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

Michael Song1,2, Xiaoyu Yang1, Xingjie Ren1, Lenka Maliskova1, Bingkun Li1, Ian R. Jones1, Chao Wang2, Fadi Jacob3,4,5,6, Kenneth Wu7, Michela Traglia8, Tsz Wai Tam1, Kirsty Jamieson1, Si-Yao Lu9,10,11, Guo-Li Ming1,2,25*, Yun Li15,16,17, Jun Yao9,10,11, Lauren A. Weiss1,8, Jesse R. Dixon18, Luke M. Judge2,19, Bruce R. Conklin7,20,21, Hongjun Song4,5,12,13,22, Li Gan3,23,24,25 and Yin Shen1,2,12,25*

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 functional studies. 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 evidence2. 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 adipocytes4. 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 disease5,6. 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.
and enabling novel therapies. Toward this goal, we used wild type human iPSCs (WTC11 line) to generate three neuronal cell types: excitatory neurons, hippocampal DG-like neurons (i3N iPSCs), and lower motor neurons (i3LMN iPSCs) with doxycycline-inducible Ngn2 at the AAVS1 safe-harbor locus. The i3N iPSCs were used to prepare homogenous cultures containing integrated, isogenic, and inducible NGN2, evidencing mature synaptic functions (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 (i3LMN iPSCs)19. 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 (Sciencell). 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).

**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 (iN iPSCs) with doxycycline-inducible Ngn2 at the AAVS1 safe-harbor locus. The iN 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 previously9,13 (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 (i3LMN iPSCs)19. 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 (Sciencell). 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).
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. 20). We first confirmed the reproducibility of contact frequency and saturation of inter-replicate correlation for our pcHi-C libraries using HiCRep16 (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 CHICAGO17, 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 tissues8 (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 promoters18,19 and the capacity of many promoters to doubly function as enhancers20,21. 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 genome22 and previous findings that promoters can be regulated by multiple enhancers23. To examine global chromatin signatures at PIRs, we leveraged chromatin states inferred by ChromHMM24 in matched human brain tissues from the Roadmap Epigenomics Project25 (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 enrichment value via Fisher’s exact test with the combined score). c, Expanded lists of enriched GO terms are available in Supplementary Table 3.
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 ($P = 9.1 \times 10^{-7}$, $t$-test). The neighboring genes MRPL10, PNPO and NFE2L1 were unaffected ($P = 9.1 \times 10^{-10}$, $t$-test). g, CRISPRi silencing of region 2, but not region 3, results in significant downregulation of CDK5RAP3 expression in excitatory neurons ($P = 2.2 \times 10^{-3}$, $t$-test). H3K27ac and CTCF binding sites mapped using CUT&RUN\textsuperscript{26} in excitatory and lower motor neurons and chromatin immunoprecipitation sequencing (ChIP-seq) in astrocytes from ENCODE\textsuperscript{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\textsuperscript{23} 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 (chr11:113,200,000–114,160,000). Monoallelic deletion of this PIR in three independent clones results in significant downregulation of *STRAP* expression in excitatory neurons (P = 6.2 × 10^-3, t = 6.92, d.f. = 3, two-tailed two-sample t-test). 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 × 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 retardation25. *CHAT* is a principal marker for lower motor neuron maturity and function, and *TLR4* is a key regulator of immune activation and synaptogenesis in astrocytes26.

