ARTICLES
https://doi.org/10.1038/s41588-019-0472-1
nature genetics

# Mapping *cis*-regulatory chromatin contacts in neural cells links neuropsychiatric disorder risk variants to target genes

Michael Song[1,2], Xiaoyu Yang[1], Xingjie Ren[1], Lenka Maliskova[1], Bingkun Li[1], Ian R. Jones[1], Chao Wang[3], Fadi Jacob[4,5,6], Kenneth Wu[7], Michela Traglia[8], Tsz Wai Tam[1], Kirsty Jamieson[1], Si-Yao Lu[9,10,11], Guo-Li Ming[4,5,12,13,14], Yun Li[15,16,17], Jun Yao[9,10,11], Lauren A. Weiss[1,8], Jesse R. Dixon[18], Luke M. Judge[7,19], Bruce R. Conklin[7,20,21], Hongjun Song[4,5,12,13,22], Li Gan[3,23,24,25] and Yin Shen[1,2,25]*

**Mutations in gene regulatory elements have been associated with a wide range of complex neuropsychiatric disorders. However, due to their cell-type specificity and difficulties in characterizing their regulatory targets, the ability to identify causal genetic variants has remained limited. To address these constraints, we perform an integrative analysis of chromatin interactions, open chromatin regions and transcriptomes using promoter capture Hi-C, assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq) and RNA sequencing, respectively, in four functionally distinct neural cell types: induced pluripotent stem cell (iPSC)-induced excitatory neurons and lower motor neurons, iPSC-derived hippocampal dentate gyrus-like neurons and primary astrocytes. We identify hundreds of thousands of long-range *cis*-interactions between promoters and distal promoter-interacting regions, enabling us to link regulatory elements to their target genes and reveal putative processes that are dysregulated in disease. Finally, we validate several promoter-interacting regions by using clustered regularly interspaced short palindromic repeats (CRISPR) techniques in human excitatory neurons, demonstrating that *CDK5RAP3*, *STRAP* and *DRD2* are transcriptionally regulated by physically linked enhancers.**

A large number of genetic variants associated with diverse human traits and diseases are located in putative regulatory regions. Genetic lesions in these regulatory elements can contribute to complex human disease by modulating gene expression and disrupting finely tuned transcriptional networks. However, deciphering the roles of noncoding variants in disease etiology remains nontrivial due to their lack of annotation in the physiologically relevant cell types. Furthermore, regulatory elements often interact with their target genes over long genomic distances, precluding a straightforward mapping of regulatory element connectivity and limiting the interpretation of noncoding variants from genome-wide association studies (GWAS). Typically, neighboring genes are assigned as risk loci for noncoding variants. However, this nearest gene model is challenged by both experimental and computational evidence[1,2]. For instance, two independent obesity-associated SNPs in the *FTO* gene have been shown not to regulate *FTO* but to regulate *IRX3* in the brain and both *IRX3* and *IRX5* in adipocytes[3,4]. The *FTO* locus in obesity illustrates the potentially intricate and cell-type-specific manner in which noncoding variants contribute to disease. However, such well-annotated cases are rare, and we still lack systematic mapping of GWAS SNPs to their regulatory targets, especially in the context of complex neuropsychiatric disorders.

Previous epigenomic annotations of the germinal zone and cortical and subcortical plates in the human brain revealed the importance of three-dimensional (3D) chromatin structure in gene regulation and disease[5,6]. However, these studies used complex, heterogeneous tissues, limiting their abilities to interpret gene regulation in a cell-type-specific manner. Therefore, charting the landscape of epigenomic regulation in well-characterized, physiologically relevant cell types should offer substantial advantages for identifying causal variants, deciphering their functions

[1]Institute for Human Genetics, University of California, San Francisco, CA, USA. [2]Pharmaceutical Sciences and Pharmacogenomics Graduate Program, University of California, San Francisco, CA, USA. [3]Gladstone Institute of Neurological Disease, University of California, San Francisco, CA, USA. [4]Department of Neuroscience, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [5]Mahoney Institute for Neurosciences, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [6]Solomon H. Snyder Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [7]Gladstone Institute of Cardiovascular Disease, University of California, San Francisco, CA, USA. [8]Department of Psychiatry, University of California, San Francisco, CA, USA. [9]Tsinghua-Peking Center for Life Sciences, Tsinghua University, Beijing, China. [10]State Key Laboratory of Membrane Biology, Tsinghua University, Beijing, China. [11]IDG/McGovern Institute for Brain Research, Tsinghua University, Beijing, China. [12]Department of Cell and Developmental Biology, University of Pennsylvania, Philadelphia, PA, USA. [13]Institute for Regenerative Medicine, University of Pennsylvania, Philadelphia, PA, USA. [14]Department of Psychiatry, University of Pennsylvania, Philadelphia, PA, USA. [15]Department of Genetics, University of North Carolina, Chapel Hill, NC, USA. [16]Department of Biostatistics, University of North Carolina, Chapel Hill, NC, USA. [17]Department of Computer Science, University of North Carolina, Chapel Hill, NC, USA. [18]Salk Institute for Biological Studies, La Jolla, CA, USA. [19]Department of Pediatrics, University of California, San Francisco, CA, USA. [20]Department of Medicine, University of California, San Francisco, CA, USA. [21]Department of Ophthalmology, University of California, San Francisco, CA, USA. [22]Epigenetics Institute, University of Pennsylvania, Philadelphia, PA, USA. [23]Helen and Robert Appel Alzheimer's Disease Research Institute, Weill Cornell Medicine, New York, NY, USA. [24]Feil Family Brain and Mind Research Institute, Weill Cornell Medicine, New York, NY, USA. [25]Department of Neurology, University of California, San Francisco, CA, USA. *e-mail: yin.shen@ucsf.edu

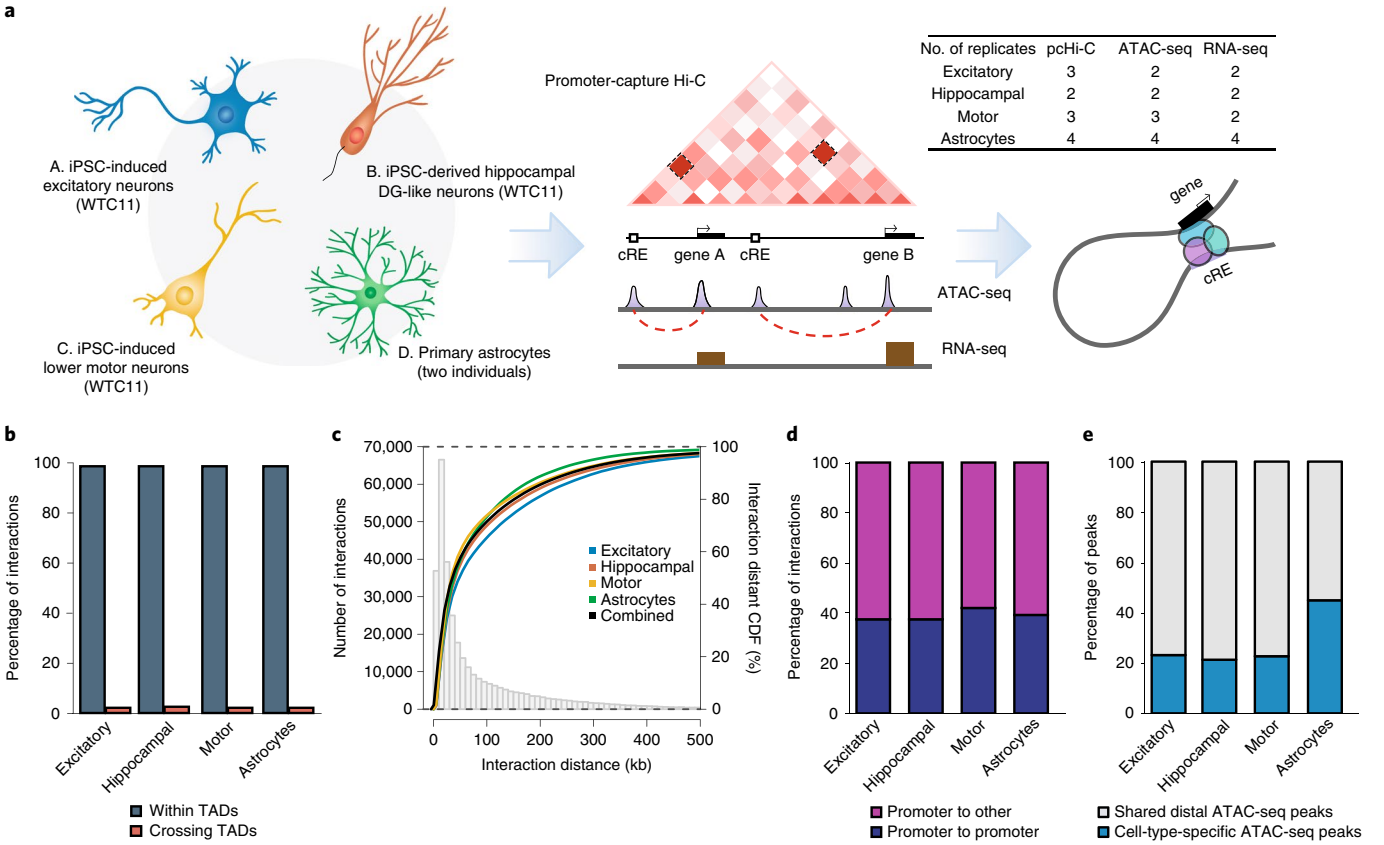


**Fig. 1 | Genome-wide mapping of physical chromatin interactions in functionally distinct neural cell types. a**, Schematic of the study design for generating four functionally distinct cell types in the CNS and performing integrative analysis of chromatin interactions using pcHi-C, open chromatin regions using ATAC-seq and transcriptomes using RNA-seq. The number of biological replicates based on independent experiments for each cell type is shown for each assay. **b**, Proportions of interactions occurring within TADs for each cell type. **c**, Histogram and empirical cumulative distribution function (CDF) plots of interaction distances for each cell type. **d**, Proportions of interactions between promoter-containing bins (blue) and between promoter- and non-promoter-containing bins (purple) for each cell type. **e**, Proportions of cell-type-specific (blue) and shared (gray) distal open chromatin peaks at PIRs for each cell type.

and enabling novel therapies. Toward this goal, we used wild type human iPSCs (WTC11 line[7]) to generate three neuronal cell types: excitatory neurons[8], hippocampal dentate gyrus (DG)-like neurons[9] and lower motor neurons[10]. Glial fibrillary acidic protein (GFAP)-positive astrocytes from the brains of two individuals were also included for their relevance to human brain development and disease. By performing integrative analysis of promoter-centric, long-range chromatin interactions, open chromatin regions and transcriptomes (Fig. 1a), we provide comprehensive annotations for promoters and distal promoter-interacting regions (PIRs) in each cell type. We identify putative gene targets for both in vivo-validated enhancer elements from the VISTA Enhancer Browser[11] and disease-associated variants, enabling the functional validation of PIRs driving diverse processes in cellular identity and disease.

## Results

**Characterizing the epigenomic landscape of long-range chromatin interactions in human neural cells.** To investigate general epigenomic features for cells in the human central nervous system (CNS), we focused on isogenic iPSC-induced excitatory neurons, iPSC-derived hippocampal DG-like neurons and iPSC-induced lower motor neurons, three neuronal subtypes which are currently impractical to isolate from primary tissue. Excitatory neurons were induced from a wild type male iPSC line (WTC11) containing an integrated, isogenic, and inducible neurogenin-2 (*Ngn2*) cassette (i$^3$N iPSCs) with doxycycline-inducible *Ngn2* at the AAVS1 safe-harbor locus[8]. The i$^3$N iPSCs were used to prepare homogenous cultures of excitatory neurons expressing the glutamatergic neuron marker VGLUT1 and the cortical neuron marker CUX1 (refs. [8,12]), though *FOXG1* expression was not detected (Supplementary Fig. 1a,b). Hippocampal DG-like neurons expressing the DG granule cell marker PROX1 were differentiated from a WTC11 line using factors as described previously[9,13] (Supplementary Fig. 1a,b). Finally, lower motor neurons were induced from WTC11 cells containing integrated, isogenic, and inducible *NGN2*, *ISL1* and *LHX3* at the AAVS1 safe-harbor locus (i$^3$LMN iPSCs)[10]. The cells exhibited homogenous expression of the lower motor neuron markers HB9 and SMI32 (Supplementary Fig. 1a,b). All three neuronal subtypes expressed the synaptic genes *SYN1* and *SYN2*, the NMDA receptor genes *GRIN1* and *GRIN2A* and the AMPA receptor genes *GRIA1* and *GRIA2*, evidencing mature synaptic functions (Supplementary Fig. 1b). We also included two batches of astrocytes isolated from 19-week-old male fetal brain samples using GFAP as a selection marker (ScienCells). Astrocytes were cultured for two or fewer passages in vitro and confirmed for positive expression of GFAP prior to harvesting (Supplementary Fig. 1a). Based on the age of the donors and transcriptional signatures of dozens of marker genes distinguishing astrocyte progenitor cells (APCs) (for example, *AGXT2L1* and *WIF1*) from mature astrocytes (for example, *TOP2A* and *TNC*)[14], the astrocytes were determined to be APCs (Supplementary Fig. 1b).

