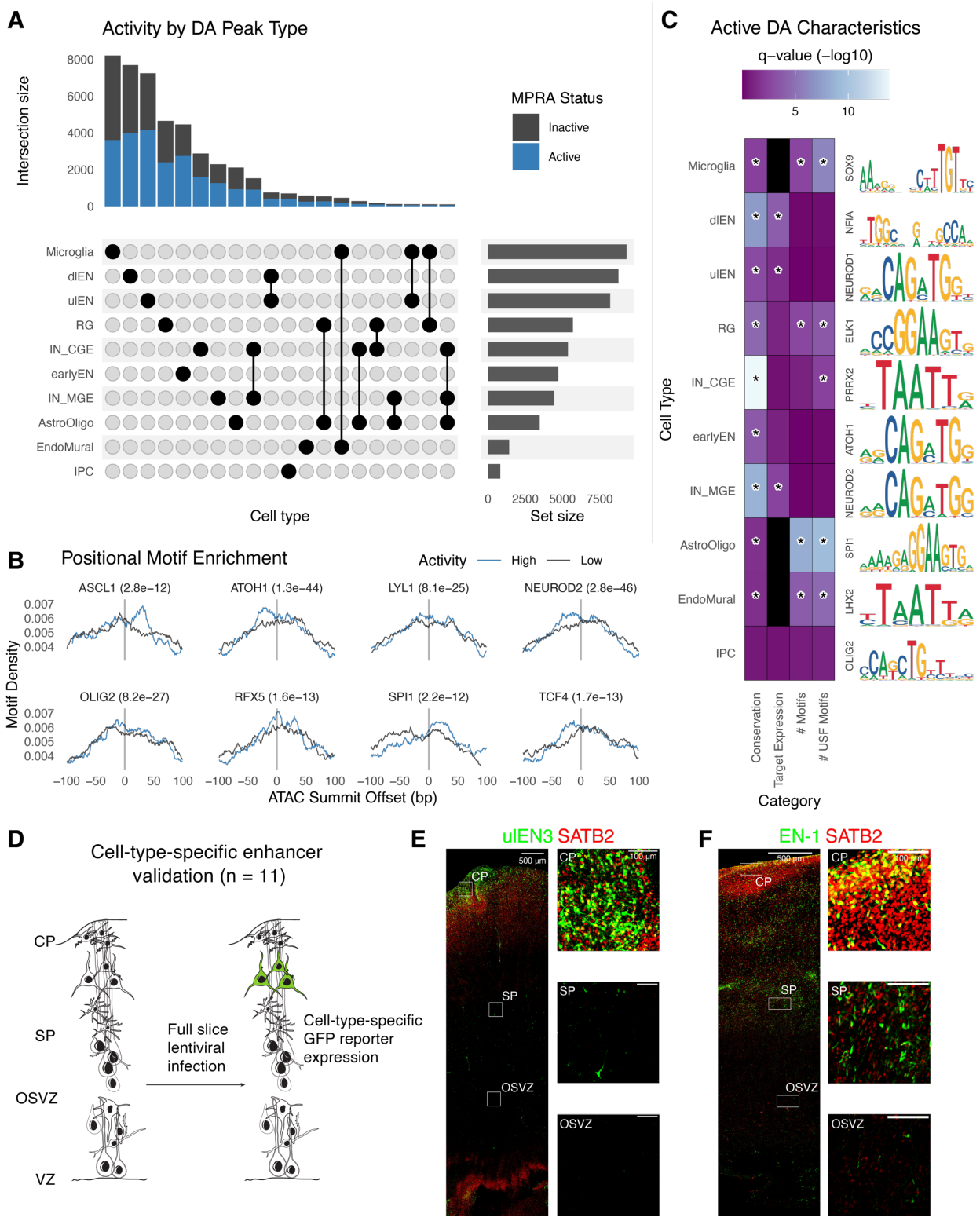

Fig. 2. Identification and validation of functional differentially accessible regions in the developing cortex. (**A**) Upset plot showing the number of DA peaks (active: blue; inactive: gray) for each cell type or combination of cell types. (**B**) The highest activity DA sequences have positional motif enrichment for neurodevelopmental TFs compared to the lowest activity sequences, exhibiting significantly more motif matches slightly up- or downstream of the ATAC-seq peak summit. (**C**) Active DA sequences have significantly higher means across several attributes

compared to inactive DA sequences (color scale: Wilcoxon test q-values, black = no data) including evolutionary conservation (phyloP), expression of PLAC-seq linked target genes in matched cell types, total number of strong motif matches (q-value < 0.01), and total number of strong USF motif matches. A representative motif enriched in active DA peaks for each cell type is shown on the right. Statistically significant comparisons (q-value < 0.05) are indicated by a star. (**D**) Experimental strategy for validating cell type specificity of active DA sequences. (**E-F**) Mid-gestation human cortex slice cultures transduced with a GFP lentivirus reporter driven by a ulEN-specific enhancer (chr5:89274678-89274948, hg38) (**E**) and a pan-excitatory neuron specific enhancer (chr2:165141999-165142269, hg38) (**F**). Expression of GFP (green) and *SATB2* (red) was visualized via immunohistochemistry staining and insets show colocalization of GFP+ along with *SATB2*+ cells in different layers.

lentiMPRA identifies functional regulatory variants associated with psychiatric disorders

We next characterized the psychiatric disorder-associated variant library. Of the 15,335 variants with both alleles passing quality control, 8,029 (52%) showed enhancer activity from at least one allele. For these active variants, we estimated the allelic effect on enhancer activity by testing for differential activity across replicates. Most variants had modest effect sizes (median absolute $\log_2$ FC = 0.069) (**Fig. 3A**). This is in line with previous MPRAs studying the effect of regulatory variants (*50*, *51*), and consistent with eQTL analyses that find ~1% of single nucleotide changes are associated with significant changes in gene expression (*52*). At a 10% false discovery rate, 164 variants showed significant allelic effects with the number of down-regulating and up-regulating variants being similar (51% versus 49%, p = 0.81, Binomial test, **Fig. 3B**) (*53*). Our subsequent analyses focus on these 164 differentially active variants (DAVs) (**Fig. 3B**, **Data S2**). Among DAVs, 26 were in LD with GWAS SNPs and 138 were within 100kb of differentially expressed disease genes, which is similar to expectation given the library design (17% GWAS and 83% eQTL). Consistent with being QTLs, DAVs are not enriched for low-frequency variants (OR=0.8, p=0.34) nor do they have elevated conservation (OR=0.88, p=0.52). Separating DAVs based on cell type showed enrichment in Astro/Oligo (OR=2.39, p=0.14, **Fig. 3C**), which is particularly striking given the relatively low proportion of this cell type in our cultures and indicates the importance of astrocytes and oligodendrocytes in psychiatric disorders.

Next, we compared our DAVs to prior studies. A recent MPRA for dementia-associated variants in human embryonic kidney cells (HEK293T) (*51*) included 96 variants that were also in our library. We found that 89 variants show no significant allelic effect in either study, and 7 altered enhancer activity in HEK293T cells but not in our primary cortical cell data. This difference could be due to the cell types and/or the thresholds used to assign differential activity. Comparing our DAVs to eQTL data from psychENCODE (*34*), we found that 55% of DAVs (n = 77) had effects in the same direction as the eQTL and effect sizes were weakly correlated (Pearson's correlation = 0.14, p = 0.102) (**fig. S5A**). Despite this, the correlation between MPRA and eQTL in non-DAVs (Pearson's r = 0.008, p = 0.366) was notably lower than that in DAVs (**fig. S5A**). This corroborates that our lentiMPRA is efficient in identifying functional variants while underscoring differences between reporter activity and endogenous gene expression.

To decode the mechanisms through which the 164 DAVs exhibit differential activity, we predicted losses and gains of TFBS motifs using motifbreakR (*54*) (threshold = 1e-5), identifying 34 DAVs (21%) in which the alternative allele alters at least one TFBS (**Fig. 3D**). We also found that there is significantly more TFBS disruption compared with non-DAVs (OR = 1.49, p =

0.047, Fisher's exact test). We then analyzed whether these predicted disrupted TFs functionally or physically interact with each other using the STRING database (*55*) and found a significant TF network centering on *SOX2* and *STAT3* (PPI enrichment p < 1e-16, **fig. S5C**).

We next predicted the putative target gene/s of DAVs using chromatin interaction data in various brain cell types (*32*, *33*, *56*) and adult brain eQTLs (*34*, *36*)(**Fig. 3A**), finding 48 DAVs (29.3%) to have chromatin loops with gene promoters and 8 of these (17%) to be eQTLs for the interacting gene. As regulatory activities vary over development, target genes predicted using adult brain eQTLs may not reflect genes regulated in early brain development, and thus we prioritized target genes predicted from chromatin interaction data. Many target genes are known risk genes or within susceptibility loci for psychiatric disorders and neural diseases. For example, we found that variant rs2193495, located in a dlEN-specific DA region that is thought to regulate the expression of T-box brain transcription factor 1 (*TBR1*), an ASD haploinsufficient associated gene and neurodevelopmental regulator (*57*), leads to reduced MPRA activity, possibly due to the creation of EOMES and MAZ binding sites (**Fig. 3E**). Another down-regulating variant, rs2154984, resides in a putative enhancer of the microtubule affinity regulating kinase 2 (*MARK2*), a risk gene whose loss-of-function variants have been associated with ASD (*58*) (**Fig. 3F**). This variant decreases the affinity of MTF1 and ZNF148 binding sites while increasing the affinity of PPARD and NR2F6 binding sites (**Fig. 3F**). Another example includes SCZ-associated variant rs10786689 that is thought to regulate nuclear factor kappa B subunit 2 (*NFKB2*) and suppressor of fused homolog (*SUFU*). This variant decreases enhancer activity, possibly due to the disruption of a SOX2 and/or SOX4 TFBS (**fig. S5D**). Since both genes were found to be up-regulated in SCZ patients (*59*, *60*), our results suggest that the alternative allele of rs10786689 could be protective. Finally, rs73392121 resides within a microglia scATAC peak and is thought to regulate NPC intracellular cholesterol transporter 1 (*NPC1*), a known cause of Niemann-Pick disease type C, with the alternative allele leading to reduced MPRA activity. Mutations in this gene lead to impaired cholesterol and lipid cellular transport, including microgliosis (*61*). These findings demonstrate that lentiMPRA can nominate candidate causal variants for known disease genes.

As a second strategy for linking DAVs to psychiatric disorders, we focused on known risk loci with multiple variants tested in our lentiMPRA. For example, in the SCZ-associated region 6p21.2 (*62*) (**Fig. 4G**), we tested 38 variants and found 2 DAVs: rs6912602 and rs9368977. DAV rs6912602 is one of the most differentially active in our lentiMPRA (3.3-fold decrease) and is an eQTL associated with reduced expression of peptidylprolyl isomerase like 1 (*PPIL1*). Partial loss-of-function variants in *PPIL1* cause neurodegenerative pontocerebellar hypoplasia in humans and mice (*63*). DAV rs9368977 increases activity in lentiMPRA, resides in open chromatin in RG and IPC, and is an eQTL for chromosome 6 open reading frame 89 (*C6orf89*). The alternative allele of rs9368977 disrupts the motifs of USFs SP3 and KLF4 (**Fig. 4H**). In the SCZ-associated locus 6p21.1 (*64*), we tested 48 variants and identified one DAV: rs1343025. The alternative allele of rs1343025 is associated with increased expression of vascular endothelial growth factor A (*VEGFA*). *VEGFA* regulates cerebral blood volume and is associated with SCZ, though the exact impact of *VEGFA* remains controversial (*65*). Another example, is the ASD risk loci is 16p11.2 (*66*), where we tested 25 variants and discovered one activity-increasing DAV, rs145650870 (**Fig. 4I**). This variant is located in an RG-specific chromatin loop for three nearby genes: tu translation elongation factor (*TUFM*), ataxin type 2 (*ATXN2L*) and

SH2B adaptor protein 1 (*SH2B1*). The alternative allele of rs145650870 creates a TFBS for EHF (**Fig. 4J**). Combined, these results show that our lentiMPRA approach could be used to prioritize variants that have an effect on regulatory activity in disease-associated loci.