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 HOMER11 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 TADs32–35. Motifs for ASCL1, ISL1, NEUROG2, OLG2 and ZIC3, TFs linked
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, NKKX6-1, ONECUT1, RXF1, RXF2, RXF5, TCF4 and TGF2. 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 pCh-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[1]. 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 their nearest genes (scenario III), 118 interact with both their nearest genes and 484 more distal genes (scenario II) and 60 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 464 more distal genes (scenario I) (Fig. 4c). The remaining 105 elements could not be resolved for interactions with their nearest genes (scenario IV), though 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 and regulates proliferation in nonneuronal cells. 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 iN 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 ($P = 1.6 \times 10^{-5}$, 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 $P = 2.2 \times 10^{-4}$ 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 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. We mined a total of 6,396 unique GWAS SNPs at a significance threshold of $10^{-4}$ 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 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. 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. 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 genotypes (scenario I), 1,243 GWAS SNPs interact with both their nearest gene 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 (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 (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 (Supplementary Fig. 7c). Finally, the MS12 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. The complex itself facilitates spliceosome assembly and is associated with spinal muscular atrophy. To validate this locus, we derived three independent iN 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. 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. 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 and performed allele-specific mapping at a resolution of 10 kb using HiC-Pro. 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 a DRD2 allele-specific deletion (Fig. 6b) 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. 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 demonstrated that significant eQTLs in cortical and hippocampal tissues from GTEx V7 (ref. 38) and 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.

**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. 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.
References

1. Mumbach, M. R. et al. Enhancer connectivity in primary human cells identifies target genes of disease-associated DNA elements. Nat. Genet. 49, 1602–1612 (2017).

2. Zhang, 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 enhancer/promoter function. Nature 513, 539–544 (2014).

5. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. Nat. Methods 10, 629–635 (2013).

6. 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).

7. Miyazaki, Y. et al. Isolation of single-base edited genome human iPS cells with high efficiency and allele selection. Nat. Methods 11, 291–293 (2014).

8. Roadmap Epigenomics, C. et al. Integrative analysis of 111 reference human epigenomes. Cell 155, 247–270 (2013).

9. Mertens, J. et al. Differential responses to lithium in hyperexcitable neurons identifies a regulator of cell-to-cell variation of gene expression. Proc. Natl Acad. Sci. USA 107, 16351–16356 (2010).

10. Zhu, Z. et al. Integration of summary data from GWAS and eQTL studies with the cis-regulatory landscape of open chromatin during human cortical neurogenesis. Cell 172, 289–304 e218 (2018).

11. Won, H. et al. Chromosome conformation elucidates regulatory relationships in developing human brain. Nature 538, 523–527 (2016).

12. Miyazaki, Y. et al. Isolation of single-base edited genome human iPS cells with high efficiency and allele selection. Nat. Methods 11, 291–293 (2014).

13. 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).

14. Yang, T. et al. HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. Genome Res. 27, 1939–1947 (2017).

15. Cairns, J. et al. CHiCAGO: robust detection of DNA looping interactions in Hi-C data. Genome Biol. 17, 127 (2016).

16. 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).

17. Zhang, Y. et al. Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations. Nature 504, 306–310 (2013).

18. Zhao, Y. et al. A tissue-specific and cell-type-based discoidal screen for cis-regulatory element identification in mammalian cells. Nat. Methods 14, 629–633 (2017).

19. Engeit, J. et al. Local regulation of gene expression by IncRNA promoters, transcription and splicing. Nature 539, 452–455 (2016).

20. Consortium, E. P. et al. An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74 (2012).

21. Shen, Y. et al. A map of the cis-regulatory sequences in the mouse genome. Nature 488, 116–120 (2012).

22. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. Nat. Methods 9, 215–216 (2012).

23. Roadmap Epigenomics, C. et al. Integrative analysis of 111 reference human epigenomes. Nature 513, 518–523 (2014).

24. Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for characterization. Nat. Methods 10, 629–635 (2013).

25. 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).

26. Servant, N. et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol. 16, 259 (2015).

27. Booth, H. D., Hirst, W. D. & Wade-Martins, R. The role of astrocyte dysfunction in Parkinson’s disease pathogenesis. Trends Neurosci. 40, 358–370 (2017).

28. Huang, C. W. et al. Conditional knockout of breast carcinoma amplified gene (BCAS2) in mouse forebrain causes dendritic malformation via beta-catenin. Sci. Rep. 6, 34927 (2016).

29. Grimmer, M. et al. Unip, 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).

30. Murakami, K. et al. Development of a conditional test for tau reversibly impairs memory. Nat. Med. 22, 1268–1276 (2016).