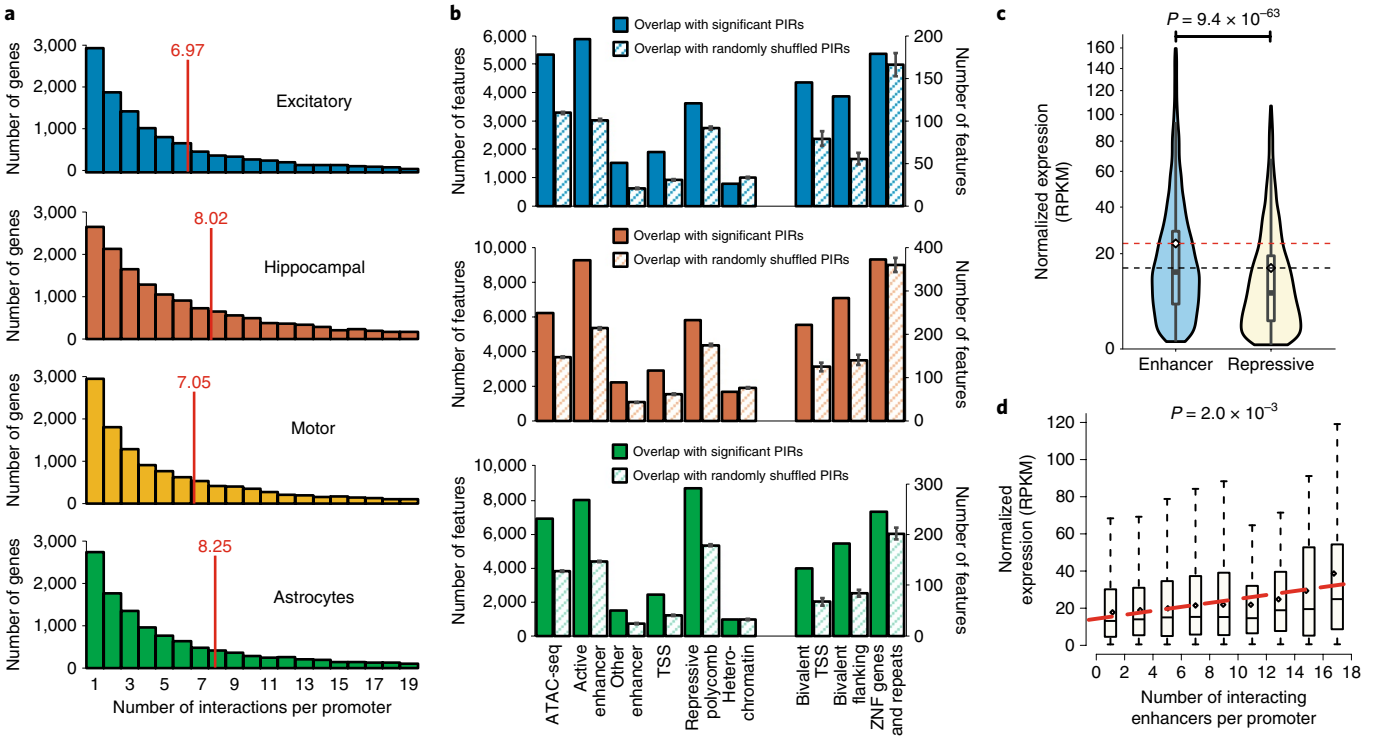


**Fig. 2 | Integrative analysis of chromatin interactions, epigenomic features and gene expression. a**, Histograms of the number of PIRs interacting with each promoter for each cell type. Means are indicated. Only protein coding and noncoding RNA promoters interacting with at least one PIR are included (15,316 promoters in excitatory neurons, 19,546 promoters in hippocampal DG-like neurons, 14,990 promoters in lower motor neurons and 15,397 promoters in astrocytes). **b**, Bar plots showing counts of epigenomic chromatin states inferred using ChromHMM in matched tissues overlapping significant (solid bars) versus randomly shuffled (striped bars) PIRs for each cell type. Means and the s.e.m. for the number of overlaps across $n=100$ sets of randomly shuffled PIRs are shown. **c**, Comparative gene expression analysis across all cell types for expressed genes (normalized reads per kilobase of transcript per million mapped reads (RPKM) > 0.5) whose promoters interact exclusively with either enhancer-PIRs ($n=6,836$ genes) or repressive-PIRs ($n=2,612$ genes) ($P=9.4\times10^{-63}$, $t=16.9$, d.f. $=6,854.6$, two-tailed two-sample $t$-test). Violin plots show the distributions of gene expression values within each group, and boxplots indicate the median, interquartile range (IQR), Q1 − 1.5 × IQR and Q3 + 1.5 × IQR. Means are indicated with dotted horizontal lines. **d**, Distributions of gene expression values across all cell types for expressed genes (normalized RPKM > 0.5) grouped according to the numbers of interactions their promoters form with enhancer-PIRs. Boxplots indicate the median, IQR, Q1 − 1.5 × IQR and Q3 + 1.5 × IQR. Linear regression was performed on the mean gene expression values for $n=9$ bins containing at least ten genes ($P=2.1\times10^{-3}$, $F_{1,7}=22.7$, $F$-test for linear regression).

We prepared promoter capture Hi-C (pcHi-C), ATAC-seq and RNA sequencing (RNA-seq) libraries using two to four biological replicates based on independent experiments for each cell type (Fig. 1a and Supplementary Table 1). Specifically, promoter-centric, long-range chromatin interactions were mapped using a set of 280,445 RNA probes targeting the promoters of 19,603 coding genes in GENCODE 19 (ref. [15]). We first confirmed the reproducibility of contact frequency and saturation of inter-replicate correlation for our pcHi-C libraries using HiCRep[16] (Supplementary Fig. 2c,d). Hierarchical clustering of ATAC-seq read density and gene expression similarly grouped the replicates by cell type (Supplementary Fig. 2a,b), evidencing minimal variations during the cell derivation process. Using CHiCAGO[17], we identified significant chromatin interactions with a score of ≥5 at 195,322 unique interacting loci across all four cell types, with 73,890, 108,156, 66,978 and 84,087 significant interactions being represented in the excitatory neurons, hippocampal DG-like neurons, lower motor neurons and astrocytes, respectively (Supplementary Table 2). Overall, 17,065 or 83.9% of coding gene promoters participate in interactions in at least one cell type (Supplementary Fig. 1c), with 80% of PIRs interacting within a distance of 160 kb (Fig. 1c and Supplementary Fig. 3a). Over 97% of interactions occur within topologically associating domains (TADs) in human fetal brain tissues[6] (Fig. 1b). Furthermore, approximately 40% of interactions occur between promoter-containing bins, while 60% occur between promoter- and non-promoter-containing bins (Fig. 1d). The observed numbers of promoter–promoter interactions can potentially be attributed to transcriptional factories of coregulated genes, the widespread colocalization of promoters[18,19] and the capacity of many promoters to doubly function as enhancers[20,21]. Finally, up to 40% of interacting distal open chromatin peaks are specific to each cell type (Fig. 1e), suggesting that PIRs are capable of orchestrating cell-type-specific gene regulation. Astrocytes exhibit the largest proportion of cell-type-specific open chromatin peaks, likely reflecting basic differences between the neuronal and glial lineages.

The majority of promoters interact with more than one PIR (Fig. 2a), consistent with the large number of regulatory elements in the human genome[22] and previous findings that promoters can be regulated by multiple enhancers[23]. To examine global chromatin signatures at PIRs, we leveraged chromatin states inferred by ChromHMM[24] in matched human brain tissues from the Roadmap Epigenomics Project[25] (dorsolateral prefrontal cortex for excitatory neurons, hippocampus middle for hippocampal DG-like neurons and normal human astrocytes for astrocytes). We show that PIRs are highly enriched for active chromatin features including open chromatin peaks, enhancers and transcriptional start sites (TSSs) while simultaneously exhibiting depletion for repressive heterochromatin marks (Fig. 2b). PIRs are also enriched for

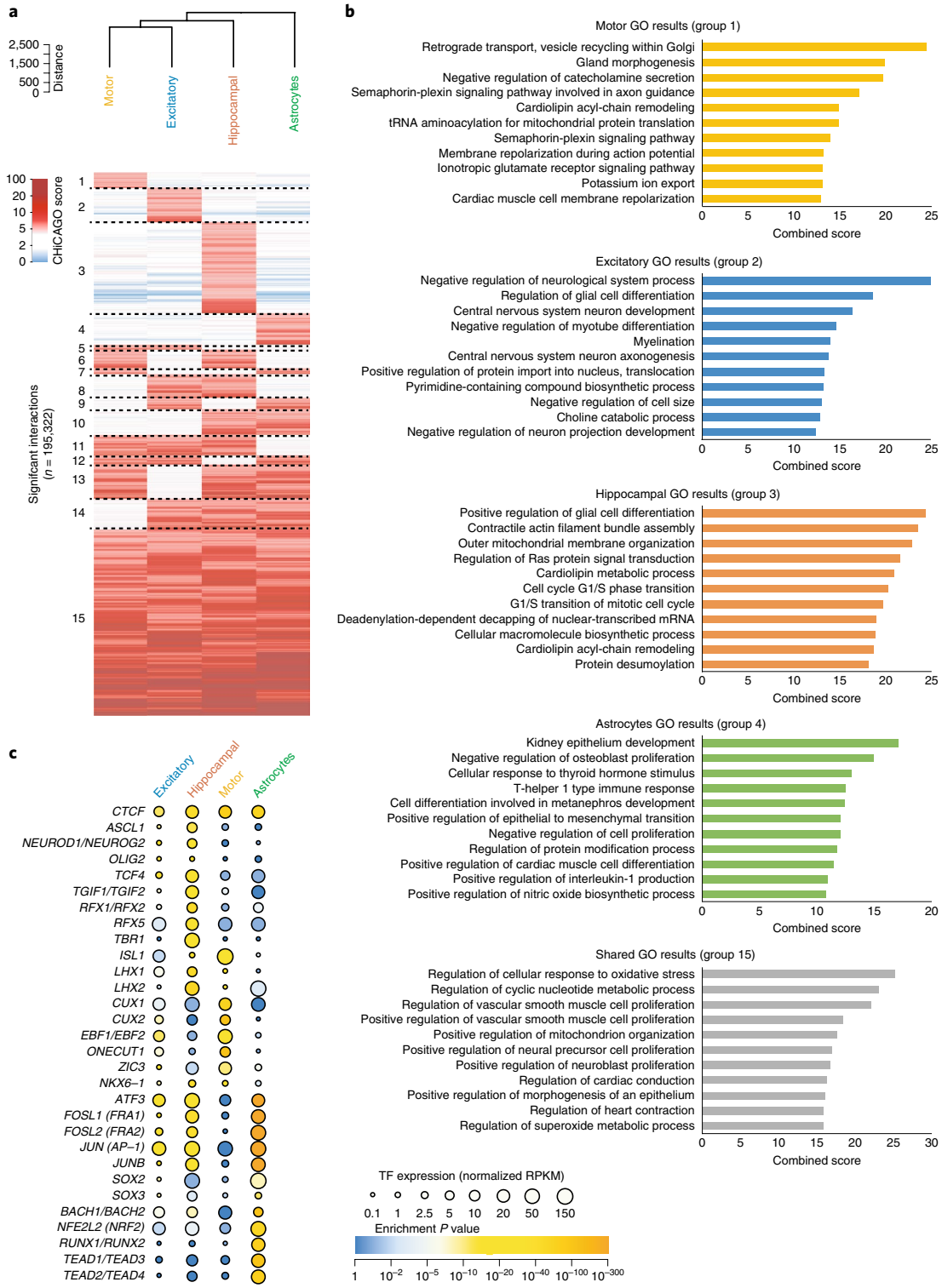


**Fig. 3 | Cell-type-specific PIRs and TF motif enrichment analysis.** **a**, Classification of significant promoter-PIR interactions with a score of ≥5 in at least one cell type based on their overall cell-type specificities. Counts of interactions in each specificity category are summarized in Supplementary Fig. 3a. Cell types are hierarchically clustered based on the Euclidean distances for their interaction scores across all interacting loci. **b**, Top enriched GO terms from the 'GO Biological Process 2018' ontology in Enrichr for genes participating in cell-type-specific (groups 1–4) versus shared (group 15) interactions with distal open chromatin peaks. In total, 459, 837, 217, 307 and 1,925 genes were used as inputs for groups 1–4 and 15, respectively. Enriched GO terms are ranked by their combined scores (calculated by multiplying the log of the *P* value via Fisher's exact test with the *z*-score of the deviation from the expected rank). Expanded lists of enriched GO terms are available in Supplementary Table 3. **c**, Enrichment of consensus TF motif sequences at open chromatin peaks in cell-type-specific PIRs by motifs (rows) and cell types (columns). In total, 1,145, 1,271, 843 and 2,566 peaks were used as inputs for the excitatory neurons, hippocampal DG-like neurons, lower motor neurons and astrocytes, respectively. The color of each dot represents the degree of enrichment (calculated using the cumulative binomial distribution in HOMER) for each motif and cell type, and the size of each dot represents the gene expression of the corresponding TFs for each motif. Entries with similar or identical consensus TF motif sequences are grouped for brevity.