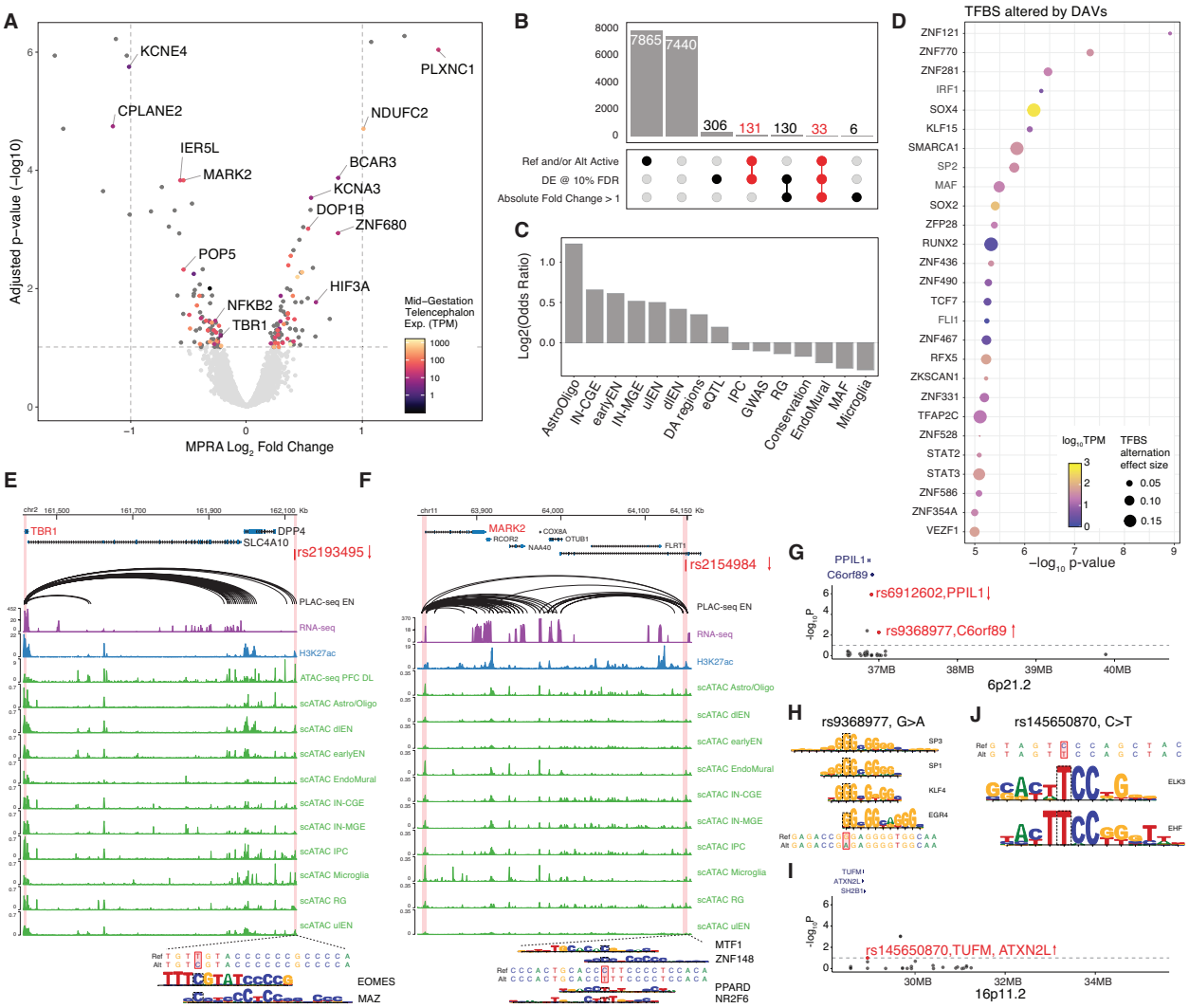

Fig. 3. Identification of differential active variants associated with psychiatric disorders. (**A**) Volcano plot showing $\log_2$ Fold Change and $-\log_{10}$ adjusted p-value for variants that have enhancer activity from at least one allele. Significant variants (FDR < 0.1) were annotated with PLAC-seq predicted target gene name and color-coded based on target gene expression in mid-gestation telencephalon. Two vertical dashed lines indicate the absolute $\log_2$FC of 1. The horizontal dashed line indicates FDR at 10%. (**B**) Upset plot showing the number of variants (bar) passing combinations of different thresholds (dots and lines below bar). The number of DAVs was highlighted in red. (**C**) Enrichment $\log_2$ odds ratio of DAVs overlapping different features, including combined or separate cell-type-specific DA regions, adult brain eQTL, GWAS of various psychiatric disorders and low-frequency variants with minor allele frequency (MAF) less than 0.01. (**D**) TFBSs predicted to be altered by DAVs using motifbreakR. Dot color represents TF expression in primary cortical cells, size represents predicted magnitude of binding affinity alternation. TFs were ranked by TFBS alternation significance ($-\log_{10}$p-value, y-axis). (**E-F**) Genomic browser tracks showing examples of causal regulatory variants and their predicted target genes. The top track shows PLAC-seq chromatin loop in EN (*32*), the second track shows bulk RNA-seq in primary cortical cells, the third track shows bulk H3K27ac ChIP-seq (*31*), followed by a track of bulk ATAC-seq in deep-layer cortex (*31*). The bottom ten tracks show scATAC-seq in human cortex (*11*). DAV rs2193495 (**E**), located in a dlEN-specific accessible region, potentially down-regulates *TBR1* expression due to the introduction of EOMES and MAZ binding sites. DAV rs2154984 (**F**) is predicted to regulate *MARK2* expression and disrupt MTF1 and ZNF148 and introduce PPARD

and NR2F6 binding sites. (**G**) Manhattan plot of SCZ-associated chromosome band 6p21.2 showing the 38 variants tested. The y-axis shows -$\log_{10}$ of adjusted p-value from MPRA. DAVs are highlighted in red and annotated with their predicted target gene. Arrows indicate the direction of allele effect observed in MPRA. (**H**) DAV rs9368977 located in 6p21.2 is predicted to disrupt binding of SP3, SP1, KLF4 and EGR4. (**I**) Manhattan plot of ASD-associated chromosome band 16p11.2 showing the 25 variants tested. The y-axis shows -$\log_{10}$ of adjusted p-value from MPRA. DAVs are highlighted in red and annotated with predicted target genes. The arrow indicates the direction of allele effect observed in MPRA. (**J**) TFBS altered in rs145650870. The alternative allele favors the binding of EHF and ELK3.

Organoids show comparable lentiMPRA activity to primary cells

Previous single-cell transcriptomic and epigenomic data along with immunohistochemical analyses suggest that cortical organoids recapitulate many of the cell types in the developing human forebrain (*11*, *46*, *67–69*). To explore the suitability of organoids as a tissue source for MPRAs, we tested both our lentiMPRA libraries in 10-week-old cortical organoids (**Fig. 4A**), which were validated for the expression of relevant cell type markers, such as *FOXG1*, *PAX6*, *EOMES*, *LHX2*, via immunostaining (**Fig. 4B**) and bulk RNA-seq (**Fig. 4C**). Following nine weeks of directed differentiation towards a dorsal forebrain fate, organoids were sectioned into 300-µm-thick slices and infected with the lentiMPRA libraries at 10 weeks. This slicing approach allowed diffusion of lentivirus into the majority of cells, providing high integration rates per cell (Multiplicity of infection (MOI) = 100). Slicing is also known to attenuate hypoxia, leading to better organoid cell health (*70*). For each library, we infected organoids derived from 2-3 iPSC lines with 2-4 technical replicates each and analyzed the data as described above for primary cells. We observed a high correlation between lentiMPRA replicates (average Pearson correlation over all tested sequences in each library, DA: 0.89, variant: 0.90) and positive controls consistently showed higher enhancer activity compared to negative controls (DA: $p = 6.6e-4$; variant: $p = 0.027$, Wilcoxon test, **fig. S2A**), confirming the high quality of our organoid data.

We compared RNA/DNA ratios between organoids and primary cortical cells and observed high correlation for both libraries (average Pearson correlation DA: 0.89, variant: 0.87, **Fig. 4D**). Similar to primary cells, roughly half of tested sequences were active (total: 31,954, DA: 23,832, variant: 8,122). The vast majority of organoid active sequences were also active in primary cells (**Fig. 4E**). To put this high level of concordance in the context of gene regulation, we performed bulk RNA-seq on three primary and three organoid samples (average replicate Pearson correlation, primary: 0.98, organoid: 0.99) and observed similar transcript levels between primary and organoid samples (average Pearson correlation 0.88), with some notable exceptions that we discuss below. Finally, we compared the activity of DA sequences stratified by the cell types in which they are accessible and found that active DA sequences were highly concordant in organoids versus primary cells for each cell type (**Fig. 4G**). The two cell types in which the lowest proportion of primary cell active DAs replicated in organoids were microglia (86.1%) and endothelial cells (86.4%), which is expected since these cell types are thought to be absent in cerebral organoids and our ability to assay activity of these DAs relies upon the permissiveness of MPRAs. These results suggest that cerebral organoids are a reasonable *in vitro* model of mid-gestation enhancer activity and gene expression, despite some differences in cell type composition.

Next, we examined the concordance of differential allelic activity between organoids and primary cells. In organoids, we observed a median absolute $\log_2$AR of 0.066, similar to that in primary cells (0.069), and detected 420 DAVs (FDR<10%), of which 74 (18% of organoid DAVs, 45% of primary DAVs) were also DAVs in primary cells (**Fig. 4F**). The larger number of DAVs identified in organoids is due to additional replicates and smaller batch effects. Consistent with this, the overlap of DAV sets was higher (53% of organoid DAVs, 54% of primary DAVs) when considering only the most differentially active organoid variants (absolute $\log_2$AR > 0.3 and activity above the median of shuffled controls, **Fig. 4F**). Despite this modest concordance in which variants were statistically significant, we observed a high correlation in DAV effect sizes in organoids versus primary cells (r = 0.91, p = 2.2e-16, **Fig. 4H**). We conclude that cerebral organoids and primary cells produce comparable lentiMPRA measurements of differential allelic activity for variants with the largest effects, with noise and cell type differences affecting measurements at and below the significance threshold for identifying DAVs.

Beyond evaluating the suitability of organoids for modeling primary cells, we were also curious about what differences in lentiMPRA results between the two settings would reveal about the biology of early neurodevelopment. Focusing first on the 2,298 DA sequences that were active only in organoids and 2,684 only in primary cells, we performed motif enrichment analysis and examined the expression level of enriched TFs (**Fig. 4I**). Organoid-specific active DA sequences were enriched for binding sites of NKX2.1, RUNX, BCL6, and ASCL2. *BCL6* is a transcriptional repressor with significantly lower expression in organoids compared to primary cells (q-value = 8.86e-7), consistent with our observation that sequences harboring BCL6 motifs tend to have higher lentiMPRA activity in organoids. One such example, includes a dlEN-specific accessible region containing a BCL6 motif that had significantly higher enhancer activity in organoids (**Fig. 4J-K**). In addition, overexpression of *BCL6* is known to inhibit apoptosis (*71*) and therefore could reflect elevated cell stress in organoids (*72*). For the primary-specific active DA peaks, we observed enrichment for GLIS3, STAT6, EHF, and HNF1B motifs. Compared to primary cells, organoids showed higher *GLIS3* expression (q-value = 9.24e-7) and we observed higher lentiMPRA activity in primary cells versus organoids for an Astro/Oligo and IN-MGE DA region containing a GLIS3 motif, suggesting that it may be functioning as a repressor in these primary-specific active sequences (**Fig. 4L-M**). Thus, motif analysis helped us identify TFs whose differential expression between primary cells and organoids is associated with significant shifts in enhancer activity, suggesting repressor versus activator roles for these TFs and underscoring their importance in regulating neurodevelopment.