31. Wang, C. et al. 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 Allen Institute and its founder P. G. Allen for making this work possible. We thank G. Hon (University of California, San Diego) for sharing pcHi-C probes and the pcHi-C protocol. Genomic analyses were performed by the UCSC Genome Bioinformatics Group (http://genome.ucsc.edu) for providing reagents. We acknowledge the ENCODE Consortium and Bradley Bernstein’s laboratory for generating the ChIP-seq data for astrocytes used in this work. We thank N. Altun, 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. NEI-NEI P30EY008162. This work was supported by the National Institutes of Health (NIH) grant nos. R01AG057497 (to Y. S., 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. NEI-NEI P30EY008162. This work was supported by the National Institutes of Health (NIH) grant nos. R01AG057497 (to Y.S., 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. NEI-NEI P30EY008162. This work was supported by the National Institutes of Health (NIH) grant nos. R01AG057497 (to Y.S., M.G. and H.S.) and R01EY027789 and U19HG009402 (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-HL135558 (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.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2019
Methods

Cell culture. Human excitatory neurons were generated using integrated, isogenic and inducible neuron 2 (Ngn2) iPSCs (iN iPSCs) with doxycycline-inducible mouse Ngn2 integrated at the AAVS1 safe-harbor locus. The iN iPSCs have a well-characterized genetic background (WTC11). A single-cell two-step predifferentiation and maturation protocol was used to generate the excitatory neurons. Briefly, iN iPSCs were incubated with 2 μg/ml doxycycline in predifferentiation media containing knockdown DMEM/F12 supplemented with 8 × 10^{-5} M TP, 1 × NEAA, 1 μg/ml mouse laminin, 10 ng/ml BDNF, 10 μg/ml B27, 0.5% FBS, and 10% 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-4-lysine and laminin plates in maturation media containing equal parts DMEM/F12 and Neurobasal-A with 2 μg/ml doxycycline and supplemented with 0.5% FBS, 20 μg/ml B27, 0.5× GlutaMax, 1× NEAA, 10 ng/ml NT3, 10 ng/ml BDNF, 10 ng/ml N2 supplement, 0.5% FBS, and 10% 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× N2 and 1× B27, 20% knockout serum replacement, and passaged once using trypsin and EDTA before harvesting.

Cell culture. Methods

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 (l’LMN iPSCs). Briefly, l’LMN iPSCs were maintained on growth factor reduced Matrigel in StemFit media (Nacalai). On day 0, 1.5×10^5 l’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-4-lysine and laminin at a density of 2.5×10^5 cells per well. From day 3 to 7, the cells were treated with a pulse of 40 μM RedU 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 ScientCell Research Laboratories (catalog no. 1800) and cultured using the recommended media (catalog no. 1801). Cells were cultured in flasks coated with poly-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 (pHi-C). In situ Hi-C libraries for excitatory neurons, hippocampal DG-like neurons, lower motor neurons and astrocytes were constructed from 1 to 2×10^6 cells (fixed in 1% PFA) using HindIII as a restriction enzyme as previously described. pHi-C was performed using biotinylated RNA probes conjugated to streptavidin-coated magnetic beads (0.5 mg, Invitrogen no. 65601) and 10 ng/ml B27. 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.

Calling significant promoter-PIR interactions. Paired-end sequencing reads were first trimmed using fastp v0.20.0 (ref. 39) running the default settings before being mapped, filtered, and deduplicated using HiCUP v0.7.1 (ref. 40) with parameter ‘-c’ v2.3.2 (ref. 41) and filtering for dities of less than 10 kb (ref. 42). 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 v1.1.8 (ref. 43) 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 PI Rs using CRISPRi deletion. To validate genomic interactions captured by pHi-C, candidate PI Rs were targeted for CRISPRi-mediated deletion in the iN iPSCs. Pairs of sgRNAs targeting the putative regulatory element as local anchors on core promoters in the candidate PI R were designed for each locus of interest. All sgRNAs were synthesized by Synthego. Cas9 protein was sourced from QB3-Berkeley. To generate deletion lines, fine-mapping ES cells and human lower motor neurons were differentiated from WTC11 iPSCs using a hippocampal fate using WNT and BMP in medium containing 1× N2 and 1× B27, 20% knockout serum replacement. Four-week-old EBs were patterned toward a neural ectoderm fate using dual SMAD inhibition and replacing weekly until harvesting and 7 to 8-week-old excitatory neurons were used for library preparation.

