cis-Regulatory Chromatin Contacts in Neural Cells Reveal Contributions of Genetic Variants to Complex Neurological Disorders

Authors: Michael Song[1,2], Xiaoyu Yang[1], Xingjie Ren[1], Chao Wang[3], Fadi Jacob[4,5], Kenneth Wu[6], Michela Traglia[7], Bingkun Li[1], Lenka Maliskova[1], Ian Jones[1], Kirsty Jamieson[1], Si-Yao Lu[8], Tse Wai Tam[1], Guo-Li Ming[4,9,10,11], Jesse Dixon[12], Lauren A. Weiss[1,7], Jun Yao[8], Luke Judge[6,13], Bruce Conklin[6], Hongjun Song[4,9,10,14], Li Gan[3], Yin Shen[1,2,15]*

Affiliation:

[1]Institute for Human Genetics, University of California, San Francisco, San Francisco, CA, USA
[2]Pharmaceutical Sciences and Pharmacogenomics Graduate Program, University of California, San Francisco, San Francisco, CA, USA
[3]Gladstone Institute for Neurological Disorders, San Francisco, CA, USA
[4]Department of Neuroscience and Mahoney Institute for Neurosciences, Perelman School for Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA.
[5]The Solomon H. Snyder Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA.
[6]Gladstone Institute for Cardiovascular Disease, San Francisco, CA, USA
[7]Department of Psychiatry, University of California, San Francisco, San Francisco, CA, USA
[8]State Key Laboratory of Membrane Biology, Tsinghua-Peking Center for Life Sciences, School of Life Sciences, Tsinghua University, Beijing, China.
[9]Department of Cell and Developmental Biology, University of Pennsylvania, Philadelphia, PA 19104, USA.
[10]Institute for Regenerative Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA.
[11]Department of Psychiatry, University of Pennsylvania, Philadelphia, PA 19104, USA.
[12]Salk Institute for Biological Studies, La Jolla, CA, USA
[13]Department of Pediatrics, University of California, San Francisco, San Francisco, CA, USA
[14]The Epigenetics Institute, University of Pennsylvania, Philadelphia, PA 19104, USA.
[15]Department of Neurology, University of California, San Francisco, San Francisco, CA, USA

*Corresponding Author: yin.shen@ucsf.edu

Abstract

Mutations in gene regulatory elements have been associated with a wide range of neurological disorders. However, due to their inherent cell type specificity and difficulties in characterizing their regulatory targets, our ability to identify causal genetic variants has remained limited. To address these constraints, we perform integrative analysis of chromatin interactions (promoter capture Hi-C), open chromatin regions (ATAC-seq), and transcriptomes (RNA-seq) in four functionally distinct cell types associated with neurological disorders: iPSC-induced cortical excitatory neurons, hippocampal dentate gyrus (DG)-like neurons, lower motor neurons, and human primary fetal astrocytes. We identify hundreds of thousands of cell type-specific, long-range *cis* interactions between promoters and distal promoter-interacting regions (PIRs), enabling us to link regulatory elements to their target genes and identify putative pathways that are dysregulated in disease. Finally, we perform validation of novel PIRs linked to neuronal function using CRISPR/Cas9 editing in human neurons, demonstrating downregulation of *CDK5RAP3*, *STRAP*, and *DRD2* upon knockdown of their corresponding enhancers. Our study demonstrates how *cis*-regulatory interactions in human neural cells can be used to elucidate the contribution of genetic variants to complex neurological disorders.

Introduction

A large number of genetic variations 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 in development and function. However, deciphering the roles of regulatory variants in disease pathogenesis remains nontrivial due to lack of annotation for regulatory sequences in the physiologically relevant cell types. Furthermore, *cis*-regulatory elements often interact with their cognate genes over long genomic distances, precluding a systematic mapping of promoter and regulatory element connectivity. As a result, our ability to derive biological insights from noncoding variants in genome wide association studies (GWAS) has remained limited. 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 *IRX3* in the human brain and both *IRX3* and *IRX5* in adipocytes, respectively[3,4]. The *FTO* locus in obesity illustrates the potential complexity and cell type specificity in

how noncoding variants contribute to disease. However, such well-annotated cases are rare, and we still lack a systematic annotation of GWAS SNPs to their regulatory targets, especially in the context of complex neurological disorders. Previous epigenomic annotations of the germinal zone (GZ) and cortical and subcortical plate (CP) in the human brain revealed the importance of 3D chromatin structure in gene regulation and disease[5,6]. However, these studies utilized complex, heterogeneous tissues, limiting their ability to interpret genome regulation and function in a cell type-specific manner. Therefore, mapping the epigenomic landscape and 3D chromatin structure in physiologically relevant, constitutive, and widely used cell types should offer significant advantages for prioritizing causal variants, interpreting their contributions to complex neurological disorders, and discovering novel treatments and biomarkers for improving patient care.