Next, we examined variants identified as DAVs only in organoids or primary cells. As most variants still showed highly comparable effect sizes in both, we focused on 61 variants showing an opposite direction of effect. We predicted TFBS losses and gains using motifbreakR and compared TF expression using bulk RNA-seq (**Fig. 4N**). Twenty-eight variants were predicted to alter TFBS and 50% of altered TFs showed differential expression between organoid and primary. For example, we found that variant rs112049982 increased enhancer activity in primary cells but decreased activity in organoids. rs112049982 was predicted to improve the binding affinity of OLIG2, a maker gene for oligodendrocytes (Oligo) and oligodendrocyte precursor cells (OPC), which showed significantly lower expression in organoids. This agrees with prior knowledge that Oligo and OPC are extremely rare populations in cerebral organoids (*11*). Thus, we speculate that the discordant effect of the alternative allele of rs112049982 in organoids

versus primary cells could be due to *OLIG2*'s differential expression. Together, these results indicate that despite organoids being a suitable *in vitro* model, differences in the *trans*-regulating environment should be carefully examined when interpreting lentiMPRA results.

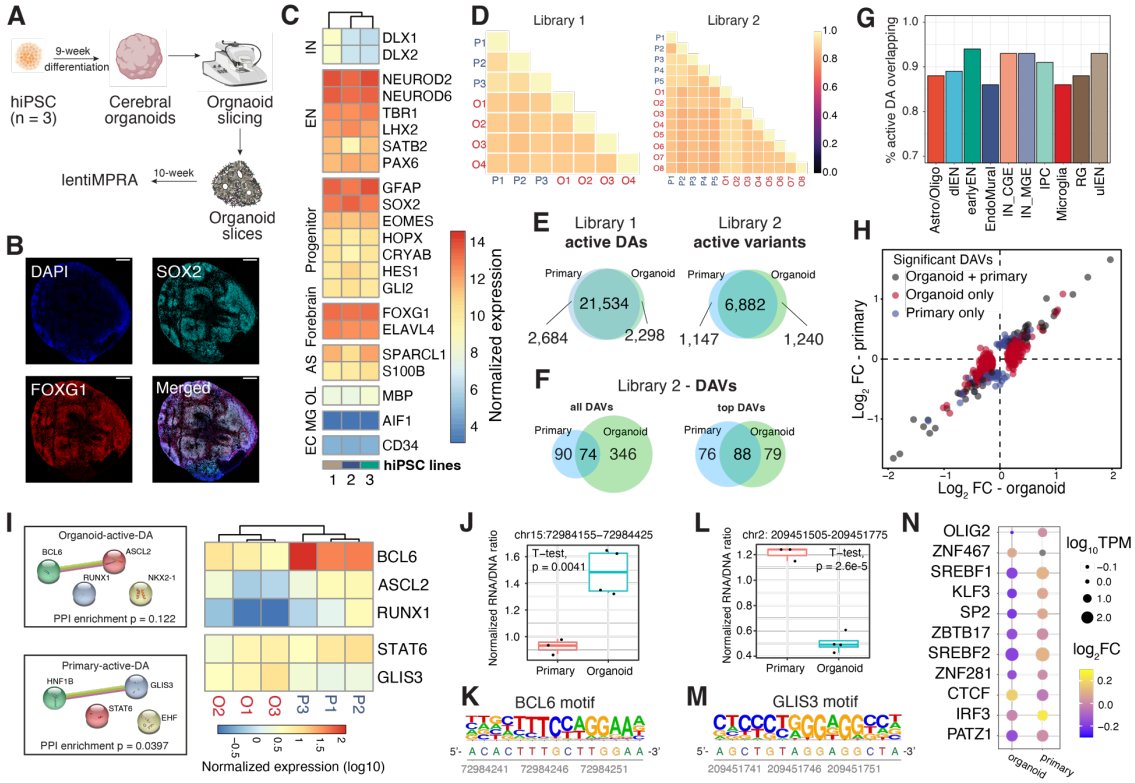

Fig. 4. Comparison of lentiMPRA results in cerebral organoids and primary cortical cells. (**A**) Schematic of the experimental workflow. (**B**). Microscopic images of 10-week-old organoid slices immunostained for SOX2 (Cyan), FOXG1 (Red), and DAPI. Scale bar, 200 µm. (**C**) Normalized transcript count of marker genes in organoids derived from 3 hiPSC lines (1 = 21792A, 2 = 1323_4, 3 = 20961B). (**D**) Correlation of log2(RNA/DNA) between replicates in organoids and primary cortical cells for library 1 and 2, respectively. (**E-F**) Venn diagrams showing the overlap between organoids and primary cells. (**E**) Left: overlap of active DA regions; Right: overlap of active variants. (**F**) Overlap of DAVs. 'Top DAVs' were identified using shuffled sequences to define active and applying a cutoff for absolute $\log_2$FC of 0.3. (**G**) The proportion of active DAs in organoids that are also active in primary cells. (**H**) lentiMPRA $\log_2$FC in organoid (x-axis) and primary cells (y-axis). The scatter plot includes variants identified as DAVs in both organoids and primary cells (grey), variants detected as DAVs only in organoids (red), and variants detected as DAV only in primary cells (blue). (**I**) Left: Protein-protein interactions (PPI) of enriched TFBS motifs in active DAs specific to organoids or primary cells. PPI network generated using STRING (*73*) database. Right: heatmap showing the normalized transcript count of enriched TFs from bulk RNA-seq data. TFs not expressed (TPM < 1) in all replicates were removed from the heatmap. (**J-K**) A DAV (chr15:72984155-72984425, hg38) that contains a BCL6 motif showed increased activity in organoids versus primary cells (**J**) and its reference sequence contains a BCL6 motif (**K**). (**L-M**) A DAV (chr2: 209451505-209451775, hg38) with GLIS3 binding motif showing increased MPRA activity in primary cells versus organoids (**L**) and the location of the GLIS3 motif in its reference sequence shown below (**M**). (**N**) TFBSs altered by DAVs that show an opposite direction of allelic effect between organoids and primary cells. Dot sizes represent normalized TF expression; color represents $\log_2$FC.

A sequence-based deep learning model of lentiMPRA activity reveals neurodevelopmental enhancer motif grammar

Our large dataset of lentiMPRA measurements provides an opportunity to characterize the mid-gestation enhancer code by modeling enhancer activity and then decoding the model's understanding of how sequence variants modulate activity. To do this, we first designed a deep learning regression model that combines a single convolutional layer to learn motif-like sequence features, followed by two recurrent layers to learn the position, spacing, and orientation of motifs (Methods). Sequences were one-hot encoded into matrices (270bp x 4 nucleotides per sequence), and the mean RNA/DNA ratio across replicates was used as the regression target variable. For each library, we trained a model on sequences from all chromosomes except chromosome 3 (used as a validation set to prevent overfitting during training) and chromosome 4 (held out completely for an independent measure of predictive performance). Controls were included in model training. The variant library also included 15,000 sequences that represent a range of expected activity levels due to varying epigenetic similarity to validated brain enhancers in the VISTA database (*74*). On chromosome 4, the DA and variant models achieved 0.82 and 0.78 Pearson correlation, respectively (**Fig. 5A**; 0.81 and 0.7 Spearman correlation). Other similar sequence-to-activity models include MPRA-DragoNN (*75*) trained on human HepG2 and K562 Sharpr-MPRA data (0.28 Spearman correlation), and DeepSTARR (*18*) trained on fruit fly STARR-seq data (0.68 Pearson correlation). Though direct comparisons are not possible due to vast differences in assay type and dataset quality, our held-out predictive performance suggests our model is learning relevant sequence features for predicting MPRA activity.

Convolutional neural networks learn *de novo* filters from DNA sequences that represent position-specific nucleotide frequencies, similar to TFBS motifs. We therefore used the set of sequences that strongly activate each filter to construct a position weight matrix (PWM) and compared these against the HOCOMOCO (v11) database (*76*) to identify significant matches to know binding site motifs (Methods). As many filters have significant matches to motifs (q-value < 0.1) for TFs that are expressed in mid-gestation telencephalon (mean TPM > 1), we estimated each filter's importance for predicting lentiMPRA activity by setting its output to zero and quantifying how much model performance decreases (deltaSSE; Methods). Top-ranked filters that match TFs with high mid-gestation telencephalon expression included TEAD1, NFATC1, STAT3, FOXJ3, POU2F1, and BCL11A (**Fig. 5B**). In addition to these TFs, our method also highlighted several USFs that function as cofactors to improve chromatin accessibility (*47*), consistent with our finding that motifs for these TFs are enriched in active versus inactive sequences (**Fig. 1D**). Thus, our model learned that both universal co-factors and TFs specific to the physiological context of our lentiMPRA experiments contribute to enhancer activity.

To complement this analysis, we performed a large-scale *in silico* mutagenesis (ISM) study. This method enabled us to quantify how individual nucleotide variants affect model predictions and did not directly rely upon a PWM database, though we did use PWM similarity to interpret high-scoring variants. Specifically, we constructed sequences with each possible alternate base at each of the 270 positions in each of the 17,069 variant-containing oligos for a total of 18.4 million alleles. We then predicted the activity of each alternate allele. Examining the distribution of the largest predicted activity change (up or down) per oligo, we found that while the QTLs tested in our lentiMPRA generally have moderate effects on activity, many of the adjacent synthetic

variants tested with ISM have larger effects (**Fig. 5C**). For 11.6% of oligos, predicted activity can be increased by 50% or more through a single nucleotide change. Conversely, 19.7% of oligos can be reduced by 50% or more through a single change. As expected, activity-increasing variants frequently create binding sites for transcriptional activators (e.g., CEBPD) or mutate binding sites for repressors (e.g., FOXK1) that are expressed at mid-gestation, while activity-decreasing variants do the opposite (**Fig. 5D-E**). All sequences contained both increasing and decreasing alleles, and in most cases the two variants with the largest absolute ISM scores have opposing effects on activity. At nucleotides with large absolute ISM scores, the three alternative alleles tend to all be increasing or decreasing as expected if the reference base is a high information content position in a TFBS (**Fig. 5F**).

As an example, we highlight the region around eQTL rs2883420 (**Fig. 5E-F**) that has strong matches for SRY-like motifs. ISM predicts that all three alternative alleles at rs2883420 increase activity (predicted RNA/DNA ~ 0.97). In lentiMPRA, the reference allele was inactive in lentiMPRA (RNA/DNA ~ 0.8), while the alternative allele made the sequence nearly active (RNA/DNA ~ 0.96), fitting with our prediction. Further examination of the sequence effects of this eQTL (**Fig. 5F**), found a strong disruption of motifs for repression-capable TFs, such as SOX2 (*77*) and SRY (*78*) (p-value < 1e-4). ISM also predicts increased activity for non-reference alleles in a TFBS-sized region surrounding rs2883420, and most of these have larger effects than the eQTL, consistent with our genome-wide observations (**Fig. 5C**). These findings indicate that our model is learning *de novo* PWM-like representations of TF motifs which together form a neurodevelopmental regulatory grammar. Such a model can be leveraged to perform ISM, revealing how variants not present in an MPRA library alter enhancer activity and transcription factor binding. This strategy could be extended to discover and design cell-type specific enhancers with precisely tuned activity levels.