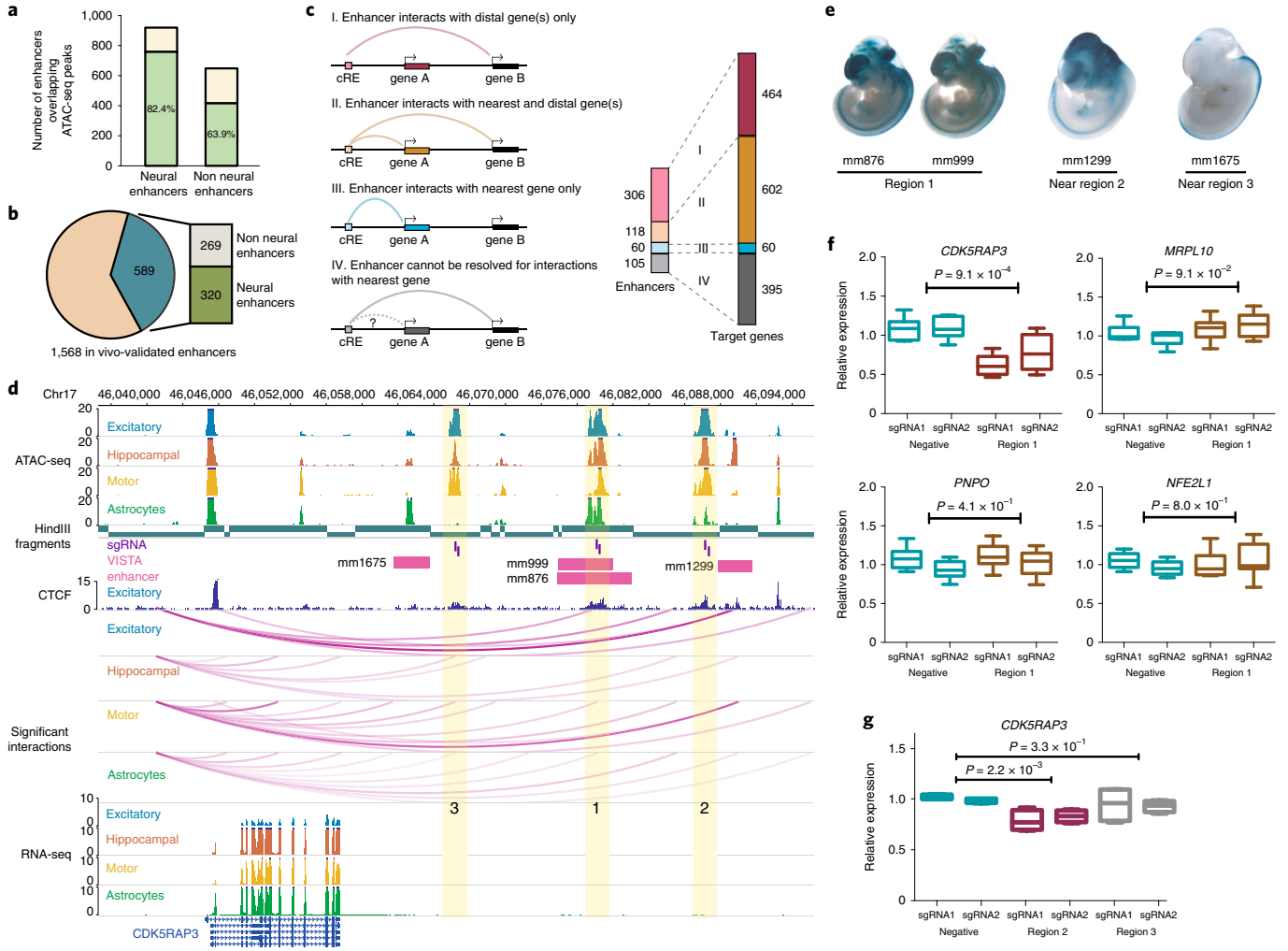


**Fig. 4 | Validation of PIRs in human neural cells.** **a**, In vivo-validated enhancer elements with neural annotations overlap a higher proportion of open chromatin peaks in the neural cells (757 of 919 elements) compared to enhancer elements with nonneural annotations (415 of 649 elements) ($P < 2.2 \times 10^{-16}$, $\chi^2 = 67.5$, d.f. = 1, Pearson's chi-squared test with Yates's correction). **b**, Counts of enhancer elements participating in chromatin interactions (589 of 1,568 elements) with neural and nonneural annotations. **c**, Counts of enhancer elements interacting exclusively with their nearest genes (blue), more distal genes (pink) or both (orange), and the number of target genes for each scenario (right). **d**, Open chromatin peaks in cell-type-specific PIRs (regions 1, 2 and 3) interact with the *CDK5RAP3* promoter. Both enhancer elements (pink) and CTCF binding sites (dark blue) in excitatory neurons are localized to all three regions, and all interactions occur within a TAD in the cortical plate (chr17:45,920,000–47,480,000). **e**, LacZ staining in mouse embryos reveals tissue-specific patterns of enhancer activity. **f**, CRISPRi silencing of region 1 results in significant downregulation of *CDK5RAP3* expression in excitatory neurons ($P = 9.1 \times 10^{-4}$, $t = 4.65$, d.f. = 10, two-tailed two-sample *t*-test). The neighboring genes *MRPL10*, *PNPO* and *NFE2L1* were unaffected ($P = 9.1 \times 10^{-2}$, $t = 1.87$, d.f. = 10, $P = 4.1 \times 10^{-1}$, $t = 0.853$, d.f. = 10 and $P = 8.0 \times 10^{-1}$, $t = 0.259$, d.f. = 10, respectively, two-tailed two-sample *t*-test). Three independent replicates per condition and two sgRNAs per replicate were used for each experiment. Boxplots indicate the median, IQR, minimum and maximum. **g**, CRISPRi silencing of region 2, but not region 3, results in significant downregulation of *CDK5RAP3* expression in excitatory neurons ($P = 2.2 \times 10^{-3}$, $t = 5.11$, d.f. = 6 and $P = 3.3 \times 10^{-1}$, $t = 1.05$, d.f. = 6, respectively, two-tailed two-sample *t*-test). Two independent replicates per condition and two sgRNAs per replicate were used for each experiment. Boxplots indicate the median, IQR, minimum and maximum.

H3K27ac and CTCF binding sites mapped using CUT&RUN[26] in excitatory and lower motor neurons and chromatin immunoprecipitation sequencing (ChIP-seq) in astrocytes from ENCODE[27] (Supplementary Fig. 3b). Promoters interacting with enhancer-PIRs exhibit elevated levels of transcription compared to those interacting with repressive-PIRs ($P = 9.4 \times 10^{-63}$, two-tailed two-sample *t*-test) (Fig. 2c and Supplementary Fig. 3c). Multiple enhancer-PIRs also present evidence for additive effects on transcription. By grouping genes according to the number of interactions their promoters form with enhancer-PIRs, a modest correlation is observed between the number of interactions and the mean gene expression in each group ($P = 2.1 \times 10^{-3}$, *F*-test for linear regression) (Fig. 2d and Supplementary Fig. 3d). Our results demonstrate that chromatin interactions identify PIRs that are enriched for regulatory features and can alter gene expression.

**PIRs contribute to cellular identity.** Chromatin interactions exhibit distinct patterns of cell-type specificity, with tens of thousands of interactions observed to be specific for each cell type (Fig. 3a and Supplementary Fig. 4a). These interactions may underlie important functional differences between the cell types, with gene ontology (GO) enrichment analysis[28] for genes interacting with cell-type-specific PIRs yielding terms associated with neuronal function in the neuronal subtypes and immune function in the

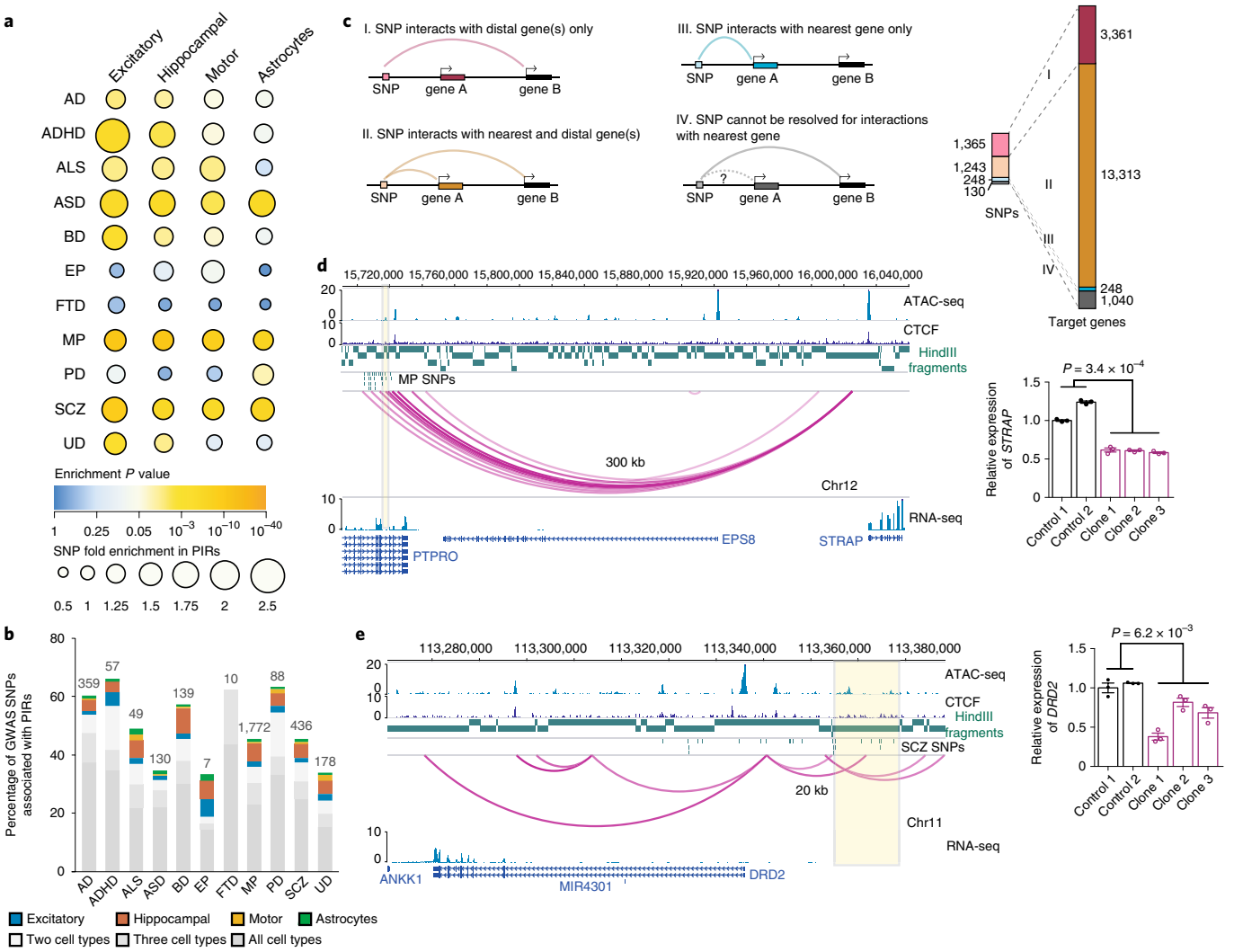


**Fig. 5 | Genetic analysis of chromatin interactions with complex neuropsychiatric disorder-associated variants.** **a**, Enrichment analysis for 11 complex neuropsychiatric disorders or traits. The color and size of each dot represent the enrichment *P* value (two-tailed one-sample *z*-test) and the raw fold enrichment (determined as the number of SNPs overlapping significant PIRs divided by the mean number of SNPs overlapping $n = 100$ sets of randomly shuffled PIRs), respectively. The total numbers of SNPs are available in Supplementary Table 6. **b**, Proportions and counts of GWAS SNPs with at least one linked SNP participating in chromatin interactions. **c**, Counts of GWAS SNPs across all diseases with at least one linked SNP interacting exclusively with their nearest genes (blue), more distal genes (pink) or both (orange), and the number of target genes for each scenario (right). **d**, PIRs with MP SNPs in an intron for *PTPRO* interact with the *STRAP* promoter. All interactions occur within a TAD in the cortical plate (chr12:14,960,000–16,040,000). Biallelic deletion of this PIR in three independent clones results in significant downregulation of *STRAP* expression in excitatory neurons ($P = 3.4 \times 10^{-4}$, $t = 18.5$, d.f. = 3, two-tailed two-sample *t*-test). Error bars represent the s.e.m. **e**, A PIR containing SCZ SNPs interacts with the *DRD2* promoter. All interactions occur within a TAD in the cortical plate (chr11:113,200,000–114,160,000). Monoallelic deletion of this PIR in three independent clones results in significant downregulation of *DRD2* expression in excitatory neurons ($P = 6.2 \times 10^{-3}$, $t = 6.92$, d.f. = 3, two-tailed two-sample *t*-test). Error bars represent the s.e.m.

astrocytes (Fig. 3b and Supplementary Table 3). Meanwhile, 58,809 or 30.1% of unique interactions are shared across all four cell types, with neural precursor cell proliferation and neuroblast proliferation ranking among the top terms for genes participating in shared interactions. In conjunction with the observed enrichment of active chromatin signatures at PIRs, their association with cell-type-specific processes suggests that PIRs harbor lineage-specific regulatory roles. Indeed, numerous promoters of differentially expressed genes, including *OPHN1* in hippocampal DG-like neurons, *CHAT* in lower motor neurons and *TLR4* in astrocytes, form specific contacts with PIRs in their respective cell types (Supplementary Fig. 4b). Notably, OPHN1 stabilizes synaptic AMPA receptors and mediates long-term depression in the hippocampus, and its loss of function has been linked to mental retardation[29]. CHAT is a principal marker for lower motor neuron maturity and function, and TLR4 is a key regulator of immune activation and synaptogenesis in astrocytes[30].