Towards this goal, we used wild type human induced pluripotent stem cells (iPSCs) from the WTC11 line to generate three subtype-specific neuronal lineages: cortical excitatory neurons[7], hippocampal dentate gyrus (DG)-like neurons[8], and lower motor neurons. Normal GFAP-positive fetal astrocytes derived from 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 using promoter capture Hi-C (pcHi-C), open chromatin regions using ATAC-seq, and transcriptomes using RNA-seq, we define maps of promoters and distal promoter-interacting regions (PIRs) across each of the four cell types. We show that PIRs are not only highly cell type-specific, but they are also enriched for regulatory chromatin signatures including open chromatin regions, active chromatin states such as enhancers, and eQTLs. In addition, our interactions with PIRs recover 34.4% of *in vivo*-validated enhancer elements on the VISTA Enhancer Browser[9] and reveal novel insights into their gene targets. We find that neurological disorder-associated SNPs from genome wide association studies exhibit differential enrichment at PIRs across the four cell types, suggesting distinct roles in disease pathophysiology for the neuronal subtypes and astrocytes. Finally, we use CRISPR/Cas9 editing to knock out several disease-associated or cell type-specific regulatory loci at PIRs interacting with the promoters of *STRAP*, *DRD2,* and *CDK5RAP3*, all of which exhibit significant downregulation in response to the genomic perturbations. In summary, our study presents a roadmap for the comprehensive annotation of human cell type-specific PIRs in the CNS, and it advances our ability to elucidate novel biological mechanisms by which noncoding variants contribute to complex neurological disorders.

Results

Characterizing the epigenomic landscape of long-range chromatin interactions in human brain cells

To study the general features of the epigenomic landscape and 3D chromatin structure for specific cell types in the human CNS, we focused on isogenic, iPSC-induced cortical excitatory neurons, hippocampal DG-like neurons, and lower motor neurons, three functionally distinct neuronal subtypes which are currently impractical to isolate *in vivo*. Cortical excitatory neurons were induced from a wild type male iPSC line (WTC11) containing an integrated, isogenic, and inducible neurogenin-2 (Ngn2) cassette (i^3N iPSCs) with doxycycline-inducible mouse Ngn2 integrated at the AAVS1 safe-harbor locus[7]. The i^3N iPSCs enabled us to obtain homogenous human cortical excitatory neurons in culture after two weeks. For our epigenomic characterizations, we used eight-week-old neurons which were positive for the cortical neuron markers CUX1 and FOXG1 as well as the glutamatergic neuron marker VGluT1[10]. In addition, whole-cell patch-clamp recordings and CNQX blocking confirmed the presence of functional glutamatergic synaptic transmissions in these neurons[7] (**Supplementary Fig. 1A)**. Hippocampal DG-like neurons were differentiated from the WTC11 line using defined factors modified from an established protocol[8,11] and exhibited marked expression for the DG granule cell marker PROX1 (**Supplementary Fig. 1A)**. Lower motor neurons (LMN) were induced from a WTC11 line containing integrated, isogenic, and inducible copies of NGN2, ISL1, and LHX3 with doxycycline-inducible mouse Ngn2 integrated at the AAVS1 safe-harbor locus (i^3LMN iPSCs)[12]. These induced cells exhibited homogenous expression of the lower motor neuron markers HB9 and SMI32 (**Supplementary Fig. 1A)**. In addition to these three neuronal subtypes, we used two batches of primary fetal astrocytes passaged once *in vitro* and isolated from two nineteen week, gastrulating male fetal brain samples utilizing GFAP as a marker (ScienCells). Based on the donor's age, as well as transcriptional signatures for dozens of marker genes distinguishing astrocyte progenitor cells (APCs) from mature astrocytes (e.g. low expression of the mature astrocyte markers *AGXT2L1* and *WIF1* and high expression of the APC markers *TOP2A* and *TNC*[13]) (**Supplementary Fig. 1b**), the astrocytes were determined to most likely be APCs. Detailed staining and characterization results for the four cell types used in this study are available in **Supplementary Fig. 1A**. We constructed pcHi-C, ATAC-seq, and RNA-seq libraries from two or three biological replicates for each of these cell types (**Supplementary Table 1**) and performed integrative analysis on the resulting datasets (**Fig. 1a)**. To map promoter-centric, long-range chromatin interactions, we performed pcHi-C using a set of 280,445 unique RNA probes targeting the promoters of