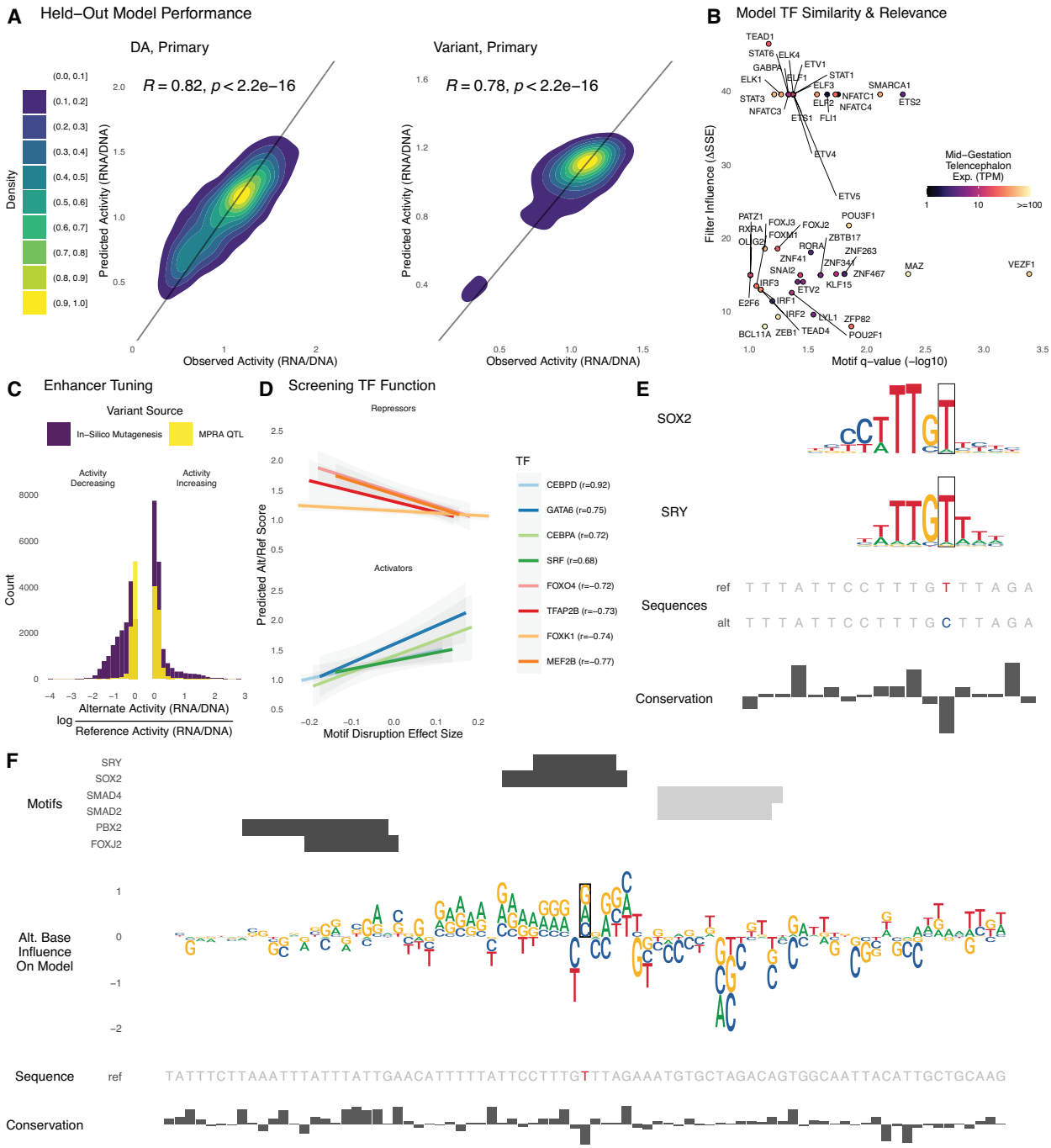

Fig. 5. Sequence determinants of lentiMPRA activity can be modeled with deep learning. For each library, we trained a deep learning model to predict lentiMPRA activity in primary cells from sequence alone. (**A**) Sequences on chromosome 4 were held out from model training and used to evaluate model performance. Predicted and measured activity have high Pearson correlation for the DA library (left) and variant library (right). (**B**) The model learned motifs of neurodevelopmental TFs and used them for accurate predictions. Predictive importance of convolutional filters (change in sum of squared errors when fixing filter output to zero) is plotted against significance of matches to HOCOMOCO motifs (TOMTOM q-value < 0.1) for TFs expressed in mid-gestation telencephalon (mean CPM > 1). (**C**) Applying ISM to the variant library, we found that the activity of most enhancers can be tuned up and down through introduction of alternative alleles. The largest activity-increasing and activity-decreasing alleles for each sequence (purple) tend to have bigger effects than the lentiMPRA measured effects for QTLs (yellow). (**D**) We combined ISM with motifbreakR TFBS disruption scores to screen TFs for repressor versus activator function in

neurodevelopment, using the most activity-increasing alternative allele for each sequence in the variant library. TF's where predicted activity is anti-correlated with motif score tend to repress expression (top) and those with a positive correlation tend to be known activators (bottom). This relationship can be used to decode if the model has learned an activator versus repressor role for TFs that function in both ways. (**E**) The reference T allele of eQTL rs2883420 (lentiMPRA RNA/DNA 0.8) matches motifs of repressors SRY and SOX2, while the alternate C allele disrupts a high information content position in both motifs, resulting in a large activity increase (lentiMPRA RNA/DNA = 0.97, predicted RNA/DNA = 0.96). (**F**) ISM predicts that the other two possible alleles at rs2883420 also increase activity (middle, sequence logo indicates magnitude and direction: up=increasing, down=decreasing). Alternative alleles at adjacent nucleotides overlapping TF motifs (top, positive strand = black, negative strand = gray) have even larger predicted effects on activity. Region shown is chr10:86,851,230-86,851,500 (hg38).

Discussion

Gene regulatory elements have a major impact on human brain development and neurodevelopmental disease. Here, we combined lentiMPRA and deep learning to annotate thousands of regulatory elements in the developing cortex and cerebral organoids. This work provides a large catalog of functional human brain developmental enhancers and disease-associated variants, along with deep learning models that can accurately predict cell-type-specific regulatory regions and variant effects. In addition, it showcases the usability of cerebral organoids as an *in vitro* model for testing regulatory activity in mid-gestation, but also highlights several differences in the *trans*-regulating environment that should be taken into account when interpreting these results.

One caveat of our lentiMPRA is that it is a bulk assay with limited capability to detect cell-type specific signals. For example, we found that abundant cell types, such as neurons and radial glia, had higher percentages of active cell-type specific DA sequences compared to rarer cell populations. For microglia in particular, this could also be due to it being a difficult cell type to infect with lentivirus (*79*), leading to its lower active DA percentage (43.9%). Our validation of eleven regions for cell-type-specific activities in developing brain tissues identified three excitatory neuron-specific enhancers showing expected cell-type specificity, while the rest were active but non-specific. Conversely, we observed lentiMPRA activity for some microglia- and endothelial-specific DA sequences in organoids, despite these cell types being absent or very rare (*11*). We hypothesize this is due to sequences being activated by TFs present in other cell types that do not activate the endogenous sequence due to repressive chromatin. Beyond enhancer assays being permissive and testing sequences outside their endogenous context, other factors that may contribute to unexpected findings for sequences with cell-type specific chromatin accessibility include the limited resolution of scATAC-seq and differences in developmental stages. Nonetheless, nearly half of the 48,861 cell-type specific open chromatin regions that we tested were active enhancers in primary cells and/or organoids.

The cerebral organoids we generated from three hiPSC lines produced lentiMPRA measurements of enhancer activity that were highly correlated with measurements for the same sequences in primary cortical cells. While differential allelic activity was highly correlated for the variants with the largest effects, at least half of the DAVs identified in organoids or primary cells were not statistically significant in the other context with some having opposite allelic effects. We mostly attribute this to differences in sample sizes and cell type proportions. However, we also found that discordant results between organoids and primary cells could shed light on differences in the cellular environment between these two contexts. We identified *BCL6* and *GLIS3* as TFs whose differential expression in primary cells versus organoids can explain why sequences containing their TFBS motifs show significantly differential activity in lentiMPRAs. By looking at whether motifs are positively or negatively correlated with activity, both this analysis and our deep learning-based ISM analysis showed how lentiMPRA data can be used to infer if TFs are acting as repressors versus activators. These computational inferences are needed, because many TFs have both repressive and activating functions (e.g.(*80–82*)), and neurodevelopmental enhancers can play both activating and repressing roles depending on the bound TFs (*46*).

Another clear signal in our computational analyses was the importance of USFs for enhancer activity in early neurodevelopment. Stripe TFs are known to function as transcriptional co-

factors improving chromatin accessibility (*47*). Motif enrichment in our lentiMPRA active sequences identified universal stripe factors alongside many known neurodevelopmental TFs. Corroborating this finding, analysis of our machine learning models identified many predictive features matching stripe motifs. These results suggest that stripe TF motifs may be useful for boosting the activity of designed enhancers, while they might reduce or overshadow signals from cell-type specific TFs. Investigations that study the role of universal stripe factors in MPRAs and *in vivo* will be an important direction for future studies.

We evaluated the regulatory effect of 17,069 brain QTLs and psychiatric disorder associated variants, identifying 164 differentially active variants. This small number is in line with other MPRAs that tested the effect of single-nucleotide variants (*14*, *51*, *53*, *83*, *84*) observing relatively small effects of single nucleotide substitutions, especially common alleles, on regulatory activity. Our deep learning model supports this conclusion; it predicts that many nucleotide changes in the open chromatin regions we tested, including alleles never or rarely seen in people, would show greater differential activity than the brain QTLs did. Another contributing factor is that regulatory variants, unlike coding variants, impact different layers of transcriptional regulation. MRPAs detect variants affecting enhancer activity or perhaps TF binding (*85*) but not those that modulate chromatin states, genome folding, splicing, or other aspects of gene expression. Finally, since about half of the DAVs we detected are in cell-type specific open chromatin regions, we expect that performing lentiMPRA on mixed cell populations limits detection of allelic effects that vary across cellular contexts.

Despite detecting only 164 high-confidence DAVs, integrative analysis of our data with publicly available chromatin interaction data linked many of these DAVs to one or more target genes expressed in neurodevelopment. Predicted target genes of many DAVs are known risk genes or within susceptibility loci, such as *TBR1* and *MARK2* for ASD or *NFKB2* and *SUFU* for SCZ. In particular, for large psychiatric disorder associated loci, our results for 6p21.1, 6p21.2 and 16p11.2 showcase the utility of lentiMPRA to identify potential disorder-associated regulatory variants in a high-throughput manner. In summary, we nominated several differentially active QTLs as potential causal variants of known disease genes/loci, paving the way for developing novel genetic diagnostic and therapeutic tools.

Overall, our work strengthens the utility of using primary cell culture, organoids, and MPRAs to investigate regulatory elements and variants involved in human brain development. In future work, it would be interesting to utilize an organoid lentiMPRA approach to test libraries from various psychiatric disorder-derived iPSCs to identify donor-specific *trans* effects on regulatory activity. iPSC derived organoids are also a promising avenue for investigating brain development in non-human primates, including comparative studies. Another technological development that could be used to expand upon this study is single-cell MPRA (*86*, *87*). While currently limited to a small number of sequences, this approach could eventually overcome some limitations we faced testing cell-type specific DAs in a bulk assay. It will also be critical to leverage CRISPR screens to assess the endogenous activity of candidate enhancer sequences, including those validated for activity with MPRAs. While CRISPR-based tools will likely improve our understanding of regulatory element cell-type specificity, they have their own caveats such as the need for high effect sizes on target gene mRNA levels. Deciphering the regulatory code of

human brain development will require integration of all these strategies. The datasets and models generated in this work are a step in that direction.

Acknowledgments

General: We would like to thank members of the Ahituv, Nowakowski, Pollen and Pollard labs for assistance with this manuscript. We would like to thank Kevin White and Sophia Gaynor for sharing control sets of MPRA. Data were generated as part of the PsychENCODE Consortium, supported by: U01DA048279, U01MH103339, U01MH103340, U01MH103346, U01MH103365, U01MH103392, U01MH116438, U01MH116441, U01MH116442, U01MH116488, U01MH116489, U01MH116492, U01MH122590, U01MH122591, U01MH122592, U01MH122849, U01MH122678, U01MH122681, U01MH116487, U01MH122509, R01MH094714, R01MH105472, R01MH105898, R01MH109677, R01MH109715, R01MH110905, R01MH110920, R01MH110921, R01MH110926, R01MH110927, R01MH110928, R01MH111721, R01MH117291, R01MH117292, R01MH117293, R21MH102791, R21MH103877, R21MH105853, R21MH105881, R21MH109956, R56MH114899, R56MH114901, R56MH114911, R01MH125516, R01MH126459, R01MH129301, R01MH126393, R01MH121521, R01MH116529, R01MH129817, R01MH117406, and P50MH106934 awarded to: Alexej Abyzov, Nadav Ahituv, Schahram Akbarian, Kristin Brennand, Andrew Chess, Gregory Cooper, Gregory Crawford, Stella Dracheva, Peggy Farnham, Michael Gandal, Mark Gerstein, Daniel Geschwind, Fernando Goes, Joachim F. Hallmayer, Vahram Haroutunian, Thomas M. Hyde, Andrew Jaffe, Peng Jin, Manolis Kellis, Joel Kleinman, James A. Knowles, Arnold Kriegstein, Chunyu Liu, Christopher E. Mason, Keri Martinowich, Eran Mukamel, Richard Myers, Charles Nemeroff, Mette Peters, Dalila Pinto, Katherine Pollard, Kerry Ressler, Panos Roussos, Stephan Sanders, Nenad Sestan, Pamela Sklar, Michael P. Snyder, Matthew State, Jason Stein, Patrick Sullivan, Alexander E. Urban, Flora M. Vaccarino, Stephen Warren, Daniel Weinberger, Sherman Weissman, Zhiping Weng, Kevin White, A. Jeremy Willsey, Hyejung Won, and Peter Zandi. Sequencing was partially carried out by the DNA Technologies and Expression Analysis Core at the UC Davis Genome Center, supported by NIH Shared Instrumentation Grant 1S10OD010786-01.