Gene expression is coordinately controlled by transcription factors (TFs) and regulatory elements such as enhancers. Therefore, PIRs provide the means for investigating mechanisms underlying cell-type-specific gene regulation. We used HOMER[31] to evaluate TF motif enrichment at cell-type-specific distal open chromatin peaks in PIRs for each cell type (Fig. 3c and Supplementary Table 4). First, the CTCF motif is highly enriched across all cell types, consistent with its role in mediating looping within TADs[32–35]. Motifs for ASCL1, ISL1, NEUROG2, OLIG2 and ZIC3, TFs linked

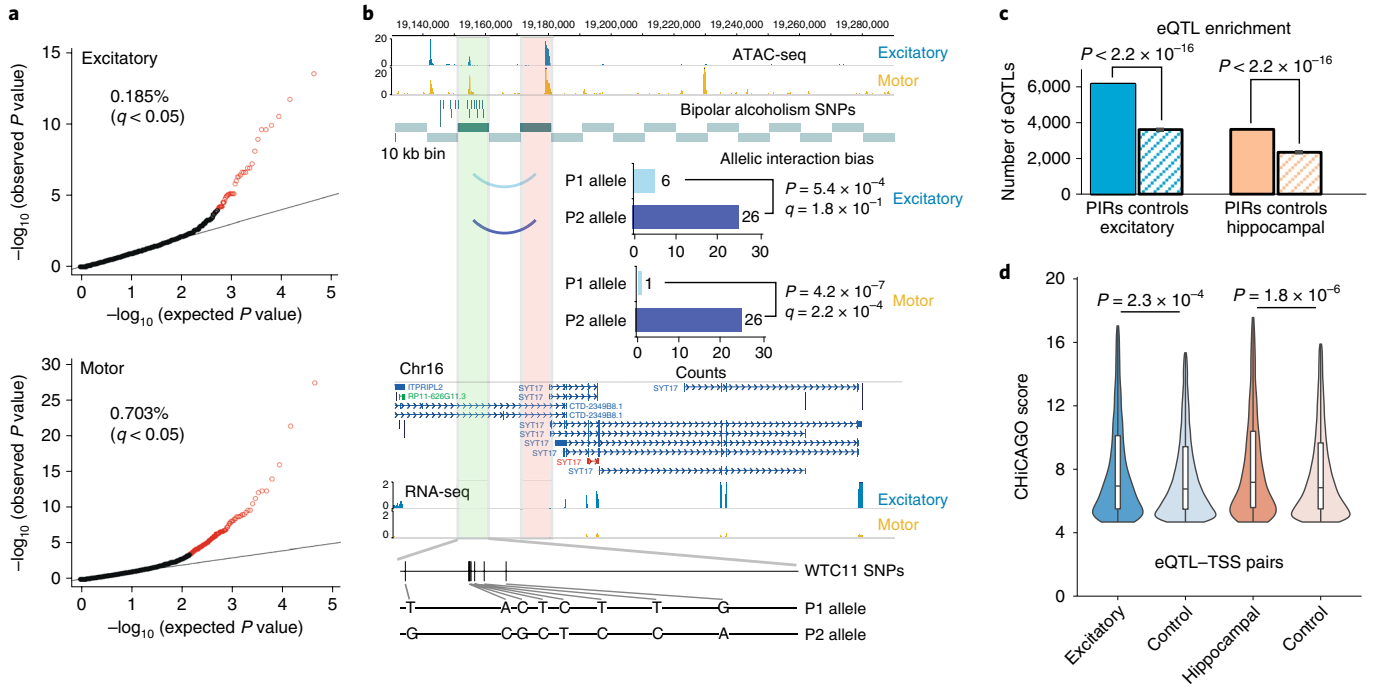


**Fig. 6 | Genetics variants contribute to chromatin interaction bias and alterations in gene expression.** **a**, Quantile-quantile plots showing the proportions of interacting 10-kb bins exhibiting significant allelic bias at an FDR cutoff of 5% (two-tailed binomial test with BH correction) in excitatory neurons ($n = 22{,}162$ bins) and lower motor neurons ($n = 21{,}479$ bins). **b**, A sample interaction with significant allelic bias in excitatory neurons ($P = 5.4 \times 10^{-4}$, two-tailed binomial test) and lower motor neurons ($P = 4.2 \times 10^{-7}$, two-tailed binomial test). The interaction occurs between the *SYT17* promoter and a PIR with bipolar alcoholism SNPs at an open chromatin peak. Heterozygous phased WTC11 variants at the PIR as well as bar graphs of allele-specific read counts are shown. **c**, Enrichment of significant eQTLs from GTEx V7 at significant versus randomly shuffled PIRs in matched tissue types for excitatory and hippocampal DG-like neurons ($P < 2.2 \times 10^{-16}$ for both cell types, two-tailed one-sample *z*-test). Means and the s.e.m. for the number of overlaps across $n = 100$ sets of randomly shuffled PIRs are shown. **d**, Distributions of interaction scores for chromatin interactions overlapping significant versus randomly sampled nonsignificant eQTL–TSS pairs in excitatory and hippocampal DG-like neurons ($P = 2.3 \times 10^{-4}$ for excitatory neurons and $P = 1.8 \times 10^{-6}$ for hippocampal DG-like neurons, two-tailed two-sample Kolmogorov–Smirnov test). Additional details are available in the Methods. Violin plots show the distributions of gene expression values within each group, and boxplots indicate the median, IQR, Q1 − 1.5 × IQR and Q3 + 1.5 × IQR.

to neuronal fate commitment, are also broadly enriched across the neuronal subtypes. Additional TFs that function in brain development include CUX1, CUX2, EBF1, EBF2, LHX1, LHX2, NKX6-1, ONECUT1, RFX1, RFX2, RFX5, TCF4 and TGIF2. The TBR1 motif is highly enriched in hippocampal DG-like neurons, consistent with its expression in the hippocampus[36]. Meanwhile, astrocytes are enriched for motifs in the Fos and Jun families, which contain key regulators for inflammatory and immune pathways. Also enriched in astrocytes are motifs for ATF3 and the RUNX and TEAD families, TFs with established roles in astrocyte differentiation, maturation and proliferation. Motif enrichment is not always accompanied by the expression of the corresponding TFs. This may reflect potential synergistic interactions between the cell types. For example, NRF2 is a key regulator of the oxidative stress response, and its activity has been shown to be repressed in neurons while also inducing a strong response in astrocytes[37]. Therefore, its shared expression may reflect the neuroprotective roles that astrocytes serve for other cell types. Alternatively, TFs do not have to be highly expressed to perform their cellular functions due to additional avenues for regulation at the posttranscriptional and posttranslational levels. Our results support the notion that PIRs contribute to cell fate commitment and are capable of identifying both known and new regulators in a cell-type-specific manner.

**Identification of regulatory targets for in vivo-validated enhancer elements using chromatin interactions.** Regulation of target genes by enhancers is thought to be mediated by physical chromatin looping. Congruent with this concept, chromatin interactions detected by pcHi-C can be used to link enhancers with their target genes. The VISTA Enhancer Browser is a database containing experimentally validated human and mouse noncoding sequences with enhancer activity[11]. To date, it contains 2,956 tested sequences, 1,568 of which exhibit enhancer activity during embryonic development. However, the regulatory targets for these enhancer elements have remained largely uncharacterized. To address this knowledge gap, we provide cell-type-specific annotations of putative target genes for each enhancer element using our chromatin interactions and open chromatin peaks (Supplementary Table 5). Across all cell types, our interactions recover 589 or 37.6% of positively tested enhancer elements with human sequences, 320 of which were further annotated as neural enhancers according to tissue-specific patterns of LacZ staining in mouse embryos (Fig. 4a,b). Of the 589 interacting positive enhancer elements, 306 interact exclusively with 464 more distal genes (scenario I), 118 interact with both their nearest genes and 484 more distal genes (scenario II) and 60 interact exclusively with their nearest genes (scenario III) (Fig. 4c). The remaining 105 elements could not be resolved for interactions with their nearest genes (scenario IV), although they were found to interact with 395 more distal genes. In total, our interactions identify 1,343 new gene targets for positive enhancer elements in the VISTA Enhancer Browser, greatly expanding our knowledge of gene regulatory relationships at these loci.

**Validation of PIRs in human neural cells using CRISPR techniques.** We validated two PIRs physically interacting up to 40 kb away with the *CDK5RAP3* promoter (regions 1 and 2) (Fig. 4d). CDK5RAP3 regulates CDK5, which functions in neuronal development[38] and regulates proliferation in nonneuronal cells[39]. Notably, both PIRs overlap open chromatin peaks as well as enhancers annotated with forebrain activity in the VISTA Enhancer Browser (mm876 and mm999 for region 1 and mm1299 for region 2) (Fig. 4e). We targeted both regions for CRISPR deletion in the i$^3$N iPSCs, followed by differentiation of the cells into excitatory neurons and quantification of any changes in gene expression by quantitative PCR (qPCR). Deleting the 2 kb open chromatin peak in region 1 led to a significant downregulation of *CDK5RAP3* expression across three independent clones ($P = 1.6 \times 10^{-2}$, two-tailed two-sample *t*-test) (Supplementary Fig. 4c). However, upon trying to delete the open chromatin peak in region 2, we observed massive cell death of iPSCs immediately following the introduction of the Cas9–single guide RNA (sgRNA) protein complex. We picked 48 individual clones from cells surviving the transfection but failed to isolate any clones with deletions, suggesting that this locus is essential for maintaining *CDK5RAP3* expression and survival in iPSCs. To circumvent this lethal phenotype for iPSCs with region 2, we used CRISPR interference (CRISPRi) to silence both regions 1 and 2 in excitatory neurons. We also silenced a third region that interacts with the *CDKRAP3* promoter in the other cell types, but not in excitatory neurons (region 3). We show that silencing of regions 1 and 2, but not region 3, leads to significant downregulation of *CDK5RAP3* expression without influencing the expression of nearby genes ($P = 9.1 \times 10^{-4}$ for region 1 and $P = 2.2 \times 10^{-3}$ for region 2, two-tailed two-sample *t*-test) (Fig. 4f,g). Interestingly, an enhancer with spinal cord activity proximal to region 3 (mm1675) interacts with the *CDK5RAP3* promoter in lower motor neurons and astrocytes, but not in the other cell types (Fig. 4d, e). Overall, these results show that chromatin interactions recapitulate cell-type-specific patterns of enhancer activity, underscoring the importance of studying epigenomic regulation in the appropriate cell types.

**Cell-type-specific enrichment and regulatory target identification for neuropsychiatric disorder risk variants at PIRs.** Previous large-scale epigenomic studies of human tissues and cell lines highlighted the importance of disease-associated variants at distal regulatory regions[25] and the need for high-throughput approaches to prioritize variants for further validation. Therefore, we used our chromatin interactions to annotate complex neuropsychiatric disorder- or trait-associated variants from the GWAS Catalog[40]. We mined a total of 6,396 unique GWAS SNPs at a significance threshold of $10^{-6}$ for 11 traits including Alzheimer's disease (AD), attention deficit hyperactivity disorder (ADHD), autism spectrum disorder (ASD), amyotrophic lateral sclerosis (ALS), bipolar disorder (BD), epilepsy (EP), frontotemporal dementia (FTD), mental process (MP), Parkinson's disease (PD), schizophrenia (SCZ) and unipolar depression (UD). We identified linked SNPs at a linkage disequilibrium (LD) threshold of 0.8 using HaploReg[41] for a total of 95,954 unique SNPs across all traits (Supplementary Table 6). We find that SNPs are enriched at PIRs in a disease- and cell-type-specific manner (Fig. 5a), with ASD, MP and SCZ SNPs enriched at PIRs across all cell types. UD SNPs are enriched exclusively in excitatory and hippocampal DG-like neurons, whereas AD, ADHD and BD SNPs also exhibit enrichment in lower motor neurons. ALS SNPs are enriched in the neuronal subtypes but not in astrocytes, consistent with the characterization of ALS as a motor neuron disease and reinforcing evidence for its role in hippocampal degeneration[42]. Interestingly, PD SNPs are enriched in astrocytes but not in the other cell types. This enrichment of PD SNPs at astrocyte-specific PIRs supports the theory that astrocytes play an initiating role in PD, based on evidence that numerous genes implicated in PD have functions unique to astrocyte biology, as well as the neuroprotective roles astrocytes provide for dopaminergic neurons in the substantia nigra[43]. EP and FTD SNPs are not enriched in any of the cell types, indicating their potential functions in alternative cell types, insufficient study power or mechanisms acting outside of chromatin-mediated gene regulation.