140 19,603 coding genes and 3,311 noncoding RNAs in the GENCODE 19 annotation (see Jung et al., under
141 revision). Using CHiCAGO[14], we identified significant chromatin interactions with scores ≥ 5 over a
142 total of 195,322 unique interacting bin pairs across the four cell types, with 73,890, 108,156, 66,978, and
143 84,087 significant interactions being represented in the cortical excitatory neurons, hippocampal DG-
144 like neurons, lower motor neurons, and fetal astrocytes, respectively (**Supplementary Table 2**). Since
145 topologically associated domains (TADs) have been shown to impose physical boundaries for promoter-
146 enhancer interactions, we first determined whether significant interactions in these neural cell types were
147 similarly restricted by TAD structure. We found that approximately 98% of detected interactions resided
148 within TADs as annotated using HiC datasets in human fetal brain tissue[6] (**Fig. 2b**). Overall, 83.9% of
149 coding gene promoters and 31.9% of noncoding RNA promoters in the GENCODE 19 annotation
150 participate in interactions for at least one cell type, with a total of 17,065 coding gene promoters (12,722
151 in cortical excitatory neurons, 16,021 in hippocampal DG-like neurons, 12,538 in lower motor neurons,
152 and 12,741 in fetal astrocytes) participating in interactions, and with 80% of the PIRs interacting within
153 a distance of 160 kb (**Fig. 1c, Supplementary Fig. 1c and Supplementary Fig. 2a**). Approximately
154 40% of interactions occur between promoters (promoter-promoter interactions), while the remaining
155 60% occur between promoter- and non-promoter-containing bins (**Fig. 1d**). The observed numbers of
156 promoter-promoter interactions can likely be attributed to transcriptional factories of co-regulated genes,
157 widespread co-localization of promoter sequences[15,16] and the fact that many promoter sequences can
158 doubly function as enhancers[17,18]. Finally, between 20 to 40% of distal ATAC-seq peaks at PIRs are cell
159 type-specific (**Fig. 1e**), indicating the potential for PIRs to orchestrate cell type-specific gene regulation
160 at open chromatin regions, and with fetal astrocytes harboring the largest proportion of cell type-specific
161 distal ATAC-seq peaks, a feature likely attributable to their non-neuronal lineage.
162
163 We observed that promoters often participated in interactions with more than one PIR (**Fig. 2a**). This
164 observation is consistent with the large number of regulatory elements in the human genome[19] and
165 previous findings that each promoter can be regulated by multiple enhancers[20]. To examine global
166 chromatin signatures at PIRs, we took advantage of publicly available ChromHMM[21] inferred chromatin
167 states from the Roadmap Epigenomics Project[22] in matched human brain tissues for three of our cell
168 types (dorsolateral prefrontal cortex for cortical excitatory neurons, hippocampus for hippocampal DG-
169 like neurons, and normal human astrocytes for fetal astrocytes). We show that PIRs are highly enriched
170 for active chromatin features including ATAC-seq peaks, enhancers, and transcriptional start sites

171 (TSSs), while simultaneously exhibiting depletion or lack of enrichment for the repressive
172 heterochromatin state (**Fig. 2b**). Next, we determined whether interactions with active or repressive
173 chromatin features could influence the activity of interacting promoters by performing global expression
174 analysis for promoters interacting with PIRs localized with either enhancer or repressive states. We
175 show that promoters interacting with enhancer-PIRs exhibit significantly increased levels of
176 transcription relative to promoters interacting with repressive-PIRs (two sample t-test, one-sided, $p<10^{-}$
177 ^{60}) (**Fig. 2c, Supplementary Fig. 2b**). Since multiple enhancers can interact with and regulate the same
178 promoter, we next asked whether interactions with multiple enhancer-PIRs could present evidence for
179 additive effects on transcription. To address this question, we grouped genes according to the numbers
180 of interactions their promoters formed with enhancer-PIRs in each cell type and discovered a positive
181 correlation between the number of enhancer-PIR interactions and the mean expression level in each
182 group (linear regression, p=0.0026) (**Fig. 2d, Supplementary Fig. 2c**). These results demonstrate that
183 integrative analysis of pcHi-C, ATAC-seq, and RNA-seq data can identify long-range, regulatory
184 chromatin interactions that are enriched for active regulatory features, and these regulatory sequences
185 are capable of modulating gene expression in a cell type-specific manner.
186
187 **Cell type-specific PIRs contribute to cellular identity**
188 We find our significant interactions between promoters and PIRs to be cell type-specific, with a total of
189 12,254, 29,875, 6,757, and 11,424 interactions that are specific to cortical excitatory neurons,
190 hippocampal DG-like neurons, lower motor neurons, and fetal astrocytes, respectively (**Supplementary**
191 **Fig. 3a**). Meanwhile, 58,809 or 30.1% of all unique interactions are shared between all neural cell types.
192 For example, interactions between the promoter of *GRIN2B* and its own intron were found to occur only
193 in hippocampal DG-like neurons (**Supplementary Fig. 3b**). In addition, *GRIN2B*, which encodes a
194 subunit of the NMDA receptor ion channel and whose disruption is associated with mental retardation
195 and epilepsy[23], was not expressed in the other three cell types. In conjunction with the observed
196 enrichment of active chromatin states at PIRs, these cell type-specific aspects of promoter-PIR
197 interactions suggest lineage-specific regulatory roles for PIRs. We also suspected that PIRs could be
198 used to establish cellular identity. To explore this idea, we performed hierarchical clustering based on
199 the mean interaction scores at each interacting promoter-PIR loci. We show that PIR interactions can
200 reliably group cell types according to their cellular identity and function, with the three neuronal
201 lineages clustering together more tightly than with the fetal astrocytes (**Fig. 3a**). These cell type-specific

interactions are enriched for at the promoters of key genes that recapitulate important biological pathways. For example, CHAT is a principal marker for lower motor neuron maturity and function, and its expression was not only observed to be the highest in lower motor neurons, but cell type-specific interactions were also observed bridging the *CHAT* promoter and distal open chromatin loci with potential regulatory functions (**Supplementary Fig. 3b**).