Funding: This work was funded in part by the National Institute of Mental Health (NIMH) grant numbers U01MH116438 (NA and KSP), R01MH109907 (NA and KSP), R01MH123179 (KSP), UF1MH130700 (TN), R01NS123263 (TN), DP2MH122400 (AAP), New York Stem Cell Foundation (TN, AAP), and R01MH125246 (NA), R56MH114911 (FMV), the National Human Genome Research Institute grant number UM1HG011966 (NA), and Coordination for the Improvement of Higher Education Personnel (CAPES/Brazil) -Finance Code 001.

Author contributions:
Conceptualization: CD, SW, PP, AP, TN, NA, KP
Methodology: CD, SW, RZ, PP, NA, KP, DA, DC, SN, FMV
Investigation: CD, SW, MS, RZ
Visualization: CD, SW, MS
Funding acquisition: NA, KP
Project administration: CD, SW, RZ, NA, KP
Supervision: TN, NA, KP

Writing - original draft: CD, SW, MS, NA, KP
Writing -review & editing: All authors

Competing interests:
NA is the cofounder and on the scientific advisory board of Regel Therapeutics and receives funding from BioMarin Pharmaceutical Incorporated.

Data and materials availability:
The source data described in this manuscript are available via the PsychENCODE Knowledge Portal (https://psychencode.synapse.org/). The PsychENCODE Knowledge Portal is a platform for accessing data, analyses, and tools generated through grants funded by the National Institute of Mental Health (NIMH) PsychENCODE Consortium. Data is available for general research use according to the following requirements for data access and data attribution: (https://psychencode.synapse.org/DataAccess). For access to content described in this manuscript see: DOI will be provided prior to publication.

References

1. K. R. Merikangas, J.-P. He, M. Burstein, S. A. Swanson, S. Avenevoli, L. Cui, C. Benjet, K. Georgiades, J. Swendsen, Lifetime prevalence of mental disorders in U.S. adolescents: results from the National Comorbidity Survey Replication--Adolescent Supplement (NCS-A). *J. Am. Acad. Child Adolesc. Psychiatry*. **49**, 980–989 (2010).

2. P. F. Sullivan, D. H. Geschwind, Defining the Genetic, Genomic, Cellular, and Diagnostic Architectures of Psychiatric Disorders. *Cell*. **177**, 162–183 (2019).

3. T. E. Bakken, J. A. Miller, S.-L. Ding, S. M. Sunkin, K. A. Smith, L. Ng, A. Szafer, R. A. Dalley, J. J. Royall, T. Lemon, S. Shapouri, K. Aiona, J. Arnold, J. L. Bennett, D. Bertagnolli, K. Bickley, A. Boe, K. Brouner, S. Butler, E. Byrnes, S. Caldejon, A. Carey, S. Cate, M. Chapin, J. Chen, N. Dee, T. Desta, T. A. Dolbeare, N. Dotson, A. Ebbert, E. Fulfs, G. Gee, T. L. Gilbert, J. Goldy, L. Gourley, B. Gregor, G. Gu, J. Hall, Z. Haradon, D. R. Haynor, N. Hejazinia, A. Hoerder-Suabedissen, R. Howard, J. Jochim, M. Kinnunen, A. Kriedberg, C. L. Kuan, C. Lau, C.-K. Lee, F. Lee, L. Luong, N. Mastan, R. May, J. Melchor, N. Mosqueda, E. Mott, K. Ngo, J. Nyhus, A. Oldre, E. Olson, J. Parente, P. D. Parker, S. Parry, J. Pendergraft, L. Potekhina, M. Reding, Z. L. Riley, T. Roberts, B. Rogers, K. Roll, D. Rosen, D. Sandman, M. Sarreal, N. Shapovalova, S. Shi, N. Sjoquist, A. J. Sodt, R. Townsend, L. Velasquez, U. Wagley, W. B. Wakeman, C. White, C. Bennett, J. Wu, R. Young, B. L. Youngstrom, P. Wohnoutka, R. A. Gibbs, J. Rogers, J. G. Hohmann, M. J. Hawrylycz, R. F. Hevner, Z. Molnár, J. W. Phillips, C. Dang, A. R. Jones, D. G. Amaral, A. Bernard, E. S. Lein, A comprehensive transcriptional map of primate brain development. *Nature*. **535**, 367–375 (2016).

4. I. Iossifov, B. J. O'Roak, S. J. Sanders, M. Ronemus, N. Krumm, D. Levy, H. A. Stessman, K. T. Witherspoon, L. Vives, K. E. Patterson, J. D. Smith, B. Paeper, D. A. Nickerson, J. Dea, S. Dong, L. E. Gonzalez, J. D. Mandell, S. M. Mane, M. T. Murtha, C. A. Sullivan, M. F. Walker, Z. Waqar, L. Wei, A. J. Willsey, B. Yamrom, Y.-H. Lee, E. Grabowska, E. Dalkic, Z. Wang, S. Marks, P. Andrews, A. Leotta, J. Kendall, I. Hakker, J. Rosenbaum, B. Ma, L. Rodgers, J. Troge, G. Narzisi, S. Yoon, M. C. Schatz, K. Ye, W. R. McCombie, J. Shendure, E. E. Eichler, M. W. State, M. Wigler, The contribution of de novo coding mutations to autism spectrum disorder. *Nature*. **515**, 216–221 (2014).

5. F. K. Satterstrom, J. A. Kosmicki, J. Wang, M. S. Breen, S. De Rubeis, J.-Y. An, M. Peng, R. Collins, J. Grove, L. Klei, C. Stevens, J. Reichert, M. S. Mulhern, M. Artomov, S. Gerges, B. Sheppard, X. Xu, A. Bhaduri, U. Norman, H. Brand, G. Schwartz, R. Nguyen, E. E. Guerrero, C. Dias, Autism Sequencing Consortium, iPSYCH-Broad Consortium, C. Betancur, E. H. Cook, L. Gallagher, M. Gill, J. S. Sutcliffe, A. Thurm, M. E. Zwick, A. D. Børglum, M. W. State, A. E. Cicek, M. E. Talkowski, D. J. Cutler, B. Devlin, S. J. Sanders, K. Roeder, M. J. Daly, J. D. Buxbaum, Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell*. **180**, 568–584.e23 (2020).

6. N. E. Clifton, E. Hannon, J. C. Harwood, A. Di Florio, K. L. Thomas, P. A. Holmans, J. T. R. Walters, M. C. O'Donovan, M. J. Owen, A. J. Pocklington, J. Hall, Dynamic expression of genes associated with schizophrenia and bipolar disorder across development. *Transl. Psychiatry*. **9**, 74 (2019).

7. Cross-Disorder Group of the Psychiatric Genomics Consortium, Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet*. **381**, 1371–1379 (2013).

8. Cross-Disorder Group of the Psychiatric Genomics Consortium, Genomic Relationships, Novel

Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. *Cell*. **179**, 1469–1482.e11 (2019).

9. M. J. Gandal, P. Zhang, E. Hadjimichael, R. L. Walker, C. Chen, S. Liu, H. Won, H. van Bakel, M. Varghese, Y. Wang, A. W. Shieh, J. Haney, S. Parhami, J. Belmont, M. Kim, P. Moran Losada, Z. Khan, J. Mleczko, Y. Xia, R. Dai, D. Wang, Y. T. Yang, M. Xu, K. Fish, P. R. Hof, J. Warrell, D. Fitzgerald, K. White, A. E. Jaffe, PsychENCODE Consortium, M. A. Peters, M. Gerstein, C. Liu, L. M. Iakoucheva, D. Pinto, D. H. Geschwind, Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science*. **362** (2018), doi:10.1126/science.aat8127.

10. S. Chatterjee, N. Ahituv, Gene Regulatory Elements, Major Drivers of Human Disease. *Annu. Rev. Genomics Hum. Genet.* **18**, 45–63 (2017).

11. R. S. Ziffra, C. N. Kim, J. M. Ross, A. Wilfert, T. N. Turner, M. Haeussler, A. M. Casella, P. F. Przytycki, K. C. Keough, D. Shin, D. Bogdanoff, A. Kreimer, K. S. Pollard, S. A. Ament, E. E. Eichler, N. Ahituv, T. J. Nowakowski, Single-cell epigenomics reveals mechanisms of human cortical development. *Nature*. **598**, 205–213 (2021).

12. A. E. Trevino, F. Müller, J. Andersen, L. Sundaram, A. Kathiria, A. Shcherbina, K. Farh, H. Y. Chang, A. M. Pașca, A. Kundaje, S. P. Pașca, W. J. Greenleaf, Chromatin and gene-regulatory dynamics of the developing human cerebral cortex at single-cell resolution. *Cell*. **184**, 5053–5069.e23 (2021).

13. F. Inoue, N. Ahituv, Decoding enhancers using massively parallel reporter assays. *Genomics*. **106**, 159–164 (2015).

14. R. Tewhey, D. Kotliar, D. S. Park, B. Liu, S. Winnicki, S. K. Reilly, K. G. Andersen, T. S. Mikkelsen, E. S. Lander, S. F. Schaffner, P. C. Sabeti, Direct Identification of Hundreds of Expression-Modulating Variants using a Multiplexed Reporter Assay. *Cell*. **165**, 1519–1529 (2016).

15. N. S. Abell, M. K. DeGorter, M. J. Gloudemans, E. Greenwald, K. S. Smith, Z. He, S. B. Montgomery, Multiple causal variants underlie genetic associations in humans. *Science*. **375**, 1247–1254 (2022).

16. S. Whalen, F. Inoue, H. Ryu, T. Fair, E. Markenscoff-Papadimitriou, K. Keough, M. Kircher, B. Martin, B. Alvarado, O. Elor, D. Laboy Cintron, A. Williams, M. A. Hassan Samee, S. Thomas, R. Krencik, E. M. Ullian, A. Kriegstein, J. L. Rubenstein, J. Shendure, A. A. Pollen, N. Ahituv, K. S. Pollard, Machine learning dissection of human accelerated regions in primate neurodevelopment. *Neuron* (2023), doi:10.1016/j.neuron.2022.12.026.

17. I. I. Taskiran, K. I. Spanier, V. Christiaens, D. Mauduit, S. Aerts, Cell type directed design of synthetic enhancers. *bioRxiv* (2022), p. 2022.07.26.501466.

18. B. P. de Almeida, F. Reiter, M. Pagani, A. Stark, DeepSTARR predicts enhancer activity from DNA sequence and enables the de novo design of synthetic enhancers. *Nat. Genet.* **54**, 613–624 (2022).

19. D. Shlyueva, G. Stampfel, A. Stark, Transcriptional enhancers: from properties to genome-wide predictions. *Nat. Rev. Genet.* **15**, 272–286 (2014).

20. F. Lu, A. Sossin, N. Abell, S. B. Montgomery, Z. He, Deep learning-assisted genome-wide characterization of massively parallel reporter assays. *Nucleic Acids Res.* **50**, 11442–11454 (2022).

21. E. D. Vaishnav, C. G. de Boer, J. Molinet, M. Yassour, L. Fan, X. Adiconis, D. A. Thompson, J. Z. Levin, F. A. Cubillos, A. Regev, The evolution, evolvability and engineering of gene regulatory DNA. *Nature*. **603**, 455–463 (2022).

22. J. Zhou, C. L. Theesfeld, K. Yao, K. M. Chen, A. K. Wong, O. G. Troyanskaya, Deep learning sequence-based ab initio prediction of variant effects on expression and disease risk. *Nat. Genet.* **50**, 1171–1179 (2018).

23. Ž. Avsec, V. Agarwal, D. Visentin, J. R. Ledsam, A. Grabska-Barwinska, K. R. Taylor, Y. Assael, J. Jumper, P. Kohli, D. R. Kelley, Effective gene expression prediction from sequence by integrating long-range interactions. *Nat. Methods*. **18**, 1196–1203 (2021).