Up to 70% of GWAS SNPs have at least one linked SNP overlapping PIRs in one or more cell type (Fig. 5b). As it is common practice to assign GWAS SNPs to their nearest genes, we counted the number of GWAS SNPs with at least one linked SNP interacting with their nearest gene across all diseases. We found that 1,365 GWAS SNPs interact exclusively with 3,361 more distal genes (scenario I), 1,243 GWAS SNPs interact with both their nearest genes and 12,070 more distal genes (scenario II) and 248 GWAS SNPs interact exclusively with their nearest genes (scenario III) (Fig. 5c and Supplementary Fig. 5a). In total, 16,471 non-neighboring gene targets are identified across all diseases (Supplementary Table 7). To prioritize variants potentially disrupting regulatory interactions, we focused on SNPs overlapping open chromatin peaks at PIRs, and find that these putative regulatory SNPs interact with genes that are relevant in the context of their respective disease etiologies (Supplementary Table 8). GO enrichment analysis for genes targeted by AD SNPs yields terms associated with amyloid-beta formation, interferon-beta production and cranial nerve development (Supplementary Fig. 6 and Supplementary Table 9). Meanwhile, genes targeted by ASD, BD, SCZ and UD SNPs are enriched for epigenetic terms including chromatin assembly, nucleosome assembly and nucleosome organization. For genes targeted by GWAS SNPs in the other diseases, enriched terms include neuronal processes such as myelin maintenance, neuron projection extension, synapse assembly, synaptic transmission and nervous system development.

Notably, a previously reported interaction between the *FOXG1* promoter and a PIR with SCZ SNPs over 700 kb away is recapitulated by our data[6] (Supplementary Fig. 5b). At a different locus, an astrocyte-specific PIR with AD SNPs targets the promoter of *CASP2*, which encodes a well-known mediator of apoptosis that is linked to neurodegeneration[44,45] (Supplementary Fig. 7a). Hippocampal DG-like neuron-specific PIRs with ASD SNPs target the promoter of *BCAS2*, whose knockdown in mice leads to microcephaly-like phenotypes with reduced learning, memory and DG volume[46] (Supplementary Fig. 7c). Finally, the *MSI2* promoter is targeted by an astrocyte-specific PIR with SCZ SNPs and by PIRs with BD SNPs in hippocampal DG-like neurons, lower motor neurons and astrocytes (Supplementary Fig. 7d). In summary, we demonstrate that an approach leveraging epigenomic data to jointly prioritize and map regulatory targets for variants enables the identification of putative processes that are disrupted in disease and development.

**Validation of PIRs containing neuropsychiatric disorder risk variants.** PIRs with MP SNPs in an intron for *PTPRO* interact over 300 kb away with the promoter of *STRAP* (Fig. 5d), which encodes a component of the survival of motor neuron (SMN) complex[47]. The complex itself facilitates spliceosome assembly and is associated with spinal muscular atrophy[48]. To validate this locus, we derived three independent i$^3$N iPSC clones containing biallelic deletions for a PIR in this region and observed significant downregulation of *STRAP* expression following differentiation of the cells into excitatory neurons ($P = 3.4 \times 10^{-4}$, two-tailed two-sample *t*-test). Targeting the same PIR with CRISPRi also consistently downregulated *STRAP* expression in excitatory neurons ($P = 7.4 \times 10^{-3}$, two-tailed two-sample *t*-test) (Supplementary Fig. 5c). Next, we focused on a PIR 20 kb upstream from the promoter of *DRD2*, which encodes the D2 subtype of the dopamine receptor. Previously, rs2514218:C > T, a noncoding variant 47 kb upstream from *DRD2*, was found to be associated with antipsychotic drug response in a cohort of patients with SBS[49]. This variant is in LD

with a cluster of SCZ SNPs overlapping open chromatin peaks in the PIR for *DRD2*. *DRD2* is also the gene associated with the Taq1A polymorphism, which has been linked to reduced dopamine receptor density as well as addiction, anxiety, depression and social problems in patients[50]. We first demonstrate that monoallelic deletion of this PIR in three independent clones leads to a significant downregulation of *DRD2* expression in excitatory neurons ($P = 6.2 \times 10^{-3}$, two-tailed two-sample *t*-test) (Fig. 5e). Next, through TOPO cloning and genotyping cDNA with allele-specific variants, we confirm that monoallelic deletion of the same PIR leads to allelic imbalance in *DRD2* expression (Supplementary Fig. 5d). By prioritizing and validating PIRs containing putative regulatory SNPs for key genes such as *DRD2*, our approach moves the field a step closer to the development of therapeutic and diagnostic strategies targeting specific risk variants in otherwise recalcitrant complex neuropsychiatric disorders.

**Genetic variants contribute to chromatin interaction bias and alterations in gene expression.** Since regulatory variants and other genetic perturbations are thought to introduce or disrupt chromatin loops between promoters and PIRs, we were interested to see if we could detect instances of allelic bias across our sets of significant promoter-PIR interactions. We used our chromatin interaction data to perform genome-wide phasing of WTC11 variants using HaploSeq[51] and performed allele-specific mapping at a resolution of 10 kb using HiC-Pro[52]. We identified 41 (0.185%) and 151 (0.703%) significantly interacting bins to exhibit allelic bias at a false discovery rate (FDR) cutoff of 5% (two-tailed binomial test with Benjamini–Hochberg (BH) correction) in the excitatory and lower motor neurons, respectively (Fig. 6a and Supplementary Table 10). In one case, allelically biased interactions were detected between a PIR containing bipolar alcoholism SNPs[53] and the promoter of *SYT17*, which encodes a member of a family of membrane-trafficking proteins that mediate synaptic function and calcium-controlled neurotransmitter release[54]. The risk allele of the lead variant (rs8062326:G > A) is associated with the WTC11 allele exhibiting reduced interaction frequency in both cell types (Fig. 6b), suggesting that regulatory variants can increase individual risk for bipolar alcoholism by disrupting interactions for *SYT17*.

Physical chromatin interactions have been theorized to mediate the effects of *cis*-acting regulatory variants, including expression quantitative trait loci (eQTLs), on gene expression. In support of this hypothesis, we first demonstrate that significant eQTLs in cortical and hippocampal tissues from GTEx V7 (ref. [55]) are enriched at PIRs for excitatory and hippocampal DG-like neurons, respectively ($P < 2.2 \times 10^{-16}$ for both cell types, two-tailed one-sample *z*-test) (Fig. 6c). Next, we show that scores for interactions overlapping significant eQTL–TSS pairs are significantly higher than scores for interactions overlapping randomly shuffled eQTL–TSS pairs ($P = 2.28 \times 10^{-4}$ for excitatory neurons and $P = 1.76 \times 10^{-6}$ for hippocampal DG-like neurons, two-tailed two-sample Kolmogorov–Smirnov test) (Fig. 6d). This indicates that chromatin interactions recapitulating regulatory relationships between significant eQTL–TSS pairs are identified by pcHi-C with increased levels of confidence. Overall, our results present orthogonal lines of evidence that chromatin interactions can not only be altered by variants in an allele-specific manner, but that variants can also modulate gene expression through the formation or disruption of regulatory chromatin loops.

## Discussion

There is a distinct lack of 3D epigenomic annotations in cell types that are relevant to disease and development, especially in the field of brain research. Past studies have relied on heterogeneous tissues comprised of cell types with disparate functions, limiting the ability to detect and interpret instances of cell-type-specific gene regulation. Neurons and glia, for example, represent lineages with divergent functions that coexist in most tissues of the CNS. At the same time, complex diseases often involve multiple dysregulated loci with cell-type-specific patterns of activity. This presents unique challenges for deciphering disease etiology, for example, in attempting to distinguish causative mechanisms from secondary reactive phenotypes when multiple cell types are involved. For these reasons, the comprehensive annotation of regulatory relationships in specific well-characterized cell populations should enable the derivation of deeper insights into complex disease biology. Chromatin interactions, in particular, are ideal for mapping promoters to distal regulatory elements, as they provide direct evidence of regulatory sequences physically contacting loci of interest. To date, several studies have characterized chromatin interactions in fetal brain tissues and cultured neural cells[6,56]. However, these studies relied on in situ Hi-C for their interaction calls, which lacks power compared to targeted approaches such as pcHi-C.

Here, we have leveraged pcHi-C, ATAC-seq and RNA-seq to comprehensively annotate previously uncharted regulatory relationships between promoters and distal regulatory elements in cell types that are relevant to complex neuropsychiatric disorders. We show that PIRs are not only cell-type specific but also enriched for regulatory chromatin signatures including open chromatin peaks and in vivo-validated enhancer elements from the VISTA Enhancer Browser. Inspection of cell-type-specific distal open chromatin peaks at PIRs reveals subtype-specific binding sites for TFs involved in the specification and maintenance of cellular identity. Furthermore, our interactions identify new gene targets for disease-associated variants and enable the prioritization of variants for validation using CRISPR techniques. We report a large number of putative regulatory variants that may provide additional insights into aspects of complex disease biology. Finally, the disease- and cell-type-specific enrichment of variants at PIRs, combined with the observation that the same PIRs can target different genes in different cell types, supports existing evidence that regulatory variants possess context-dependent functional specificities.

The integrative analysis in this study has several limitations including a lack of cell-type-specific annotations for genomic and epigenomic features occurring at PIRs. For example, the analysis of chromatin state and eQTL enrichment at PIRs used data in matched tissues from the Roadmap Epigenomics Project and GTEx V7, respectively. Furthermore, while studying chromatin interactions in healthy cells enables the detection of regulatory interactions in the absence of dysregulation, the epigenomic characterization of patient-derived cells will be important to glean specific insights into how the 3D epigenome is altered in disease. Additional experiments are necessary to determine how the haploinsufficiency of proteins such as STRAP and DRD2 may contribute to phenotypes in disease. Finally, in vitro cultured cells can at present only approximate the full set of cellular responses occurring in vivo, especially for complex structures such as the brain, and they may reflect different developmental stages than expected based on their time in culture. Future approaches isolating specific cell types from tissues, using single-cell sequencing, or using advanced organoid models will be essential for drilling down more deeply into mechanisms driving cellular identity and disease. The epigenomic characterization of additional cell types should continue to yield rich insights into the landscape of transcriptional regulation, contributing to an improved understanding of complex disease biology[57].

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at https://doi.org/10.1038/s41588-019-0472-1.