Gene regulation is coordinately controlled by both transcription factors (TFs) and cis-regulatory elements. Therefore, PIRs identified in interactions for each cell type provide a unique opportunity to investigate potential mechanisms driving cell type-specific gene regulation. We collected cell type-specific distal open chromatin peaks at PIRs for each cell type and used HOMER[24] to systematically evaluate TF motif enrichment in the set of putative regulatory open chromatin loci for each cell type. We report the most highly enriched TF motifs across all cell types along with the expression levels of their corresponding TFs (**Fig. 3b**). The CTCF motif is highly enriched in all cell types, consistent with its role to mediate proper looping between promoters and distal regulatory elements[25-28]. Other TF motifs display various patterns of cell type-specific enrichment and expression, with motifs for NEUROG2 and the RFX factors enriched for in the excitatory neuronal lineages, and NKX6-1 enriched across all three neuronal cell types. Other neurologically relevant TFs involved in cell fate determination include CUX2 and HNF6, which are enriched for in cortical excitatory neurons and lower motor neurons, ISL1 in lower motor neurons, and the RUNX and TEAD families which are enriched for in fetal astrocytes. In addition, the motif for TBR1 is enriched in hippocampal DG-like neurons, which is consistent with its established role in NMDA receptor assembly to drive long term potentiation and maintain plasticity in the hippocampus. Motif enrichment is not always correlated with the expression of the corresponding TFs, which may reflect synergistic interactions between cell types in the complex, heterogeneous tissues of the human brain. For example, NRF2 is a crucial 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[29]. Therefore, its shared expression across all cell types may reflect the supportive, neuroprotective role that astrocytes provide for neurons and the specialization of cellular functions in the human brain. Alternatively, TFs do not have to be expressed at high levels to perform their cellular functions due to the existence of additional avenues for regulation at the post-transcriptional and protein levels. Overall, our results demonstrate that PIRs contribute to cell type-specific functions and have the capacities both to recapitulate known and to reveal novel regulators.

233

234 ***In vivo* validation of PIRs and linking enhancer elements to their target genes**

235 Regulation of target genes by enhancers and regulatory elements is thought to be mediated by physical
236 chromatin looping. Congruent with this concept, interactions detected by pcHi-C have the potential to
237 link enhancer elements to their target genes. The VISTA Enhancer Browser is a database containing
238 experimentally validated human and mouse noncoding sequences with enhancer activities. To date, it
239 contains 2,845 tested elements, 1,574 of which exhibited positive enhancer activity during embryonic
240 development[9]. However, the regulatory targets for these validated enhancer elements are largely not
241 known. To address this knowledge gap, we provide detailed, cell type-specific annotations of putative
242 target genes for each positive enhancer element based on our significant promoter-PIR interactions
243 (**Supplementary Table 3**). In total, our interactions recover 532 of 1,547 or 34.4% of *in vivo* validated
244 enhancer elements with human sequences, and 331 of which were further annotated as neural enhancers
245 according to tissue-specific patterns of LacZ staining in mouse embryos (**Fig. 4a**). Of the 532 positive
246 enhancer elements at PIRs that were found to interact with promoters, 134 were found to interact with
247 their nearest neighboring genes, whereas another 230 interacted exclusively with more distal genes (**Fig.
248 4b**). (The 168 remaining elements were too close to their nearest neighboring genes to reliably resolve
249 interactions.) Therefore, while only 25% of the validated elements were found to interact with their
250 nearest neighboring genes, we were able to utilize pcHi-C data to identify an additional 827 cell type-
251 specific regulatory targets for the group of 532 positive enhancer elements at PIRs.

252

253 **Validation of PIRs detected in human brain cells using CRISPR**

254 To validate that the PIRs we detected were indeed regulating the expression of their cognate genes, even
255 across long distances, we performed CRISPRi experiments in i^3N iPSC-induced neurons to determine if
256 targeting the distal regulatory element would influence gene expression. We also used an orthogonal
257 technique in CRISPR/Cas9 genome editing to delete PIRs in i^3N iPSCs, followed by differentiation of
258 the i^3N iPSCs into cortical excitatory neurons and quantification of any differences in gene expression
259 relative to the wild type by qPCR. Two of the PIRs located in an intergenic region on chromosome 17
260 were found to physically interact over 30 kb away with the promoter of *CDK5RAP3* in cortical
261 excitatory neurons (**Fig. 4c**). CDK5RAP3 is a potential regulator of CDK5, which plays a role in
262 neuronal differentiation and migration[30] and regulates cell proliferation in non-neuronal cells[31].
263 Interestingly, one of the two interacting PIRs was found to overlap a positively annotated 6 kb brain