In situ Hi-C libraries for excitatory neurons, hippocampal DG-like neurons, lower motor neurons and astrocytes were constructed from 1 to 2×10^6 cells (fixed in 1% PFA) using HindIII as a restriction enzyme as previously described. pHi-C was performed using biotinylated RNA probes conjugated to streptavidin-coated magnetic beads (0.5 mg, Invitrogen no. 65601) and 10 ng/ml B27. 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.

Validation of PI Rs using CRISPRi. Excitatory neurons induced from iN iPSCs were infected with lentivirus carrying dsCas9 KRAB-blast (Addgene no. 12239) and puromycin-selected lines were sourced from QB3-Berkeley. A seq-opti vector (Addgene no. 106280) was used for sgRNA expression. sgRNAs were co-transfected with lentivirus packaging plasmids pMD2.G (Addgene no. 12235) and psPAX (Addgene no. 12260) into 293 T cells with PolyJet (SigaGen 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-15 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: TCCTCCGAGGGATTTCGTGCGGA and reverse primer: CCCCCCATCTGCAGTCGTTGCAC). Integration of the WPRE fragment was quantified by PCR comparison with a cell line containing 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 20X SSPE, 1.05 μl EDTA, 10 μl 50x 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 μl μl 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 mM NaCl and 1 mM EDTA. The beads were washed once with 0.1% 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).

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 peaks to perform motif enrichment analysis using HOMER 4.10 (ref. 44) running the default settings. The cumulative binomial distribution was used for motif scoring. The genome
SNP enrichment analysis and target gene identification. GWAS SNPs for a total of 11 neuropyschiatric disorders were mined from the GWAS Catalog (December 2018) using a threshold of 10−8. See Table 1 for detailed summary of the studies included for each trait. The GWAS SNPs were expanded to sets of linked SNPs using HaploReg v.4.1 (ref. 4) 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 number of randomly shuffled PIRs 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 or not 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 unsolvable (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 neutral annotations: neural tube, hindbrain, cranial nerve, midbrain, forebrain, mesenchyme derived from neural crest, dorsal root ganglion and trigeminal V.

Allelic bias analysis. We used the WTC11 phasing data along with the allele-specific mapping capabilities of Hi-C Pro v.2.11.0 (ref. 5) 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 use 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. 6) to a version of the hg19 reference genome where all sites with heterozygous phased SNPs were masked. Unfiltered Hi-C 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 ≥ 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 < 10−4 are reported in Supplementary Table 10.

eQTL enrichment analysis. One-dimensional enrichment of significant eQTLs from GTEx V7 (ref. 7) 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 SNPs from all GTEx tissues which included 478,903 eQTLs associated with 6,146 significant cis-eQTL genes (eGenes) for the cortex and 221,676 eQTLs associated with 3,262 eGenes for the hippocampus. To determine 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 1,491,828 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 in 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/legrity?eGenes=19&session=8OCx2rkpEA (brain_pchic_nature_genetics_00). Tracks include ATAC-seq signal, chromatin interactions with score ≥ 5 and range plus and minus strand signals 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, 884–889 (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, 1153–1159 (2008).
63. Li, N. & Stephens, M. Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. Genetics 165, 2213–2233 (2003).
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.

Conf/confirmed

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

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

Software and code

Policy information about availability of computer code

Data collection
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/324397

Data can be visualized on the WashU Epigenome Browser using the session bundle ID (session ID in parentheses): 6e375740-8e71-11e9-be37-cb77c4bb5fc (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://
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**

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).

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).

We used the following antibodies for CUT&RUN: H3K27ac from Active Motif, 39122, lot: 20618011, 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.

CUX1 (Abcam, ab54583, lot: GR3224721-2)

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
referred 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-FrFl, 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, IHC; 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.

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 irradiated 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
None of the cell lines used are commonly misidentified lines.