Received: 4 May 2018; Accepted: 21 June 2019;
Published online: 31 July 2019

## References


1. Mumbach, M. R. et al. Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. *Nat. Genet.* **49**, 1602–1612 (2017).
2. Zhu, Z. et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481–487 (2016).
3. Claussnitzer, M. et al. FTO obesity variant circuitry and adipocyte browning in humans. *N. Engl. J. Med.* **373**, 895–907 (2015).
4. Smemo, S. et al. Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature* **507**, 371–375 (2014).
5. de la Torre-Ubieta, L. et al. The dynamic landscape of open chromatin during human cortical neurogenesis. *Cell* **172**, 289–304 e218 (2018).
6. Won, H. et al. Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature* **538**, 523–527 (2016).
7. Miyaoka, Y. et al. Isolation of single-base genome-edited human iPS cells without antibiotic selection. *Nat. Methods* **11**, 291–293 (2014).
8. Wang, C. et al. Scalable production of iPSC-derived human neurons to identify Tau-lowering compounds by high-content screening. *Stem Cell Rep.* **9**, 1221–1233 (2017).
9. Mertens, J. et al. Differential responses to lithium in hyperexcitable neurons from patients with bipolar disorder. *Nature* **527**, 95–99 (2015).
10. Fernandopulle, M. S. et al. Transcription Factor-Mediated Differentiation of Human iPSCs into Neurons. *Curr. Protoc. Cell Biol.* **79**, e51 (2018).
11. Visel, A., Minovitsky, S., Dubchak, I. & Pennacchio, L. A. VISTA Enhancer Browser–a database of tissue-specific human enhancers. *Nucleic Acids Res.* **35**, D88–92 (2007).
12. Zhang, Y. et al. Rapid single-step induction of functional neurons from human pluripotent stem cells. *Neuron* **78**, 785–798 (2013).
13. Yu, D. X. et al. Modeling hippocampal neurogenesis using human pluripotent stem cells. *Stem Cell Rep.* **2**, 295–310 (2014).
14. Zhang, Y. et al. Purification and characterization of progenitor and mature human astrocytes reveals transcriptional and functional differences with mouse. *Neuron* **89**, 37–53 (2016).
15. Jung, I. et al. A compendium of promoter-centered long-range chromatin interactions in the human genome. *Nat. Genet.* (in the press).
16. Yang, T. et al. HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. *Genome Res.* **27**, 1939–1949 (2017).
17. Cairns, J. et al. CHiCAGO: robust detection of DNA looping interactions in Capture Hi-C data. *Genome Biol.* **17**, 127 (2016).
18. Jackson, D. A., Hassan, A. B., Errington, R. J. & Cook, P. R. Visualization of focal sites of transcription within human nuclei. *EMBO J.* **12**, 1059–1065 (1993).
19. Zhang, Y. et al. Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations. *Nature* **504**, 306–310 (2013).
20. Diao, Y. et al. A tiling-deletion-based genetic screen for cis-regulatory element identification in mammalian cells. *Nat. Methods* **14**, 629–635 (2017).
21. Engreitz, J. M. et al. Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature* **539**, 452–455 (2016).
22. Consortium, E. P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
23. Shen, Y. et al. A map of the cis-regulatory sequences in the mouse genome. *Nature* **488**, 116–120 (2012).
24. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* **9**, 215–216 (2012).
25. Roadmap Epigenomics, C. et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
26. Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife* **6**, e21856 (2017).
27. Davis, C. A. et al. The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res.* **46**, D794–D801 (2018).
28. Kuleshov, M. V. et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–97 (2016).
29. Nadif Kasri, N., Nakano-Kobayashi, A. & Van Aelst, L. Rapid synthesis of the X-linked mental retardation protein OPHN1 mediates mGluR-dependent LTD through interaction with the endocytic machinery. *Neuron* **72**, 300–315 (2011).
30. Shen, Y. et al. Postnatal activation of TLR4 in astrocytes promotes excitatory synaptogenesis in hippocampal neurons. *J. Cell Biol.* **215**, 719–734 (2016).
31. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
32. Ren, G. et al. CTCF-mediated enhancer-promoter interaction is a critical regulator of cell-to-cell variation of gene expression. *Mol. Cell* **67**, 1049–1058 e1046 (2017).
33. Guo, Y. et al. CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. *Cell* **162**, 900–910 (2015).
34. Handoko, L. et al. CTCF-mediated functional chromatin interactome in pluripotent cells. *Nat. Genet.* **43**, 630–638 (2011).
35. Hou, C., Dale, R. & Dean, A. Cell type specificity of chromatin organization mediated by CTCF and cohesin. *Proc. Natl Acad. Sci. USA* **107**, 3651–3656 (2010).
36. Chuang, H. C., Huang, T. N. & Hsueh, Y. P. Neuronal excitation upregulates Tbr1, a high-confidence risk gene of autism, mediating Grin2b expression in the adult brain. *Front Cell Neurosci.* **8**, 280 (2014).
37. Liddell, J. R. Are astrocytes the predominant cell type for activation of Nrf2 in aging and neurodegeneration? *Antioxid.* **6**, E65 (2017).
38. Yin, X., Warner, D. R., Roberts, E. A., Pisano, M. M. & Greene, R. M. Novel interaction between nuclear co-activator CBP and the CDK5 activator binding protein - C53. *Int. J. Mol. Med.* **16**, 251–256 (2005).
39. Xie, Y. H. et al. Cloning and characterization of human IC53-2, a novel CDK5 activator binding protein. *Cell Res.* **13**, 83–91 (2003).
40. Buniello, A. et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005–D1012 (2019).
41. Ward, L. D. & Kellis, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* **40**, D930–934 (2012).
42. Abdulla, S. et al. Hippocampal degeneration in patients with amyotrophic lateral sclerosis. *Neurobiol. Aging* **35**, 2639–2645 (2014).
43. Booth, H. D. E., Hirst, W. D. & Wade-Martins, R. The role of astrocyte dysfunction in Parkinson's disease pathogenesis. *Trends Neurosci.* **40**, 358–370 (2017).
44. Tiwari, M., Lopez-Cruzan, M., Morgan, W. W. & Herman, B. Loss of caspase-2-dependent apoptosis induces autophagy after mitochondrial oxidative stress in primary cultures of young adult cortical neurons. *J. Biol. Chem.* **286**, 8493–8506 (2011).
45. Zhao, X. et al. Caspase-2 cleavage of tau reversibly impairs memory. *Nat. Med.* **22**, 1268–1276 (2016).
46. Huang, C. W. et al. Conditional knockout of breast carcinoma amplified sequence 2 (BCAS2) in mouse forebrain causes dendritic malformation via beta-catenin. *Sci. Rep.* **6**, 34927 (2016).
47. Grimmler, M. et al. Unrip, a factor implicated in cap-independent translation, associates with the cytosolic SMN complex and influences its intracellular localization. *Hum. Mol. Genet* **14**, 3099–3111 (2005).
48. Burghes, A. H. & Beattie, C. E. Spinal muscular atrophy: why do low levels of survival motor neuron protein make motor neurons sick? *Nat. Rev. Neurosci.* **10**, 597–609 (2009).
49. Zhang, J. P. et al. Association of a schizophrenia risk variant at the DRD2 locus with antipsychotic treatment response in first-episode psychosis. *Schizophr. Bull.* **41**, 1248–1255 (2015).
50. Eisenstein, S. A. et al. Prediction of striatal D2 receptor binding by DRD2/ANKK1 TaqIA allele status. *Synapse* **70**, 418–431 (2016).
51. Selvaraj, S., R Dixon, J., Bansal, V. & Ren, B. Whole-genome haplotype reconstruction using proximity-ligation and shotgun sequencing. *Nat. Biotechnol.* **31**, 1111–1118 (2013).
52. Servant, N. et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259 (2015).
53. Lydall, G. J. et al. Confirmation of prior evidence of genetic susceptibility to alcoholism in a genome-wide association study of comorbid alcoholism and bipolar disorder. *Psychiatr. Genet.* **21**, 294–306 (2011).
54. Sudhof, T. C. Calcium control of neurotransmitter release. *Cold Spring Harb. Perspect. Biol.* **4**, a011353 (2012).
55. Consortium, G. T. et al. Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
56. Rajarajan, P. et al. Neuron-specific signatures in the chromosomal connectome associated with schizophrenia risk. *Science* **362**, eaat4311 (2018).
57. Straws in a haystack. *Nat. Genet.* **50**, 631 (2018).


## Acknowledgements

We thank A. Schmitt and B. Ren (Ludwig Institute for Cancer Research, University of California, San Diego) for sharing pcHi-C probes and the pcHi-C protocol. Genomic analysis of the WTC11 line in this study was made possible by the whole-genome sequencing data generated by the Allen Institute for Cell Science. We thank the Institute and its founder P.G. Allen for making this work possible. We thank G. Hon (University of Texas Southwestern Medical Center) and S. Henikoff (Fred Hutchinson Cancer Research Center, Howard Hughes Medical Institute) for providing reagents. We acknowledge the ENCODE Consortium and Bradley Bernstein's laboratory for generating the ChIP-seq data for astrocytes used in this study. We thank N. Ahituv, Y. Guo, R.D. Hawkins, M. McManus and B. Ren for providing critical feedback on the manuscript. We thank Y. Qu for her contributions to the illustrations for Fig. 1. This work was made possible in part by the Core Grant for Vision Research and the Research to Prevent Blindness Unrestricted Grant no. NIH-NEI P30EY002162. This work was supported by the National Institutes of Health (NIH) grant nos. R01AG057497 (to Y.S., L.G. and H.S.) and R01EY027789 and UM1HG009402 (to Y.S.), the UCSF Weill Institute

for Neuroscience Innovation Award, the Hillblom Foundation, and the American Federation for Aging Research New Investigator Award in Alzheimer's Disease to Y.S., NIH grant nos. R01EY028249, R01HL130533, R01-HL135358 (to B.R.C.), P01NS097206 and U19MH106434 (to H.S.) and R01MH105128, R35NS097370, and U19AI131130 (to G.L.M.). M.S. is supported by NIH grant no. T32GM007175. F.J. is supported by NIH grant no. T32GM007309.

## Author contributions

M.S. and Y.S. designed the study. M.S., X.Y., X.R., L.M., I.J., T.W.T. and K.J. performed the experiments. M.S., B.L., I.J., M.T., L.W. and Y.L. performed data analysis. J.D. contributed to genomic phasing using HaploSeq. S.L., J.Y., K.W., B.R.C., F.J., G.L.M., H.S., L.D., C.W. and L.G. provided cells. M.S. and Y.S. prepared the manuscript with assistance from all authors.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at https://doi.org/10.1038/s41588-019-0472-1.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Correspondence and requests for materials** should be addressed to Y.S.

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

## Methods

**Cell culture.** Human excitatory neurons were generated using integrated, isogenic and inducible neurogenin-2 (Ngn2) iPSCs ($i^3$N iPSCs) with doxycycline-inducible mouse Ngn2 integrated at the AAVS1 safe-harbor locus. The $i^3$N iPSCs have a well-characterized wild type genetic background (WTC11)[7]. A simplified, two-step predifferentiation and maturation protocol was used to generate the excitatory neurons[6]. Briefly, $i^3$N iPSCs were incubated with 2 μg ml$^{-1}$ doxycycline in predifferentiation media containing knockout DMEM/F12 supplemented with 1× N-2, 1× NEAA, 1 μg ml$^{-1}$ mouse laminin, 10 ng ml$^{-1}$ brain-derived neurotrophic factor (BDNF) and 10 ng ml$^{-1}$ NT3. ROCK inhibitor(10 μM) was included in the predifferentiation media for the first day. Media were changed daily for 3 d. For maturation, predifferentiated precursor cells were dissociated and subplated on poly-D-lysine and laminin plates in maturation media containing equal parts DMEM/F12 and Neurobasal-A with 2 μg ml$^{-1}$ doxycycline and supplemented with 0.5× B-27, 0.5× N-2, 1× NEAA, 0.5× GlutaMax, 1 μg ml$^{-1}$ mouse laminin, 10 ng ml$^{-1}$ BDNF and 10 ng ml$^{-1}$ NT3. The doxycycline was omitted from all subsequent media changes. Half of the media was changed weekly for the first 2 weeks, then the amount of media was doubled on day 21. Thereafter, a third of the media was replaced weekly until harvesting and 7 to 8-week-old excitatory neurons were used for library preparation.

Human hippocampal DG-like neurons were generated from dissociated hippocampal organoids (unpublished). Briefly, WTC11 iPSCs were grown on MEF feeder cells and patterned toward a neural ectoderm fate using dual SMAD inhibition as floating embryoid bodies (EBs) in medium containing 20% knockout serum replacement. Four-week-old EBs were patterned toward a hippocampal fate using WNT and BMP in medium containing 1× N-2. After patterning, organoids were dissociated using a neural tissue dissociation kit (MiltenyiBiotech), plated on PDL- and laminin-coated plates, and cultured for 4 weeks in media containing 1× B-27, 10 ng ml$^{-1}$ BDNF, 10 ng ml$^{-1}$ GDNF, 0.5 mM cAMP and 200 μM ascorbic acid.

Human lower motor neurons were differentiated from WTC11 iPSCs using a doxycycline-inducible transgene expressing NGN2, ISL1 and LHX3 integrated at the AAVS1 safe-harbor locus ($i^3$LMN iPSCs)[10]. Briefly, $i^3$LMN iPSCs were maintained on growth factor reduced Matrigel in StemFit media (Nacalai). On day 0, $1.5 \times 10^6$ $i^3$LMN iPSCs were plated on 10-cm dishes, followed 24 h later by exchange into neural induction media containing doxycycline and compound E. On day 3, the precursor cells were transferred to 12-well plates coated with poly-D-lysine and laminin at a density of $2.5 \times 10^5$ cells per well. From day 3 to 4, the cells were treated with a pulse of 40 μM BrdU for 24 h to suppress the proliferation of undifferentiated cells. Media were exchanged on day 4 and every 3 d thereafter. The cells were harvested 10 d postdifferentiation for library preparation.

Human primary astrocytes (P0) were purchased from ScienCell Research Laboratories (catalog no. 1800) and cultured using the recommended media (catalog no. 1801). Cells were cultured in flasks coated with poly-L-lysine (2 μg cm$^2$) and passaged once using trypsin and EDTA before harvesting.

All cells used in the present study were verified as being free from mycoplasma contamination.

**Promoter capture Hi-C (pcHi-C).** In situ Hi-C libraries for excitatory neurons, hippocampal DG-like neurons, lower motor neurons and astrocytes were constructed from 1 to $2 \times 10^6$ cells (fixed in 1% PFA) using HindIII as a restriction enzyme as previously described[58]. pcHi-C was performed using biotinylated RNA probes according to an established protocol[15]. Briefly, sets of 120-base pair (bp) probes with 30-bp overhangs were designed to capture all promoter-containing and adjacent HindIII fragments. Three probes were targeted to each side of a restriction site for a total of 12 probes targeting each promoter-containing HindIII fragment. Promoters (defined as the sequences up to 500 bp upstream and downstream of each TSS) for 19,603 of the 20,332 protein coding genes in GENCODE 19 were captured using this approach. While noncoding RNA promoters were not explicitly targeted by this design, HindIII fragments containing 3,311 of the 14,069 noncoding RNA promoters in GENCODE 19 were also captured by the probes.