264 enhancer element (*mm999*) from the VISTA Enhancer Browser, for which no knowledge of the target
265 gene was previously available. Both the deletion and CRISPRi targeting of an of a 2 kb open chromatin
266 peak (region #1) within the enhancer sequence resulted in significant down-regulation of *CDK5RAP3*
267 expression in human cortical excitatory neurons (two sample t-test, two-sided, $p=7.7\times10^{-7}$), but not the
268 neighboring genes (**Fig. 4c** and **Supplementary Fig. 3c**). When we tried to delete the other PIR for the
269 *CDK5RAP3* promoter (region #2), we observed massive cell death of iPSCs immediately following
270 introduction of the Cas9-sgRNA protein complex. Among the cells that survived the transfection, we
271 picked 48 individual clones but failed to isolate any clones with deletions, suggesting that this region is
272 required for maintaining *CDK5RAP3* expression in iPSCs, consistent with *CDK5RAP3*'s role in
273 promoting cell survival.
274
275 **Cell type-specific enrichment and identification of novel target genes for neurological disorder-**
276 **associated variants**
277 Previous large-scale epigenomic studies in human tissues and cell lines highlighted the importance of
278 disease-associated variants at distal regulatory regions[22] and the need for high-throughput approaches to
279 prioritize variants for further validation. We used our physical chromatin interaction data to annotate
280 genes targeted by noncoding GWAS SNPs in a cell type-specific manner. We downloaded a total of
281 3,362 tag SNPs from the GWAS Catalog at a significance threshold of 10^{-6} for ten neurological
282 disorders including Alzheimer's disease (AD), attention deficit hyperactivity disorder (ADHD),
283 amyotrophic lateral sclerosis (ALS), autism spectrum disorder (ASD), bipolar disorder (BP), epilepsy
284 (EP), frontotemporal dementia (FTD), Parkinson's disease (PD), schizophrenia (SCZ), and unipolar
285 depression (UD)[32]. Tag SNPs reported in the studies were imputed at a linkage disequilibrium (LD)
286 threshold of 0.8 using HaploReg[33] and filtered to obtain a total of 36,728 noncoding SNP loci across all
287 disorders (**Supplementary Table 4**). We found the imputed SNPs to be enriched for in our significant
288 PIRs compared with shuffled distance-matched PIRs in both a disease- and cell type-specific manner
289 (**Fig. 5a**). While PIRs for all four cell types exhibit moderate to high levels of enrichment for SNPs in
290 ADHD, ASD, and SCZ, SNPs for EP and FTD are not significantly enriched in PIRs for any of the cell
291 types as compared with background controls. Possible explanations are that noncoding SNPs in EP and
292 FTD may be disrupting functions in alternative cell types, acting outside of influencing chromatin-
293 mediated transcriptional regulation, or deriving from studies lacking power or reflecting a highly
294 heterogeneous disease biology. Meanwhile, SNPs for ALS and BD were enriched for in the neuronal

295 lineages but not the astrocytes. We also find that AD SNPs are highly enriched for in cortical excitatory
296 neuron and hippocampal DG-like neuron PIRs but not in PIRs for fetal astrocytes, whereas PD SNPs
297 display a complementary pattern of enrichment in fetal astrocyte PIRs but not PIRs for neuronal
298 cells. This enrichment pattern is consistent with the notion that AD is largely a disease of the forebrain,
299 whereas PD is more a disease of the midbrain, in which cortical excitatory neurons and hippocampal
300 DG-like neurons are absent. These findings also support the roles of excessive NMDA receptor activity
301 leading to neurodegeneration in glutamatergic neurons and early damage to adult hippocampal tissues in
302 well-known aspects of AD pathophysiology. Conversely, the enrichment of PD SNPs in fetal astrocyte
303 PIRs supports the emerging theory that astrocytes may play an initiating role in PD[34], based on evidence
304 that numerous genes implicated in PD have unique functions in astrocyte biology, as well as the
305 neuroprotective roles that astrocytes serve for dopaminergic neurons in the substantia nigra.
306
307 We find that 30 to 60% of noncoding tag SNPs have at least one SNP in LD overlapping with our
308 significant PIRs in one or more cell types (**Fig. 5b**). As it is common practice in association studies to
309 assign noncoding tag SNPs to their nearest neighboring genes, we were interested in determining what
310 proportion of noncoding tag SNPs had at least one SNP in LD interacting with their nearest neighboring
311 genes. Overall, we find that 549 noncoding tag SNPs have at least one SNP in LD interacting with their
312 nearest neighboring gene, whereas 399 noncoding tag SNPs interact exclusively with more distal genes
313 (**Fig. 5c, Supplementary Fig. 4a**). Furthermore, our interactions provide 4,823 additional cell type
314 specific gene targets other than the nearest neighboring gene for the set of all noncoding tag SNPs.
315 These results are summarized in **Supplementary Table 5**. Notably, one such interaction recapitulated a
316 previously published significant interaction between the *FOXG1* promoter and a PIR containing
317 noncoding SCZ SNPs over 700 kb away[6] (**Supplementary Fig. 4b**). To prioritize the list of noncoding
318 SNPs for variants possessing putative regulatory function and to address the challenge of determining
319 which SNPs are causal in potentially complex LD blocks, we focused on disease-associated SNPs in
320 PIRs that are simultaneously localized with open chromatin regions. The resulting curated lists of
321 disease-associated SNPs overlapping with ATAC-seq peaks and interacting with regulatory targets in
322 one or more cell types are reported in **Supplementary Table 5**. PIRs for these regulatory SNPs were
323 found to interact in a cell type-specific manner with genes possessing putative relevant function in the
324 context of their derivative diseases, demonstrating that an approach jointly prioritizing non-coding SNPs