24. G. Fudenberg, D. R. Kelley, K. S. Pollard, Predicting 3D genome folding from DNA sequence with Akita. *Nat. Methods*. **17**, 1111–1117 (2020).

25. J. Zhou, Sequence-based modeling of three-dimensional genome architecture from kilobase to chromosome scale. *Nat. Genet.* **54**, 725–734 (2022).

26. D. R. Kelley, J. Snoek, J. L. Rinn, Basset: learning the regulatory code of the accessible genome with deep convolutional neural networks. *Genome Res.* **26**, 990–999 (2016).

27. Q. Yin, M. Wu, Q. Liu, H. Lv, R. Jiang, DeepHistone: a deep learning approach to predicting histone modifications. *BMC Genomics*. **20**, 193 (2019).

28. J. Zhou, O. G. Troyanskaya, Predicting effects of noncoding variants with deep learning-based sequence model. *Nat. Methods*. **12**, 931–934 (2015).

29. K. M. Chen, A. K. Wong, O. G. Troyanskaya, J. Zhou, A sequence-based global map of regulatory activity for deciphering human genetics. *Nat. Genet.* **54**, 940–949 (2022).

30. C. Chen, J. Hou, X. Shi, H. Yang, J. A. Birchler, J. Cheng, DeepGRN: prediction of transcription factor binding site across cell-types using attention-based deep neural networks. *BMC Bioinformatics*. **22**, 38 (2021).

31. E. Markenscoff-Papadimitriou, S. Whalen, P. Przytycki, R. Thomas, F. Binyameen, T. J. Nowakowski, A. R. Kriegstein, S. J. Sanders, M. W. State, K. S. Pollard, J. L. Rubenstein, A Chromatin Accessibility Atlas of the Developing Human Telencephalon. *Cell*. **182**, 754–769.e18 (2020).

32. M. Song, M.-P. Pebworth, X. Yang, A. Abnousi, C. Fan, J. Wen, J. D. Rosen, M. N. K. Choudhary, X. Cui, I. R. Jones, S. Bergenholtz, U. C. Eze, I. Juric, B. Li, L. Maliskova, J. Lee, W. Liu, A. A. Pollen, Y. Li, T. Wang, M. Hu, A. R. Kriegstein, Y. Shen, Cell-type-specific 3D epigenomes in the developing human cortex. *Nature*. **587**, 644–649 (2020).

33. M. Song, X. Yang, X. Ren, L. Maliskova, B. Li, I. R. Jones, C. Wang, F. Jacob, K. Wu, M. Traglia, T. W. Tam, K. Jamieson, S.-Y. Lu, G.-L. Ming, Y. Li, J. Yao, L. A. Weiss, J. R. Dixon, L. M. Judge, B. R. Conklin, H. Song, L. Gan, Y. Shen, Mapping cis-regulatory chromatin contacts in neural cells links neuropsychiatric disorder risk variants to target genes. *Nat. Genet.* **51**, 1252–1262 (2019).

34. D. Wang, S. Liu, J. Warrell, H. Won, X. Shi, F. C. P. Navarro, D. Clarke, M. Gu, P. Emani, Y. T. Yang, M. Xu, M. J. Gandal, S. Lou, J. Zhang, J. J. Park, C. Yan, S. K. Rhie, K. Manakongtreecheep, H. Zhou, A. Nathan, M. Peters, E. Mattei, D. Fitzgerald, T. Brunetti, J. Moore, Y. Jiang, K. Girdhar,

G. E. Hoffman, S. Kalayci, Z. H. Gümüş, G. E. Crawford, PsychENCODE Consortium, P. Roussos, S. Akbarian, A. E. Jaffe, K. P. White, Z. Weng, N. Sestan, D. H. Geschwind, J. A. Knowles, M. B. Gerstein, Comprehensive functional genomic resource and integrative model for the human brain. *Science*. **362** (2018), doi:10.1126/science.aat8464.

35. D. Liang, A. L. Elwell, N. Aygün, O. Krupa, J. M. Wolter, F. A. Kyere, M. J. Lafferty, K. E. Cheek, K. P. Courtney, M. Yusupova, M. E. Garrett, A. Ashley-Koch, G. E. Crawford, M. I. Love, L. de la Torre-Ubieta, D. H. Geschwind, J. L. Stein, Cell-type-specific effects of genetic variation on chromatin accessibility during human neuronal differentiation. *Nat. Neurosci.* **24**, 941–953 (2021).

36. D. M. Werling, S. Pochareddy, J. Choi, J.-Y. An, B. Sheppard, M. Peng, Z. Li, C. Dastmalchi, G. Santpere, A. M. M. Sousa, A. T. N. Tebbenkamp, N. Kaur, F. O. Gulden, M. S. Breen, L. Liang, M. C. Gilson, X. Zhao, S. Dong, L. Klei, A. E. Cicek, J. D. Buxbaum, H. Adle-Biassette, J.-L. Thomas, K. A. Aldinger, D. R. O'Day, I. A. Glass, N. A. Zaitlen, M. E. Talkowski, K. Roeder, M. W. State, B. Devlin, S. J. Sanders, N. Sestan, Whole-Genome and RNA Sequencing Reveal Variation and Transcriptomic Coordination in the Developing Human Prefrontal Cortex. *Cell Rep.* **31**, 107489 (2020).

37. D. Demontis, R. K. Walters, J. Martin, M. Mattheisen, T. D. Als, E. Agerbo, G. Baldursson, R. Belliveau, J. Bybjerg-Grauholm, M. Bækvad-Hansen, F. Cerrato, K. Chambert, C. Churchhouse, A. Dumont, N. Eriksson, M. Gandal, J. I. Goldstein, K. L. Grasby, J. Grove, O. O. Gudmundsson, C. S. Hansen, M. E. Hauberg, M. V. Hollegaard, D. P. Howrigan, H. Huang, J. B. Maller, A. R. Martin, N. G. Martin, J. Moran, J. Pallesen, D. S. Palmer, C. B. Pedersen, M. G. Pedersen, T. Poterba, J. B. Poulsen, S. Ripke, E. B. Robinson, F. K. Satterstrom, H. Stefansson, C. Stevens, P. Turley, G. B. Walters, H. Won, M. J. Wright, ADHD Working Group of the Psychiatric Genomics Consortium (PGC), Early Lifecourse & Genetic Epidemiology (EAGLE) Consortium, 23andMe Research Team, O. A. Andreassen, P. Asherson, C. L. Burton, D. I. Boomsma, B. Cormand, S. Dalsgaard, B. Franke, J. Gelernter, D. Geschwind, H. Hakonarson, J. Haavik, H. R. Kranzler, J. Kuntsi, K. Langley, K.-P. Lesch, C. Middeldorp, A. Reif, L. A. Rohde, P. Roussos, R. Schachar, P. Sklar, E. J. S. Sonuga-Barke, P. F. Sullivan, A. Thapar, J. Y. Tung, I. D. Waldman, S. E. Medland, K. Stefansson, M. Nordentoft, D. M. Hougaard, T. Werge, O. Mors, P. B. Mortensen, M. J. Daly, S. V. Faraone, A. D. Børglum, B. M. Neale, Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat. Genet.* **51**, 63–75 (2019).

38. I. E. Jansen, J. E. Savage, K. Watanabe, J. Bryois, D. M. Williams, S. Steinberg, J. Sealock, I. K. Karlsson, S. Hägg, L. Athanasiu, N. Voyle, P. Proitsi, A. Witoelar, S. Stringer, D. Aarsland, I. S. Almdahl, F. Andersen, S. Bergh, F. Bettella, S. Bjornsson, A. Brækhus, G. Bråthen, C. de Leeuw, R. S. Desikan, S. Djurovic, L. Dumitrescu, T. Fladby, T. J. Hohman, P. V. Jonsson, S. J. Kiddle, A. Rongve, I. Saltvedt, S. B. Sando, G. Selbæk, M. Shoai, N. G. Skene, J. Snaedal, E. Stordal, I. D. Ulstein, Y. Wang, L. R. White, J. Hardy, J. Hjerling-Leffler, P. F. Sullivan, W. M. van der Flier, R. Dobson, L. K. Davis, H. Stefansson, K. Stefansson, N. L. Pedersen, S. Ripke, O. A. Andreassen, D. Posthuma, Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat. Genet.* **51**, 404–413 (2019).

39. Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium, Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. *Mol. Autism*. **8**, 21 (2017).

40. N. Mullins, A. J. Forstner, K. S. O'Connell, B. Coombes, J. R. I. Coleman, Z. Qiao, T. D. Als, T. B. Bigdeli, S. Børte, J. Bryois, A. W. Charney, O. K. Drange, M. J. Gandal, S. P. Hagenaars, M. Ikeda, N. Kamitaki, M. Kim, K. Krebs, G. Panagiotaropoulou, B. M. Schilder, L. G. Sloofman, S. Steinberg, V. Trubetskoy, B. S. Winsvold, H.-H. Won, L. Abramova, K. Adorjan, E. Agerbo, M. Al

Eissa, D. Albani, N. Alliey-Rodriguez, A. Anjorin, V. Antilla, A. Antoniou, S. Awasthi, J. H. Baek, M. Bækvad-Hansen, N. Bass, M. Bauer, E. C. Beins, S. E. Bergen, A. Birner, C. Bøcker Pedersen, E. Bøen, M. P. Boks, R. Bosch, M. Brum, B. M. Brumpton, N. Brunkhorst-Kanaan, M. Budde, J. Bybjerg-Grauholm, W. Byerley, M. Cairns, M. Casas, P. Cervantes, T.-K. Clarke, C. Cruceanu, A. Cuellar-Barboza, J. Cunningham, D. Curtis, P. M. Czerski, A. M. Dale, N. Dalkner, F. S. David, F. Degenhardt, S. Djurovic, A. L. Dobbyn, A. Douzenis, T. Elvsåshagen, V. Escott-Price, I. N. Ferrier, A. Fiorentino, T. M. Foroud, L. Forty, J. Frank, O. Frei, N. B. Freimer, L. Frisén, K. Gade, J. Garnham, J. Gelernter, M. Giørtz Pedersen, I. R. Gizer, S. D. Gordon, K. Gordon-Smith, T. A. Greenwood, J. Grove, J. Guzman-Parra, K. Ha, M. Haraldsson, M. Hautzinger, U. Heilbronner, D. Hellgren, S. Herms, P. Hoffmann, P. A. Holmans, L. Huckins, S. Jamain, J. S. Johnson, J. L. Kalman, Y. Kamatani, J. L. Kennedy, S. Kittel-Schneider, J. A. Knowles, M. Kogevinas, M. Koromina, T. M. Kranz, H. R. Kranzler, M. Kubo, R. Kupka, S. A. Kushner, C. Lavebratt, J. Lawrence, M. Leber, H.-J. Lee, P. H. Lee, S. E. Levy, C. Lewis, C. Liao, S. Lucae, M. Lundberg, D. J. MacIntyre, S. H. Magnusson, W. Maier, A. Maihofer, D. Malaspina, E. Maratou, L. Martinsson, M. Mattheisen, S. A. McCarroll, N. W. McGregor, P. McGuffin, J. D. McKay, H. Medeiros, S. E. Medland, V. Millischer, G. W. Montgomery, J. L. Moran, D. W. Morris, T. W. Mühleisen, N. O'Brien, C. O'Donovan, L. M. Olde Loohuis, L. Oruc, S. Papiol, A. F. Pardiñas, A. Perry, A. Pfennig, E. Porichi, J. B. Potash, D. Quested, T. Raj, M. H. Rapaport, J. R. DePaulo, E. J. Regeer, J. P. Rice, F. Rivas, M. Rivera, J. Roth, P. Roussos, D. M. Ruderfer, C. Sánchez-Mora, E. C. Schulte, F. Senner, S. Sharp, P. D. Shilling, E. Sigurdsson, L. Sirignano, C. Slaney, O. B. Smeland, D. J. Smith, J. L. Sobell, C. Søholm Hansen, M. Soler Artigas, A. T. Spijker, D. J. Stein, J. S. Strauss, B. Świątkowska, C. Terao, T. E. Thorgeirsson, C. Toma, P. Tooney, E.-E. Tsermpini, M. P. Vawter, H. Vedder, J. T. R. Walters, S. H. Witt, S. Xi, W. Xu, J. M. K. Yang, A. H. Young, H. Young, P. P. Zandi, H. Zhou, L. Zillich, HUNT All-In Psychiatry, R. Adolfsson, I. Agartz, M. Alda, L. Alfredsson, G. Babadjanova, L. Backlund, B. T. Baune, F. Bellivier, S. Bengesser, W. H. Berrettini, D. H. R. Blackwood, M. Boehnke, A. D. Børglum, G. Breen, V. J. Carr, S. Catts, A. Corvin, N. Craddock, U. Dannlowski, D. Dikeos, T. Esko, B. Etain, P. Ferentinos, M. Frye, J. M. Fullerton, M. Gawlik, E. S. Gershon, F. S. Goes, M. J. Green, M. Grigoroiu-Serbanescu, J. Hauser, F. Henskens, J. Hillert, K. S. Hong, D. M. Hougaard, C. M. Hultman, K. Hveem, N. Iwata, A. V. Jablensky, I. Jones, L. A. Jones, R. S. Kahn, J. R. Kelsoe, G. Kirov, M. Landén, M. Leboyer, C. M. Lewis, Q. S. Li, J. Lissowska, C. Lochner, C. Loughland, N. G. Martin, C. A. Mathews, F. Mayoral, S. L. McElroy, A. M. McIntosh, F. J. McMahon, I. Melle, P. Michie, L. Milani, P. B. Mitchell, G. Morken, O. Mors, P. B. Mortensen, B. Mowry, B. Müller-Myhsok, R. M. Myers, B. M. Neale, C. M. Nievergelt, M. Nordentoft, M. M. Nöthen, M. C. O'Donovan, K. J. Oedegaard, T. Olsson, M. J. Owen, S. A. Paciga, C. Pantelis, C. Pato, M. T. Pato, G. P. Patrinos, R. H. Perlis, D. Posthuma, J. A. Ramos-Quiroga, A. Reif, E. Z. Reininghaus, M. Ribasés, M. Rietschel, S. Ripke, G. A. Rouleau, T. Saito, U. Schall, M. Schalling, P. R. Schofield, T. G. Schulze, L. J. Scott, R. J. Scott, A. Serretti, C. Shannon Weickert, J. W. Smoller, H. Stefansson, K. Stefansson, E. Stordal, F. Streit, P. F. Sullivan, G. Turecki, A. E. Vaaler, E. Vieta, J. B. Vincent, I. D. Waldman, T. W. Weickert, T. Werge, N. R. Wray, J.-A. Zwart, J. M. Biernacka, J. I. Nurnberger, S. Cichon, H. J. Edenberg, E. A. Stahl, A. McQuillin, A. Di Florio, R. A. Ophoff, O. A. Andreassen, Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nat. Genet.* **53**, 817–829 (2021).

41. Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium, Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50**, 668–681 (2018).

42. International Obsessive Compulsive Disorder Foundation Genetics Collaborative and OCD Collaborative Genetics Association Studies, Revealing the complex genetic architecture of obsessive-compulsive disorder using meta-analysis. *Mol. Psychiatry*. **23**, 1181–1188 (2018).

43. Schizophrenia Working Group of the Psychiatric Genomics Consortium, Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature*. **604**, 502–508 (2022).

44. Tourette Association of America International Consortium for Genetics, the Gilles de la Tourette GWAS Replication Initiative, the Tourette International Collaborative Genetics Study, and the Psychiatric Genomics Consortium Tourette Syndrome Working Group, Interrogating the Genetic Determinants of Tourette's Syndrome and Other Tic Disorders Through Genome-Wide Association Studies. *Am. J. Psychiatry*. **176**, 217–227 (2019).

45. H. Mostafavi, J. P. Spence, S. Naqvi, J. K. Pritchard, Limited overlap of eQTLs and GWAS hits due to systematic differences in discovery. *bioRxiv* (2022), p. 2022.05.07.491045.

46. A. Amiri, G. Coppola, S. Scuderi, F. Wu, T. Roychowdhury, F. Liu, S. Pochareddy, Y. Shin, A. Safi, L. Song, Y. Zhu, A. M. M. Sousa, PsychENCODE Consortium, M. Gerstein, G. E. Crawford, N. Sestan, A. Abyzov, F. M. Vaccarino, Transcriptome and epigenome landscape of human cortical development modeled in organoids. *Science*. **362** (2018), doi:10.1126/science.aat6720.

47. Y. Zhao, S. V. Vartak, A. Conte, X. Wang, D. A. Garcia, E. Stevens, S. Kyoung Jung, K.-R. Kieffer-Kwon, L. Vian, T. Stodola, F. Moris, L. Chopp, S. Preite, P. L. Schwartzberg, J. M. Kulinski, A. Olivera, C. Harly, A. Bhandoola, E. F. Heuston, D. M. Bodine, R. Urrutia, A. Upadhyaya, M. T. Weirauch, G. Hager, R. Casellas, "Stripe" transcription factors provide accessibility to co-binding partners in mammalian genomes. *Mol. Cell* (2022), doi:10.1016/j.molcel.2022.06.029.

48. M. J. Blow, D. J. McCulley, Z. Li, T. Zhang, J. A. Akiyama, A. Holt, I. Plajzer-Frick, M. Shoukry, C. Wright, F. Chen, V. Afzal, J. Bristow, B. Ren, B. L. Black, E. M. Rubin, A. Visel, L. A. Pennacchio, ChIP-Seq identification of weakly conserved heart enhancers. *Nat. Genet.* **42**, 806–810 (2010).

49. R. C. Bandler, C. Mayer, G. Fishell, Cortical interneuron specification: the juncture of genes, time and geometry. *Curr. Opin. Neurobiol.* **42**, 17–24 (2017).

50. C. V. Weiss, L. Harshman, F. Inoue, H. B. Fraser, D. A. Petrov, N. Ahituv, D. Gokhman, The cis-regulatory effects of modern human-specific variants. *Elife*. **10** (2021), doi:10.7554/eLife.63713.

51. Y. A. Cooper, N. Teyssier, N. M. Dräger, Q. Guo, J. E. Davis, S. M. Sattler, Z. Yang, A. Patel, S. Wu, S. Kosuri, G. Coppola, M. Kampmann, D. H. Geschwind, Functional regulatory variants implicate distinct transcriptional networks in dementia. *Science*. **377**, eabi8654 (2022).

52. GTEx Consortium, Laboratory, Data Analysis &Coordinating Center (LDACC)—Analysis Working Group, Statistical Methods groups—Analysis Working Group, Enhancing GTEx (eGTEx) groups, NIH Common Fund, NIH/NCI, NIH/NHGRI, NIH/NIMH, NIH/NIDA, Biospecimen Collection Source Site—NDRI, Biospecimen Collection Source Site—RPCI, Biospecimen Core Resource—VARI, Brain Bank Repository—University of Miami Brain Endowment Bank, Leidos Biomedical—Project Management, ELSI Study, Genome Browser Data Integration &Visualization—EBI, Genome Browser Data Integration &Visualization—UCSC Genomics Institute, University of California Santa Cruz, Lead analysts:, Laboratory, Data Analysis &Coordinating Center (LDACC):, NIH program management:, Biospecimen collection:, Pathology:, eQTL manuscript working group:, A. Battle, C. D. Brown, B. E. Engelhardt, S. B. Montgomery, Genetic effects on gene expression across human tissues. *Nature*. **550**, 204–213 (2017).

53. M. Kircher, C. Xiong, B. Martin, M. Schubach, F. Inoue, R. J. A. Bell, J. F. Costello, J. Shendure, N. Ahituv, Saturation mutagenesis of twenty disease-associated regulatory elements at single base-pair

resolution. *Nat. Commun.* **10**, 3583 (2019).

54. S. G. Coetzee, G. A. Coetzee, D. J. Hazelett, motifbreakR: an R/Bioconductor package for predicting variant effects at transcription factor binding sites. *Bioinformatics*. **31**, 3847–3849 (2015).

55. D. Szklarczyk, A. L. Gable, D. Lyon, A. Junge, S. Wyder, J. Huerta-Cepas, M. Simonovic, N. T. Doncheva, J. H. Morris, P. Bork, L. J. Jensen, C. von Mering, STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2018).

56. A. Nott, I. R. Holtman, N. G. Coufal, J. C. M. Schlachetzki, M. Yu, R. Hu, C. Z. Han, M. Pena, J. Xiao, Y. Wu, Z. Keulen, M. P. Pasillas, C. O'Connor, C. K. Nickl, S. T. Schafer, Z. Shen, R. A. Rissman, J. B. Brewer, D. Gosselin, D. D. Gonda, M. L. Levy, M. G. Rosenfeld, G. McVicker, F. H. Gage, B. Ren, C. K. Glass, Brain cell type-specific enhancer-promoter interactome maps and disease-risk association. *Science*. **366**, 1134–1139 (2019).

57. J. H. McDermott, D. D. D. Study, J. Clayton-Smith, T. A. Briggs, The TBR1-related autistic-spectrum-disorder phenotype and its clinical spectrum. *Eur. J. Med. Genet.* **61**, 253–256 (2018).

58. X. Zhou, P. Feliciano, C. Shu, T. Wang, I. Astrovskaya, J. B. Hall, J. U. Obiajulu, J. R. Wright, S. C. Murali, S. X. Xu, L. Brueggeman, T. R. Thomas, O. Marchenko, C. Fleisch, S. D. Barns, L. G. Snyder, B. Han, T. S. Chang, T. N. Turner, W. T. Harvey, A. Nishida, B. J. O'Roak, D. H. Geschwind, SPARK Consortium, J. J. Michaelson, N. Volfovsky, E. E. Eichler, Y. Shen, W. K. Chung, Integrating de novo and inherited variants in 42,607 autism cases identifies mutations in new moderate-risk genes. *Nat. Genet.* **54**, 1305–1319 (2022).

59. D. W. Volk, A. E. Moroco, K. M. Roman, J. R. Edelson, D. A. Lewis, The Role of the Nuclear Factor-κB Transcriptional Complex in Cortical Immune Activation in Schizophrenia. *Biol. Psychiatry*. **85**, 25–34 (2019).

60. J. Wang, J. Liu, S. Li, X. Li, J. Yang, X. Dang, C. Mu, Y. Li, K. Li, J. Li, R. Chen, Y. Liu, D. Huang, Z. Zhang, X.-J. Luo, Genetic regulatory and biological implications of the 10q24.32 schizophrenia risk locus. *Brain* (2022), , doi:10.1093/brain/awac352.

61. M. Baudry, Y. Yao, D. Simmons, J. Liu, X. Bi, Postnatal development of inflammation in a murine model of Niemann-Pick type C disease: immunohistochemical observations of microglia and astroglia. *Exp. Neurol.* **184**, 887–903 (2003).

62. Y. Zhang, T. Lu, H. Yan, Y. Ruan, L. Wang, D. Zhang, W. Yue, L. Lu, Replication of association between schizophrenia and chromosome 6p21-6p22.1 polymorphisms in Chinese Han population. *PLoS One*. **8**, e56732 (2013).

63. G. Chai, A. Webb, C. Li, D. Antaki, S. Lee, M. W. Breuss, N. Lang, V. Stanley, P. Anzenberg, X. Yang, T. Marshall, P. Gaffney, K. J. Wierenga, B. H.-Y. Chung, M. H.-Y. Tsang, L. S. Pais, A. K. Lovgren, G. E. VanNoy, H. L. Rehm, G. Mirzaa, E. Leon, J. Diaz, A. Neumann, A. P. Kalverda, I. W. Manfield, D. A. Parry, C. V. Logan, C. A. Johnson, D. T. Bonthron, E. M. A. Valleley, M. Y. Issa, S. F. Abdel-Ghafar, M. S. Abdel-Hamid, P. Jennings, M. S. Zaki, E. Sheridan, J. G. Gleeson, Mutations in Spliceosomal Genes PPIL1 and PRP17 Cause Neurodegenerative Pontocerebellar Hypoplasia with Microcephaly. *Neuron*. **109**, 241–256.e9 (2021).