To perform the hybridization, 500 ng of each situ Hi-C library was first mixed with 2.5 μg human Cot-1 DNA (Invitrogen no. 15279011), 2.5 μg salmon sperm DNA (Invitrogen no. 15632011) and 0.5 nmol each of the p5 and p7 IDT xGen Universal Blocking Oligos in a total volume of 10 μl and then denatured for 5 min at 95 °C and prewarmed at 65 °C. Next, a hybridization buffer mix was prepared by combining 25 μl 20× SSPE, 1 μl 0.5 M EDTA, 10 μl 50× Denhardt's solution and 13 μl 1% SDS and prewarming the mix to 65 °C. Finally, 500 ng of the probes was mixed with 1 μl 20 U μl$^{-1}$ SUPERase-In (Invitrogen no. AM2696) in a total volume of 6 μl, prewarmed to 65 °C, and combined with the library and hybridization buffer mixes. The final solution was transferred to a humidified hybridization chamber and incubated for 24 h at 65 °C. Dynabeads MyOne Streptavidin T1 magnetic beads (0.5 mg, Invitrogen no. 65601) were used to pull down the captured fragments in a binding buffer consisting of 10 mM Tris–HCl pH 7.5, 1 M NaCl and 1 mM EDTA. The beads were washed once with 1× SSC and 0.1% SDS for 30 min at 25 °C, followed by three washes with prewarmed 0.1× SSC and 0.1% SDS for 10 min at 65 °C. The final library was eluted in 20 μl nuclease-free water, amplified, then sent for paired-end sequencing on the HiSeq 4000 (50-bp reads), the HiSeq X Ten (150-bp reads), or the NovaSeq 6000 (150-bp reads).

**Calling significant promoter-PIR interactions.** Paired-end sequencing reads were first trimmed using fastp v.0.20.0 (ref. [59]) running the default settings before being mapped, filtered and deduplicated using HiCUP v.0.71 (ref. [60]) with bowtie2 v.2.3.2 (ref. [61]) and filtering for ditags between 100 and 1,200 bp. In addition, the sequencing depth of all libraries was normalized so that each replicate had the same number of usable reads (defined as the number of on-target *cis*-pairs interacting over a distance of 10 kb). Significant promoter-PIR interactions were called using CHiCAGO v.1.1.8 (ref. [17]) running the default settings and retaining baited fragments that are supported by at least 250 reads (minNPerBaits = 250). Interactions between HindIII fragments with a score (defined as the negative log *P* value) of 5 or greater in each cell type were determined to be significant. All data processing metrics are reported in Supplementary Table 1. In cases where CHiCAGO reported the same interaction in different orientations, the two interactions were merged, retaining the higher score of the two interactions. Interchromosomal interactions were omitted from the analysis. To call overlaps between our sets of significant interactions and genomic and epigenomic features including promoters, open chromatin peaks, chromatin states, disease-associated variants and eQTLs, interacting bins were expanded to a minimum width of 5 kb or retained as the original widths of the HindIII fragments if they exceeded 5 kb. Interactions overlapping HindIII fragments larger than 100 kb were omitted from our analysis. An interaction was considered to be shared between cell types if both of its interacting ends intersected the corresponding ends of an interaction in another cell type. Otherwise, an interaction was determined to be cell-type-specific.

**Validation of PIRs using CRISPR deletion.** To validate genomic interactions captured by pcHi-C, candidate PIRs were targeted for CRISPR-mediated deletion in the $i^3$N iPSCs. Pairs of sgRNAs targeting the putative regulatory element as localized by open chromatin peaks in the candidate PIR were designed for each locus of interest. All sgRNAs were synthesized by Synthego. Cas9 protein was sourced from QB3-Berkeley. To generate deletion lines, CRISPR/Cas9 nucleofections were performed using the LONZA Human Stem Cell Nucleofector Kit. For each nucleofection, approximately 500,000 $i^3$N iPSCs were transfected with Cas9:sgRNA RNP complex (consisting of 12 μg Cas9, 10 μg sgRNA 1 and 10 μg sgRNA 2) using program 'A-023' on the LONZA 4D-Nucleofector. The nucleofected cells were then seeded onto Matrigel-coated six-well plates containing Essential 8 Medium (ThermoFisher no. A15169-01) with Y-27632 added for recovery following nucleofection. After 48 h, the cells were split into new six-well plates at a concentration of approximately 50 cells per well for picking single colonies. Clones picked from the six-well plates containing homozygous deletions were confirmed by qPCR and induced into excitatory neurons for quantifying the expression of genes targeted by the deleted PIRs. For each experiment, we used three deletion clones and two wild type clones. Total RNA from the excitatory neurons was extracted using a Qiagen AllPrep DNA/RNA Mini Kit, and cDNA was synthesized using a Bio-RAD iScript cDNA Synthesis Kit. qPCR for targeted genes was performed with FastStart Essential DNA Green Master reaction mix (Roche) on the LightCycler 96 System (Roche). The mean values from three technical replicates were used for statistical testing. Detailed information on all the primers used is available in Supplementary Table 11.

**Validation of PIRs using CRISPRi.** Excitatory neurons induced from $i^3$N iPSCs were infected with lentivirus carrying dCas9-KRAB-blast (Addgene no. 89567) and colonies with high expression of dCas9 were picked. The CROP-seq-opti vector (Addgene no. 106280) was used for sgRNA expression. sgRNAs were cotransfected with lentivirus packaging plasmids pMD2.G (Addgene no. 12259) and psPAX (Addgene no. 12260) into 293 T cells with PolyJet (SignaGen Laboratories no. SL100688) according to the manufacturer's instructions. Virus-containing media was collected for 72 h, filtered through a 0.45 μm filter (Millipore no. SLHV033RS), and concentrated with an Amicon Ultra centrifugal filter (Millipore no. UFC801024). The virus was titrated in the excitatory neurons by qPCR 72 h postinfection. The internal qPCR control targeted on the intronic region (forward primer: TCCTCCGGAGTTATTCTTGGCA and reverse primer: CCCCCCATCTGATCTGTTTCAC). Integration of the WPRE fragment was quantified in comparison with a cell line containing a known copy number of WPRE. For CRISPRi silencing of putative regulatory elements, excitatory neurons were treated with lentivirus expressing sgRNAs (multiplicity of infection ~3) for two or three replicates per condition representing independent differentiation events. Two independent sets of sgRNAs were used for each replicate. Cells were collected for mRNA extraction 7 d posttransfection, and gene expression was determined using qPCR. The mean values from three technical replicates were used for statistical testing. Detailed information on all the primers used is available in Supplementary Table 11.

**Motif enrichment analysis.** We took the sets of all cell-type-specific distal open chromatin peaks participating in significant promoter-PIR interactions between promoter-containing and non-promoter-containing bins for each cell type and used the sequences in 250 bp windows around the peak summits to perform motif enrichment analysis using HOMER 4.10 (ref. [31]) running the default settings. The cumulative binomial distribution was used for motif scoring. The entire genome

was used as a background. Significance and expression values for each detected motif and its corresponding TFs are reported in Supplementary Table 3.

**VISTA enhancer analysis and target gene identification.** Human and mouse enhancer regions with orthologous human sequences and positive annotations in the VISTA Enhancer Browser[11] were downloaded and analyzed for overlap with our sets of significant promoter-PIR interactions for each cell type. Of the 2,956 tested elements in their database (January 2019), 1,568 were positive elements with orthologous human sequences (976 human elements and 892 mouse elements with orthologous human sequences). Positive elements expanded to a minimum width of 5 kb and participating in significant interactions are reported in Supplementary Table 5. The 'hs' or 'mm' prefixes for elements indicate the species of DNA origin (human or mouse). For determining whether positive elements interacted with their nearest or more distal genes, we only considered protein coding and noncoding RNA genes in GENCODE 19. To evaluate cases where interactions between positive elements and their nearest genes were unresolvable ('same fragment ambiguity'), we determined if a promoter for the nearest gene overlapped at least one HindIII fragment that the positive element did not also overlap. The following terms were considered to be neural annotations: neural tube, hindbrain, cranial nerve, midbrain, forebrain, mesenchyme derived from neural crest, dorsal root ganglion and trigeminal V.

**SNP enrichment analysis and target gene identification.** GWAS SNPs for a total of 11 neuropsychiatric disorders were mined from the GWAS Catalog[39] (December 2018) using a $P$ value threshold of $10^{-6}$. See Supplementary Table 5 for a detailed summary of the studies included for each trait. The GWAS SNPs were expanded to sets of linked SNPs using HaploReg 4.1 (ref. [41]) at an LD threshold of 0.8 according to the reported study population(s) for each SNP. All SNPs were fitted to hg19 and filtered for duplicates by position. Disease- and cell-type-specific enrichment for SNPs was calculated as the number of SNPs overlapping significant PIRs divided by the mean number of SNPs overlapping randomly shuffled PIRs with matching distance distributions. $n = 100$ sets of randomly shuffled PIRs were sampled in each case. To determine whether a GWAS SNP interacted with a target gene, we determined whether it or any of its linked SNPs (expanded to a minimum width of 1 kb) interacted with a promoter associated with the nearest gene. To evaluate cases where interactions between GWAS SNPs and their nearest genes were unresolvable (same fragment ambiguity), we determined if a promoter for the nearest gene overlapped at least one HindIII fragment that a GWAS SNP or any of its linked SNPs did not also overlap. Finally, we derived a list of SNPs for which the SNP was located within 2 kb of the center of an open chromatin peak at a PIR, indicating additional evidence for a functional regulatory variant at that locus. These SNPs are referred to as 'putative regulatory SNPs'.

**Phasing of the WTC11 genome.** The raw WTC11 genome sequence can be downloaded from http://genome.ucsc.edu/cgi-bin/hgTracks?db=hg38&hubClear=https://s3-us-west-2.amazonaws.com/downloads.allencell.org/genome-sequence/ucsc_hubs/WTC_genome_hub/hub.txt. Phasing of the WTC11 genome was performed as previously described[51]. Briefly, WTC11 variants were split by chromosome and phase-informative reads from pcHi-C were extracted using extractHAIRS with the minimum mapping quality set to 10 and the maximum insert size set to 30,000,000 bp (ref. [62]). Phasing was performed with Hapcut using a maximum of 101 iterations. Next, we extracted the maximum variants phased (MVP) haplotype block from the output of Hapcut to use as a seed haplotype. We modified the 'neighborhood correction' aspect of phasing by filtering phased variants whose predicted phase would have a marginal probability below 0.99 according to an in-house implementation of a hidden Markov model (HMM) as described previously[63,64] with a reference haplotype set from the 1000 Genomes Project. Missing variants were imputed using the aforementioned HMM with the reference haplotype set from the 1000 Genomes Project. The WTC11 SNP phasing data are available at the Gene Expression Omnibus under the accession number GSE113483.

**Allelic bias analysis.** We used the WTC11 phasing data along with the allele-specific mapping capabilities of HiC-Pro v.2.11.0 (ref. [52]) to quantify allelic bias between significantly interacting 10-kb bins genome-wide in the excitatory and lower motor neurons. We selected these two cell types because they used homogenous induction of TFs for differentiation, therefore minimizing the noise introduced by conventional differentiation techniques. Briefly, reads were mapped using bowtie v.1.2.1.1 (ref. [65]) to a version of the hg19 reference genome where all sites with heterozygous phased SNPs were masked. Unfiltered HiC-Pro contact maps were used for this analysis. Next, nucleotides at masked polymorphic sites were used to assign reads to either allele. Reads reporting conflicting allele assignments or unexpected bases were filtered out. Reads with at least one allele-specific mate were used to construct allele-specific Hi-C contact maps at a resolution of 10 kb. The allele-specific Hi-C contact maps were intersected with the set of all significant promoter-PIR interactions with score $\geq 3$ to assess allelic bias between interacting 10-kb bins. Only interacting bins with ten or more reads across both alleles were retained ($n = 22{,}162$ bins for excitatory neurons and $n = 21{,}479$ bins for lower motor neurons). A two-tailed binomial test was used to assess allelic bias across each set of interacting bins, and the resulting $P$ values were adjusted using the BH correction to filter out significantly biased loci at an FDR cutoff of 5%. Allelically biased interactions with $P$ values $< 10^{-3}$ are reported in Supplementary Table 10.