325 according to their localization with open chromatin regions and specifying their regulatory targets using
326 pcHi-C holds significant promise for identifying novel regulatory mechanisms of disease.
327
328 **Validation of PIRs containing neurological disorder-associated regulatory variants**
329 To validate PIRs associated with variants for complex neurological disorders, we selected two
330 regulatory loci targeting the *STRAP* and *DRD2* genes for knockdown via both CRISPRi and
331 CRISPR/Cas9-mediated deletion. STRAP influences the cellular distribution of the survival of motor
332 neuron (SMN) complex, which in turn facilitates spliceosome assembly and is associated with spinal
333 muscular atrophy[35]. Notably, interactions were detected across all cell types between the promoter of
334 *STRAP* and a distal PIR over 300 kb away containing regulatory BD SNPs in an intron for *PTPRO* (**Fig.**
335 **5d**). We derived three clones containing bi-allelic deletions for this candidate PIR and observed
336 significant down-regulation of *STRAP* expression in human cortical excitatory neurons (two sample t-
337 test, two-sided, p=9.4×10^{-13}). In addition, CRISPRi targeting of this PIR consistently revealed
338 significant down regulation of STRAP gene (**Supplementary Fig. 4c).** Next, we focused on a novel PIR
339 20 kb upstream from the promoter of *DRD2*, a gene encoding the D2 subtype of the dopamine receptor.
340 Previously, rs2514218, a noncoding variant 47 kb upstream from DRD2, was found to be significantly
341 associated with anti-psychotic drug response in a cohort of 100 schizophrenia patients[36]. This risk
342 variant is actually in LD with a cluster of regulatory SCZ SNPs that are localized to open chromatin
343 regions within our PIR of interest. *DRD2* is also the gene associated with the well-known Taq1A
344 polymorphism that is linked to addiction, anxiety, depression, and social problems, and for which the A1
345 allele is associated with reduced dopamine receptor density in neurons. We first demonstrate that mono-
346 allelic deletion of this PIR in three independent clones leads to significant down-regulation of *DRD2*
347 transcription (two sample t-test, two-sided, p=2.7×10^{-7}) (**Fig. 5e, Supplementary Fig. 4d**). Next, we
348 confirm that mono-allelic deletion of the *DRD2* PIR leads to allelic imbalance in the expression of
349 *DRD2* in deletion clones, both by direct sequencing and through the more quantitative approach of
350 performing TOPO cloning and sequencing the PCR products of SNP-containing *DRD2* cDNA. *DRD2* is
351 a key gene possessing multifaceted roles in human brain function, and it has been implicated in
352 numerous other neurological disorders including bipolar disorder, addiction, migraine, and obesity[37]. By
353 identifying and validating novel regulatory loci for genes such as *DRD2*, our integrative approach
354 enables the discovery of novel therapeutic strategies, such as modulating gene expression through

355 targeting regulatory variants, and offers significant promise for developing novel insights and treatments
356 for otherwise intractable complex neurological disorders.
357
358 **Genetics variants contribute to chromatin interaction bias and alterations in gene expression**
359 Since regulatory variants and other genetic perturbations are thought to disrupt chromatin looping
360 between promoters and PIRs, we were interested to see if we could detect any instances of allelic bias
361 across our sets of significant promoter-PIR interactions. Therefore, we took advantage of our pcHi-C
362 data and performed genome-wide phasing of WTC11[38] variants using HaploSeq[39]. The resulting phased
363 WTC11 genome was used to identify interactions exhibiting allelic bias between 10 kb genomic bins
364 using HiC-Pro. We identified 22 of 3,962 or 0.56% of the interacting bins as being significantly biased
365 at a false discovery rate of 5%, suggesting that genetic diversity can indeed contribute to imbalanced
366 interactions between the two alleles (**Fig. 6a**). One notable loci presented biased interactions between a
367 PIR containing regulatory BD SNPs and the promoter of *SYT17* (q=0.0196). *SYT17* encodes one of the
368 synaptotagmins, membrane-trafficking proteins which mediate synaptic function and regulate calcium-
369 controlled neurotransmitter release[40]. The detection of chromatin interactions with allelic bias at this
370 locus suggests that regulatory SNPs can increase individual risk for BD by altering regulatory
371 interactions for *SYT17* **(Fig. 6b**).
372
373 In addition to determining whether regulatory variants could impart allelic bias at interaction loci, we
374 were interested to learn if we could find evidence that physical chromatin interactions can mediate the
375 effects of regulatory variants on gene expression. Expression quantitative trait loci (eQTLs) are genetic
376 loci at which variation correlates with expression of genes in a quantitative manner. Chromatin
377 interactions present a possible mechanism through which distal variants interact with and regulate
378 transcription for their targets genes. We first determined whether sets of significant eQTL-TSS pairs (q
379 < 0.05) for cortical, hippocampal, and cervical spinal cord tissues from GTEx V7[41] overlapped with our
380 significant chromatin interactions in a cell type-specific manner. We noticed a significant enrichment of
381 eQTL loci from matched tissues at PIRs (**Fig 6c**). Our results confirm that significant chromatin
382 interactions have the capacity to mediate quantitative expression differences by eQTL loci in a cell type-
383 specific manner.
384
385 **Discussion**