64. J. Shi, D. F. Levinson, J. Duan, A. R. Sanders, Y. Zheng, I. Pe'er, F. Dudbridge, P. A. Holmans, A. S. Whittemore, B. J. Mowry, A. Olincy, F. Amin, C. R. Cloninger, J. M. Silverman, N. G. Buccola,

W. F. Byerley, D. W. Black, R. R. Crowe, J. R. Oksenberg, D. B. Mirel, K. S. Kendler, R. Freedman, P. V. Gejman, Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature*. **460**, 753–757 (2009).

65. A. Rampino, T. Annese, S. Torretta, R. Tamma, R. M. Falcone, D. Ribatti, Involvement of vascular endothelial growth factor in schizophrenia. *Neurosci. Lett.* **760**, 136093 (2021).

66. S. J. Sanders, X. He, A. J. Willsey, A. G. Ercan-Sencicek, K. E. Samocha, A. E. Cicek, M. T. Murtha, V. H. Bal, S. L. Bishop, S. Dong, A. P. Goldberg, C. Jinlu, J. F. Keaney 3rd, L. Klei, J. D. Mandell, D. Moreno-De-Luca, C. S. Poultney, E. B. Robinson, L. Smith, T. Solli-Nowlan, M. Y. Su, N. A. Teran, M. F. Walker, D. M. Werling, A. L. Beaudet, R. M. Cantor, E. Fombonne, D. H. Geschwind, D. E. Grice, C. Lord, J. K. Lowe, S. M. Mane, D. M. Martin, E. M. Morrow, M. E. Talkowski, J. S. Sutcliffe, C. A. Walsh, T. W. Yu, Autism Sequencing Consortium, D. H. Ledbetter, C. L. Martin, E. H. Cook, J. D. Buxbaum, M. J. Daly, B. Devlin, K. Roeder, M. W. State, Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron*. **87**, 1215–1233 (2015).

67. A. A. Pollen, A. Bhaduri, M. G. Andrews, T. J. Nowakowski, O. S. Meyerson, M. A. Mostajo-Radji, E. Di Lullo, B. Alvarado, M. Bedolli, M. L. Dougherty, I. T. Fiddes, Z. N. Kronenberg, J. Shuga, A. A. Leyrat, J. A. West, M. Bershteyn, C. B. Lowe, B. J. Pavlovic, S. R. Salama, D. Haussler, E. E. Eichler, A. R. Kriegstein, Establishing Cerebral Organoids as Models of Human-Specific Brain Evolution. *Cell*. **176**, 743–756.e17 (2019).

68. E. Di Lullo, A. R. Kriegstein, The use of brain organoids to investigate neural development and disease. *Nat. Rev. Neurosci.* **18**, 573–584 (2017).

69. A. Fiorenzano, E. Sozzi, M. Birtele, J. Kajtez, J. Giacomoni, F. Nilsson, A. Bruzelius, Y. Sharma, Y. Zhang, B. Mattsson, J. Emnéus, D. R. Ottosson, P. Storm, M. Parmar, Single-cell transcriptomics captures features of human midbrain development and dopamine neuron diversity in brain organoids. *Nat. Commun.* **12**, 7302 (2021).

70. X. Qian, Y. Su, C. D. Adam, A. U. Deutschmann, S. R. Pather, E. M. Goldberg, K. Su, S. Li, L. Lu, F. Jacob, P. T. T. Nguyen, S. Huh, A. Hoke, S. E. Swinford-Jackson, Z. Wen, X. Gu, R. C. Pierce, H. Wu, L. A. Briand, H. I. Chen, J. A. Wolf, H. Song, G.-L. Ming, Sliced Human Cortical Organoids for Modeling Distinct Cortical Layer Formation. *Cell Stem Cell*. **26**, 766–781.e9 (2020).

71. T. Kurosu, T. Fukuda, T. Miki, O. Miura, BCL6 overexpression prevents increase in reactive oxygen species and inhibits apoptosis induced by chemotherapeutic reagents in B-cell lymphoma cells. *Oncogene*. **22**, 4459–4468 (2003).

72. A. Bhaduri, M. G. Andrews, W. Mancia Leon, D. Jung, D. Shin, D. Allen, D. Jung, G. Schmunk, M. Haeussler, J. Salma, A. A. Pollen, T. J. Nowakowski, A. R. Kriegstein, Cell stress in cortical organoids impairs molecular subtype specification. *Nature*. **578**, 142–148 (2020).

73. D. Szklarczyk, A. L. Gable, K. C. Nastou, D. Lyon, R. Kirsch, S. Pyysalo, N. T. Doncheva, M. Legeay, T. Fang, P. Bork, L. J. Jensen, C. von Mering, The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* **49**, D605–D612 (2021).

74. A. Visel, S. Minovitsky, I. Dubchak, L. A. Pennacchio, VISTA Enhancer Browser--a database of tissue-specific human enhancers. *Nucleic Acids Res.* **35**, D88–92 (2007).

75. R. Movva, P. Greenside, G. K. Marinov, S. Nair, A. Shrikumar, A. Kundaje, Deciphering regulatory DNA sequences and noncoding genetic variants using neural network models of massively parallel reporter assays. *PLoS One*. **14**, e0218073 (2019).

76. I. V. Kulakovskiy, I. E. Vorontsov, I. S. Yevshin, R. N. Sharipov, A. D. Fedorova, E. I. Rumynskiy, Y. A. Medvedeva, A. Magana-Mora, V. B. Bajic, D. A. Papatsenko, F. A. Kolpakov, V. J. Makeev, HOCOMOCO: towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. *Nucleic Acids Res.* **46**, D252–D259 (2018).

77. Y.-R. Liu, Z. A. Laghari, C. A. Novoa, J. Hughes, J. R. M. Webster, P. E. Goodwin, S. P. Wheatley, P. J. Scotting, Sox2 acts as a transcriptional repressor in neural stem cells. *BMC Neurosci.* **15**, 95 (2014).

78. M. Desclozeaux, F. Poulat, P. de Santa Barbara, J. P. Capony, P. Turowski, P. Jay, C. Méjean, B. Moniot, B. Boizet, P. Berta, Phosphorylation of an N-terminal motif enhances DNA-binding activity of the human SRY protein. *J. Biol. Chem.* **273**, 7988–7995 (1998).

79. M. E. Maes, G. Colombo, R. Schulz, S. Siegert, Targeting microglia with lentivirus and AAV: Recent advances and remaining challenges. *Neurosci. Lett.* **707**, 134310 (2019).

80. M. Bienz, TCF: transcriptional activator or repressor? *Curr. Opin. Cell Biol.* **10**, 366–372 (1998).

81. J. J. Westendorf, Transcriptional co-repressors of Runx2. *J. Cell. Biochem.* **98**, 54–64 (2006).

82. S. Kim, N.-K. Yu, B.-K. Kaang, CTCF as a multifunctional protein in genome regulation and gene expression. *Exp. Mol. Med.* **47**, e166 (2015).

83. J. Choi, T. Zhang, A. Vu, J. Ablain, M. M. Makowski, L. M. Colli, M. Xu, R. C. Hennessey, J. Yin, H. Rothschild, C. Gräwe, M. A. Kovacs, K. M. Funderburk, M. Brossard, J. Taylor, B. Pasaniuc, R. Chari, S. J. Chanock, C. J. Hoggart, F. Demenais, J. H. Barrett, M. H. Law, M. M. Iles, K. Yu, M. Vermeulen, L. I. Zon, K. M. Brown, Massively parallel reporter assays of melanoma risk variants identify MX2 as a gene promoting melanoma. *Nat. Commun.* **11**, 2718 (2020).

84. J. C. McAfee, S. Lee, J. Lee, J. L. Bell, O. Krupa, J. Davis, K. Insigne, M. L. Bond, D. H. Phanstiel, M. I. Love, J. L. Stein, S. Kosuri, H. Won, Systematic investigation of allelic regulatory activity of schizophrenia-associated common variants. *bioRxiv* (2022), , doi:10.1101/2022.09.15.22279954.

85. W. B. Hamilton, Y. Mosesson, R. S. Monteiro, K. B. Emdal, T. E. Knudsen, C. Francavilla, N. Barkai, J. V. Olsen, J. M. Brickman, Dynamic lineage priming is driven via direct enhancer regulation by ERK. *Nature*. **575**, 355–360 (2019).

86. S. Zhao, C. K. Y. Hong, C. A. Myers, D. M. Granas, M. A. White, J. C. Corbo, B. A. Cohen, A single-cell massively parallel reporter assay detects cell-type-specific gene regulation. *Nat. Genet.*, 1–9 (2023).

87. J.-B. Lalanne, S. G. Regalado, S. Domcke, D. Calderon, B. Martin, T. Li, C. C. Suiter, C. Lee, C. Trapnell, J. Shendure, Multiplex profiling of developmental enhancers with quantitative, single-cell expression reporters. *bioRxiv* (2022), p. 2022.12.10.519236.

88. I. Abugessaisa, S. Noguchi, A. Hasegawa, A. Kondo, H. Kawaji, P. Carninci, T. Kasukawa, refTSS: A Reference Data Set for Human and Mouse Transcription Start Sites. *J. Mol. Biol.* **431**, 2407–2422 (2019).

89. S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, P. C. Sham, PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

90. S. T. Sherry, M. H. Ward, M. Kholodov, J. Baker, L. Phan, E. M. Smigielski, K. Sirotkin, dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29**, 308–311 (2001).

91. M. J. Hubisz, K. S. Pollard, Exploring the genesis and functions of Human Accelerated Regions sheds light on their role in human evolution. *Curr. Opin. Genet. Dev.* **29**, 15–21 (2014).

92. K. C. Keough, S. Whalen, F. Inoue, P. F. Przytycki, T. Fair, C. Deng, M. Steyert, H. Ryu, K. Lindblad-Toh, E. Karlsson, Zoonomia Consortium, T. Nowakowski, N. Ahituv, A. Pollen, K. S. Pollard, Three-dimensional genome re-wiring in loci with Human Accelerated Regions. *bioRxiv* (2022), p. 2022.10.04.510859.

93. S. A. Sloan, J. Andersen, A. M. Pașca, F. Birey, S. P. Pașca, Generation and assembly of human brain region-specific three-dimensional cultures. *Nat. Protoc.* **13**, 2062–2085 (2018).

94. Y. Chen, C. A. Tristan, L. Chen, V. M. Jovanovic, C. Malley, P.-H. Chu, S. Ryu, T. Deng, P. Ormanoglu, D. Tao, Y. Fang, J. Slamecka, H. Hong, C. A. LeClair, S. Michael, C. P. Austin, A. Simeonov, I. Singeç, A versatile polypharmacology platform promotes cytoprotection and viability of human pluripotent and differentiated cells. *Nat. Methods*. **18**, 528–541 (2021).

95. M. G. Gordon, F. Inoue, B. Martin, M. Schubach, V. Agarwal, S. Whalen, S. Feng, J. Zhao, T. Ashuach, R. Ziffra, A. Kreimer, I. Georgakopoulos-Soares, N. Yosef, C. J. Ye, K. S. Pollard, J. Shendure, M. Kircher, N. Ahituv, lentiMPRA and MPRAflow for high-throughput functional characterization of gene regulatory elements. *Nat. Protoc.* **15**, 2387–2412 (2020).

96. J. M. Gaspar, NGmerge: merging paired-end reads via novel empirically-derived models of sequencing errors. *BMC Bioinformatics*. **19**, 536 (2018).

97. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. **25**, 1754–1760 (2009).

98. M. E. Ritchie, B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, G. K. Smyth, limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).

99. S. Heinz, C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh, C. K. Glass, Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell*. **38**, 576–589 (2010).

100. J. Reimand, M. Kull, H. Peterson, J. Hansen, J. Vilo, g:Profiler--a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Res.* **35**, W193–200 (2007).

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

102. D. Smedley, S. Haider, B. Ballester, R. Holland, D. London, G. Thorisson, A. Kasprzyk, BioMart--biological queries made easy. *BMC Genomics*. **10**, 22 (2009).

103. M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving,

M. Isard, Others, "TensorFlow: A System for Large-Scale Machine Learning" in *12th USENIX symposium on operating systems design and implementation (OSDI 16)* (2016), pp. 265–283.

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

105. C. Soneson, M. I. Love, M. D. Robinson, Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res.* **4**, 1521 (2015).

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

107. S. Mesman, R. Bakker, M. P. Smidt, Tcf4 is required for correct brain development during embryogenesis. *Mol. Cell. Neurosci.* **106**, 103502 (2020).