**eQTL enrichment analysis.** One-dimensional enrichment of significant eQTLs from GTEX V7 (ref. [55]) at significant versus randomly shuffled PIRs in matched tissue types for excitatory neurons (Brain–Cortex, $n = 136$) and hippocampal DG-like neurons (Brain–Hippocampus, $n = 111$) was performed in the same manner as the chromatin state and SNP enrichment analysis. We used the full set of significant eQTL SNP-gene associations available from GTEx which included 478,903 eQTLs associated with 6,146 significant cis-eQTL genes (eGenes) for the cortex and 221,876 eQTLs associated with 3,262 eGenes for the hippocampus. To determine the two-dimensional enrichment of eQTL–TSS pairs in our significant interaction sets, we first filtered out eQTL–TSS pairs that were within 10 kb of each other or on the same HindIII fragment as this would be below the minimum detectable resolution by pcHi-C. Next, we sampled a set of nonsignificant eQTL–TSS pairs with a matching distance distribution as the set of significant eQTL–TSS pairs for each cell type, controlling for the number of genes around which the eQTL–TSS pairs were centered. We sampled three times the number of nonsignificant eQTL–TSS pairs as the number of significant eQTL–TSS pairs (424,912 significant pairs for 5,826 TSSs and 1,274,736 nonsignificant pairs for 17,570 TSSs in excitatory neurons, and 197,155 significant pairs for 3,083 TSSs and 591,465 nonsignificant pairs for 9,238 TSSs in hippocampal DG-like neurons). Similar results were obtained when using the same number of significant and nonsignificant eQTL–TSS pairs, or when using the same sets of eGenes (data not shown). We compared the distributions of interaction scores for significant interactions supporting the significant and nonsignificant eQTL–TSS pairs by overlapping the eQTL–TSS pairs with our significant interactions (two-tailed two-sample Kolmogorov–Smirnov test).

**Statistics and reproducibility.** Statistical analyses were performed using R 3.4.3 and Microsoft Excel. All of the statistical tests used are described in the relevant sections of the manuscript. $P$ values are provided as exact values where possible and otherwise are reported as a range.

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

## Data availability

All datasets used in this study (pcHi-C, ATAC-seq, RNA-seq, CUT&RUN, and chromosome-wide SNP phasing data) are available at the Gene Expression Omnibus under the accession number GSE113483. Open chromatin peaks and gene expression results for each cell type are also available on Zenodo through the following link: https://zenodo.org/record/3243977

Data can be visualized on the WashU Epigenome Browser using the session bundle ID (session ID in parentheses): 6e375740-8e71-11e9-be37-cb77c4bbb5fc (brain_pchic_nature_genetics_00).

Alternatively, the data can also be visualized on the legacy WashU Epigenome Browser (session ID in parentheses): http://epigenomegateway.wustl.edu/legacy/?genome=g19&session=8OCs2rkpEA (brain_pchic_nature_genetics_00).

Tracks include ATAC-seq signal, chromatin interactions with score ≥5 and RNA-seq plus and minus strand signal for each cell type. HindIII fragments, in vivo-validated enhancer elements, GENCODE 19 genes and GWAS SNPs are also displayed.

## References


58. Rao, S. S. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
59. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
60. Wingett, S. et al. HiCUP: pipeline for mapping and processing Hi-C data. *F1000Res* **4**, 1310 (2015).
61. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
62. Bansal, V. & Bafna, V. HapCUT: an efficient and accurate algorithm for the haplotype assembly problem. *Bioinformatics* **24**, i153–159 (2008).
63. Li, N. & Stephens, M. Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics* **165**, 2213–2233 (2003).
64. Delaneau, O., Zagury, J. F. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat. Methods* **10**, 5–6 (2013).
65. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).

nature research

Corresponding author(s): Yin Shen

Last updated by author(s): Jun 17, 2019

# Reporting Summary

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

## Statistics

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

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

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

## Software and code

Policy information about availability of computer code

| | |
|---|---|
| Data collection | Sequencing data was obtained from the HiSeq 4000, HiSeq X Ten, and NovaSeq 6000 systems (Illumina). |
| Data analysis | All software and code used in the text are now reported in the text along with their version numbers. We used the following software: fastp 0.20.0, HiCUP 0.71, bowtie 2.3.2, CHiCAGO 1.1.8, HOMER 4.10, HaploReg 4.1, HiC-Pro 2.11.0, bowtie 1.2.1.1, https://github.com/kundajelab/atac_dnase_pipelines (June 2018), STAR 2.7.0f, RSEM 1.3.1, TrimGalore 0.4.5, edgeR 3.20.9, Picard Tools 1.141, MACS2 2.1.1, HiCRep 1.4.0, DiffBind 2.6.6, DESeq2 1.18.1. A copy of the custom code used for all the data analysis and figure generation in this study can be viewed and downloaded at the following GitHub repository: https://github.com/stayingsong/brain_pchic. |

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

## Data

Policy information about availability of data

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

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

All datasets used in this study (pcHi-C, ATAC-seq, RNA-seq, CUT&RUN, and chromosome-wide SNP phasing data) are available at the Gene Expression Omnibus under the accession number GSE113483. Open chromatin peaks and gene expression results for each cell type are also available on Zenodo through the following link: https://zenodo.org/record/3243977.
Data can be visualized on the WashU Epigenome Browser using the session bundle ID (session ID in parentheses): 6e375740-8e71-11e9-be37-cb77c4bbb5fc (brain_pchic_nature_genetics_00).
Alternatively, the data can also be visualized on the legacy WashU Epigenome Browser at the following link (session ID in parentheses): http://

epigenomegateway.wustl.edu/legacy/?genome=hg19&session=8OCs2rkpEA (brain_pchic_nature_genetics_00).
Tracks include ATAC-seq signal, chromatin interactions with score >= 5, and RNA-seq plus and minus strand signal for each cell type. HindIII fragments, in vivo-validated enhancer elements, GENCODE 19 genes, and GWAS SNPs are also displayed.

# Field-specific reporting

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

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

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

# Life sciences study design

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

| | |
|---|---|
| Sample size | Our study design, which ranges from between two to two and four to four pairwise comparisons across the various assays, has approximately 80% power to detect a mean difference of 2.39 to 5.66 standard deviations at a nominal significance threshold of 0.05. We chose this study design because effect sizes of this magnitude are compatible with our research goals. |
| Data exclusions | No data were excluded from the analyses. |
| Replication | All attempts at replication were successful and are described in the text. |
| Randomization | Randomization and blinding is not relevant to our study because we do not apply any differential treatment, intervention, or perturbation to our samples. Instead, we compare the epigenomic profiles for different cell types. |
| Blinding | Randomization and blinding is not relevant to our study because we do not apply any differential treatment, intervention, or perturbation to our samples. Instead, we compare the epigenomic profiles for different cell types. |

# Reporting for specific materials, systems and methods

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

## Materials & experimental systems

| n/a | Involved in the study |
|---|---|
| ☐ | ☒ Antibodies |
| ☐ | ☒ Eukaryotic cell lines |
| ☒ | ☐ Palaeontology |
| ☒ | ☐ Animals and other organisms |
| ☒ | ☐ Human research participants |
| ☒ | ☐ Clinical data |

## Methods

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

## Antibodies

| | |
|---|---|
| Antibodies used | We used the following primary antibodies for immunofluorescence: CUX1 (CASP) (Abcam, ab54583, lot: GR3224721-2, 1:500 dilution), MAP2 (Abcam, ab5392, lot: GR3242762-1, 1:1000 dilution), PROX1 (Millipore, MAB5654, lot: 3075604, 1:500 dilution), HB9 (Millipore, ABN174, lot: 3050643, 1:500 dilution), SMI32 (Abcam, ab7795, lot: GR299862-23, 1:1000 dilution), and GFAP (Abcam, ab7260, lot: GR3240356-1, 1:1000 dilution).<br><br>Secondary antibodies for immunofluorescence included: Alexa Fluor 594 goat anti-chicken IgG (Thermo Fisher Scientific, A11042, lot: 1977707, 1:500 dilution), Alexa Fluor 568 donkey anti-mouse IgG (Thermo Fisher Scientific, A10037, lot: 1917938, 1:500 dilution), Alexa Fluor 488 donkey anti-rabbit IgG (Thermo Fisher Scientific, A21206, lot: 1981155, 1:500 dilution), and Alexa Fluor 488 donkey anti-mouse IgG (Thermo Fisher Scientific, A21202, lot: 2018296, 1:500 dilution).<br><br>We used the following antibodies for CUT&RUN: H3K27ac from Active Motif, 39122, lot: 22618011, 1:100 dilution, and CTCF from Millipore, 07-729, lot: 305960, 1:100 dilution. |
| Validation | All primary antibodies for immunofluorescence recognize human proteins and are verified for staining on the manufacturers' websites.<br><br>CUX1 (Abcam, ab54583, lot: GR3224721-2)<br>Mouse monoclonal [2A10] to Protein CASP; Recombinant fragment (GST-tag) corresponding to Human Protein CASP aa 521-621; Suitable for: IHC-P, ICC, WB, IHC-FoFr, ICC/IF, Sandwich ELISA, Flow Cyt; Reacts with: Mouse, Human; This product has been |

referenced in 10 publications.

MAP2 (Abcam, ab5392, lot: GR3242762-1)
Chicken polyclonal to MAP2; Recombinant fragment corresponding to Human MAP2. Mix of recombinant human constructs of projection domain sequences, amino acids 235-1588; Suitable for: ELISA, IHC-Fr, IHC-FoFr, IHC-P, WB, ICC/IF, IHC (PFA fixed); Reacts with: Mouse, Rat, Sheep, Cow, Dog, Human, Cynomolgus monkey, Common marmoset, Aplysia; This product has been referenced in 274 publications.

PROX1 (Millipore, MAB5654, lot: 3075604)
Monoclonal Antibody; Protein A Purfied; Recombinant human Prox1 protein; Anti-Prox1 Antibody, clone 4G10 is an antibody against Prox1 for use in WB, IH; Validated using positive control (mouse dentate granule neurons) and negative control (secondary antibody only). Also validated in fresh formaldehyde-fixed human hippocampal tissue with specific staining in the dentate granule layer.

HB9 (Millipore, ABN174, lot: 3050643)
Polyclonal Antibody; Affinity Purified; KLH-conjugated linear peptide corresponding to human MNX1; Anti-MNX1 (HB9) Antibody detects level of MNX1 (HB9) & has been published & validated for use in Western Blotting & IHC.

SMI32 (Abcam, ab7795, lot: GR299862-23)
Mouse monoclonal [NF-01] to Neurofilament heavy polypeptide; This antibody recognizes a phosphorylated epitope on heavy neurofilament protein (210 kDa) of various species; Suitable for: ELISA, IHC-Fr, ICC, IHC-P, WB, IHC - Wholemount, ICC/IF, Flow Cyt; Reacts with: Mouse, Rat, Cow, Human, Pig; Predicted to work with: a wide range of other species, Mammals; This product has been referenced in 11 publications.

GFAP (Abcam, ab7260, lot: GR3240356-1)
Rabbit polyclonal to GFAP; Specifically recognizes mammalian GFAP on western blots and immunocytochemically. Detects a band of 55kDa corresponding to GFAP and also a GFAP derived 48kDa band; Suitable for: IHC-FoFr, IHC-Fr, IHC-FrFl, ICC/IF, WB, IHC-P, IHC - Wholemount, ICC; Reacts with: Mouse, Rat, Cat, Dog, Human, Common marmoset; Predicted to work with: Cow, Pig, Mammals; This product has been referenced in 343 publications.

Alexa Fluor 568 goat anti-chicken IgG (Thermo Fisher Scientific, A11042, lot: 1977707)
Goat / IgG Polyclonal to Chicken; This product has been referenced in 26 publications.

Alexa Fluor 568 donkey anti-mouse IgG (Thermo Fisher Scientific, A10037, lot: 1917938)
Donkey / IgG Polyclonal to Mouse; This product has been referenced in 32 publications.

Alexa Fluor 488 donkey anti-rabbit IgG (Thermo Fisher Scientific, A21206, lot: 1981155)
Donkey / IgG Polyclonal to Rabbit; This product has been referenced in 70 publications.

Alexa Fluor 488 donkey anti-mouse IgG (Thermo Fisher Scientific, A21202, lot: 2018296)
Donkey / IgG Polyclonal to Mouse; This product has been referenced in 68 publications.

H3K27ac and CTCF antibodies were validated by the ENCODE project and have been used in many publications.

## Eukaryotic cell lines

Policy information about cell lines

| | |
|---|---|
| Cell line source(s) | We used excitatory neurons (i3N iPSCs) from co-author Dr. Li Gan's lab (Mertens et al., 2016 Nature and Wang et al., 2017 Stem Cell Reports), hippocampal DG-like neurons from co-author Dr. Hongjun Song's lab, and lower motor neurons (i3LMN iPSCs) from co-author Dr. Bruce Conklin's lab (Fernandopulle et al., 2018 Curr Protoc Cell Biol). Two batches of primary astrocytes derived from two individuals were purchased from ScienCell. The mouse embryonic fibroblasts (MEF) were isolated in Dr. Hongjun Song's lab. They were derived from E13 embryos extracted from pregnant CF1 mice (Charles River, Strain Code: 023). The MEFs were irrated with 3000 rads before plating them as feeder cells. |
| Authentication | We checked the expression of key marker genes for each cell type using immunofluorescence and RNA-seq. For excitatory neurons, we used VGLUT1 and CUX1, for hippocampal DG-like neurons, we used SOX2 and PROX1, for lower motor neurons, we used HB9 and SMI32, and for astrocytes, we used GFAP. |
| Mycoplasma contamination | All cells used in the present study were verified as mycoplasma contamination free. |
| Commonly misidentified lines (See ICLAC register) | None of the cell lines used are commonly misidentified lines. |