386 By generating the first cell type-specific *cis*-regulatory chromatin interaction maps in human neural
387 cells, we are able to provide a comprehensive annotation for previously uncharted regulatory
388 relationships between promoters and distal regulatory elements in the human CNS. We demonstrate that
389 PIRs are highly cell type-specific and directly linked to gene regulation. We also show that PIRs can be
390 used to gain insight into the mechanisms underlying cellular identity, as they are enriched for in
391 interactions with genes recapitulating key cell type-specific functions. Despite the similarity between the
392 closely related neural cell types used in this study, examination of cell type-specific distal open
393 chromatin regions at PIRs enables the intricate identification of TF-binding sites responsible for both the
394 specification and maintenance of cell identity and function.
395
396 GWAS SNPs for complex diseases have been challenging to interpret in the past due to the lack of
397 systematic annotation for noncoding regions. Promoter-PIR maps enable both the identification of novel
398 targets for regulatory sequences and the functional interpretation and prioritization of variants associated
399 with complex neurological disorders. We demonstrate through genome editing that PIRs can regulate
400 neurologically relevant genes with diverse functions in neuronal survival, synaptic activity, drug
401 response, and disease pathogenesis. We believe a large number of interesting candidate variants have
402 been reported in this study that would be interesting to validate in the future, and which may reveal
403 important disease biology. Finally, regulatory variants at PIRs linked to these genes may potentially
404 serve as therapeutic targets or biomarkers to aid in the development of novel treatments against
405 previously intractable complex neurological disorders. The epigenomic characterization of additional
406 cell types should continue to yield rich insights into the landscape of transcriptional regulation and
407 contribute towards the advancement of human health[42].
408
409 **Methods**
410 **Cell culture**
411 Human cortical excitatory neurons were generated using integrated, isogenic, and inducible neurogenin-
412 2 (Ngn2) iPSCs (i^3N iPSCs) with doxycycline-inducible mouse Ngn2 integrated at the AAVS1 safe-
413 harbor locus. The i^3N iPSC line has a well-characterized wild type genetic background (WTC11)[38]. A
414 simplified, two-step pre-differentiation and maturation protocol was used to generate the i^3Neurons[7].
415 Briefly, i^3N iPSCs were incubated with 2 μg/ml doxycycline in pre-differentiation media containing
416 KnockOut DMEM/F12 supplemented with 1x N-2, 1x NEAA, 1 μg/ml mouse laminin, 10 ng/ml BDNF,

417 and 10 ng/ml NT3. In addition, 10 μM Rock inhibitor was included in the pre-differentiation media for
418 the first day only. Media was changed daily for three days. For maturation, pre-differentiated i^3N
419 precursor cells were dissociated and subplated on poly-D-lysine/laminin plates in maturation media
420 containing equal parts DMEM/F12 and Neurobasal-A with 2 μg/ml doxycycline and supplemented with
421 0.5x B-27, 0.5x N-2, 1x NEAA, 0.5x GlutaMax, 1 μg/ml mouse laminin, 10 ng/ml BDNF, and 10 ng/ml
422 NT3. The doxycycline was omitted from all subsequent media changes. Half of the media was half
423 changed weekly over the first two weeks, then the amount of media was doubled on day 21. Thereafter,
424 a third of the media was replaced weekly.
425
426 Human hippocampal DG-like neurons were generated from dissociated hippocampal organoids
427 (unpublished). Briefly, human iPSC cells from the WTC11 line were grown on MEF feeder cells and
428 patterned towards a neural ectoderm fate using dual-SMAD inhibition as floating embryoid bodies
429 (EBs) in media containing 20% Knockout serum replacement. Next, EBs were patterned towards a
430 hippocampal fate using WNT and BMP in media containing N-2. After patterning, organoids were
431 dissociated using a neural dissociation kit (MiltenyiBiotech), plated on PDL/laminin coated plates, and
432 cultured for 4 weeks in media containing B-27, BDNF, GDNF, cAMP, and ascorbic acid.
433
434 WTC11 iPSCs with a doxycycline inducible transgene expressing NGN2, ISL1, and LHX3 integrated
435 into the AAVS1 locus were differentiated into motor neurons as previously reported[12]. Briefly, iPSCs
436 were maintained on growth factor-reduced Matrigel in StemFit media (Nacalai USA). On day 0, 1.5
437 $x10^6$ iPSCs were plated on 10cm dishes, followed 24 hours later by exchange into neural induction
438 media containing doxycycline and compound E. On day 3, the precursor cells were replated onto 12-
439 well plates coated with poly-D-lysine and laminin at a density of 2.5×10^5 cells per well. From day 3 to
440 4, the cells were treated with a pulse of 40 μM BrdU for 24 hours to suppress proliferation of
441 undifferentiated cells. Media was exchanged on day 4 and every three days thereafter until harvest.
442
443 Human primary fetal astrocytes (P0) were purchased from ScienCell Research Laboratories (Cat#1800)
444 and cultured using the prescribed media (Cat#1801). Briefly, cells were cultured in flasks coated with
445 poly-L-lysine ($2\mu/cm^2$) and passaged once with trypsin/EDTA before harvesting.
446
447 All cells used in the present study were verified as mycoplasma contamination free.

448

449 **Immunofluorescence**

450 Cells were fixed with 4% paraformaldehyde for 15 minutes at room temperature when mature (indicated
451 in cell culture section). Fixed cells were washed repeatedly with PBS containing 0.1% Triton X-100
452 (PBS-T), then blocked in a solution of 5% bovine serum albumin (BSA, Sigma-Aldrich) in PBS-T at
453 room temperature for 1 hour. Primary antibodies against Cux1 (ab54583, Abcam), MAP2 (ab5392,
454 Abcam), HB9 (ABN174, Millipore), SMI32 (ab7795, Abcam), PROX1 (ab101851, Abcam), β-Tubulin
455 (T8328, Sigma-Aldrich), GFAP (sc-9065, Santa Cruz Biotechnology), were diluted in 5% BSA solution
456 and incubated overnight at 4°C. Secondary antibodies, including Alexa Fluor 568 goat anti-chicken IgG,
457 Alexa Fluor 568 goat anti-mouse IgG, Alexa Fluor 488 donkey anti-rabbit IgG, Alexa Fluor 488 donkey
458 anti-mouse IgG (Molecular Probes), were diluted in 5% BSA solution and incubated for 1-2 hours at
459 room temperature. Images were acquired by Leica TCS SP8 confocal microscope with 40× oil-
460 immersion objective lens.

461 **Promoter capture Hi-C (pcHi-C)**

462 *In situ* Hi-C libraries for the cortical excitatory neurons, hippocampal DG-like neurons, and fetal
463 astrocytes were constructed from 1-2 million cells using HindIII as a restriction enzyme as previously
464 described[43]. pcHi-C was performed using biotinylated RNA probes prepared as previously described
465 (Jung et al., under review). Briefly, sets of 120 bp probes with 30 bp overhangs were designed to capture
466 the sequences adjacent to restriction sites flanking each promoter-containing HindIII fragment. Three
467 probes were used to target each side of a restriction site, for a total of 12 probes targeting each promoter-
468 containing HindIII fragment. In total, promoters (defined as the regions up to 500 bp upstream and 500
469 bp downstream of each transcription start site) for 19,603 of 20,332 protein coding genes in GENCODE
470 19 were captured using this approach. While noncoding RNA promoters were not explicitly baited by
471 this design, fragments containing 3,311 of 14,069 noncoding RNA promoters in GENCODE 19 were
472 also targeted by the probes.

473

474 To perform the hybridization, 500 ng of the finished *in situ* Hi-C libraries were first mixed with 2.5 μg
475 human Cot-1 DNA (Invitrogen #15279011), 2.5 μg salmon sperm DNA (Invitrogen #15632011), and
476 0.5 nmol each of the p5 and p7 IDT xGen Universal Blocking Oligos in a 10 μL volume, then denatured
477 at 95°C for 5 min before holding at 65°C. Next, a hybridization buffer mix was prepared by mixing 25

μL 20x SSPE, 1 μL 0.5 M EDTA, 10 μL 50x Denhardt's solution, and 13 μL 1% SDS and pre-warmed to 65°C. Finally, 500 ng of the probe mix was combined with 1 μL 20 U/μL SUPERase-In (Invitrogen #AM2696) in a 6 μL volume, pre-warmed to 65°C, and promptly mixed with the library and hybridization buffer mixes. The final solution was transferred to a humidified hybridization chamber and incubated for 24 hours at 65°C. 0.5 mg Dynabeads MyOne Streptavidin T1 (Invitrogen #65601) was 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, and the beads were washed once with 1x SSC and 0.1% SDS for 30 minutes at 25°C and three times with pre-warmed 0.1X SSC and 0.1% SDS at 65°C and 10 minutes per wash. The final library was eluted in 20 μL nuclease-free water and sent for paired-end sequencing on either the HiSeq 4000 (50 bp reads), HiSeq X Ten, or NovaSeq instruments (150 bp reads).

Detailed description of capture probe design and experimental procedures can be found in Jung et al. (under revision).

ATAC-seq

ATAC-seq was carried out as previously described using the Nextera DNA Library Prep Kit (Illumina #FC-121-1030)[44]. Cortical excitatory neurons and fetal astrocytes were first harvested using Accutase and rinsed twice with PBS. Following resuspension into nuclei extraction buffer (10 mM Tris-HCl pH 7.5, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% Igepal CA630, and 1x proteinase inhibitor) on ice, 50,000 cells were counted out, exchanged into 1x Buffer TD, and incubated with 2.5 μL TDE1 for 30 minutes at 37°C. The libraries were purified on Qiagen MinElute spin columns, amplified using Nextera primers, then size selected for fragments between 300 and 1000 bp using AMPure XP beads prior to being sent for single-end sequencing on the HiSeq 4000 (50 bp reads).

Sequencing reads were mapped to hg19 and processed using the ENCODE pipeline. For cell types with 3 or more biological replicates, IDR was used to threshold the peaks for downstream analysis. Overlapping peaks were used for cell types with 2 biological replicates. ATAC-seq peaks were expanded to a minimum width of 500 bp for all downstream analysis. All data processing metrics are reported in **Supplementary Table 1**.

RNA-seq

509 RNA from cortical excitatory neurons and fetal astrocytes was extracted using the RNeasy Mini Kit
510 (Qiagen #74104). Approximately 500 ng extracted RNA was used to construct libraries using the
511 TruSeq Stranded mRNA Library Prep Kit (Illumina #20020594). Libraries were sent for single-end
512 sequencing on the HiSeq 4000 (50 bp reads). Raw sequencing reads were aligned to hg19/GRCh37
513 using STAR running the standard ENCODE parameters, and transcript quantification was performed in
514 a strand-specific manner using RSEM based on the GENCODE 19 annotation. The edgeR package in R
515 was used to calculate TMM-normalized RPKM values for each gene based on the expected counts and
516 normalized gene lengths reported by RSEM. All data processing metrics are reported in **Supplementary**
517 **Table 1**.
518
519 **Validation of pcHi-C interaction loci using CRISPR/Cas9-mediated deletion**
520 To validate genomic interactions captured by pcHi-C, several putative regulatory loci interacting with
521 neurological gene targets were deleted in the human i^3N iPSC line using CRISPR/Cas9 genome-editing.
522 For each locus, we designed a pair of sgRNAs to delete the candidate enhancer region as localized by
523 open chromatin or enhancer signal. sgRNAs were synthesized by Synthego and Cas9 protein was
524 sourced from QB3-Berkeley. To generate deletion lines, CRISPR/Cas9 nucleofections were performed
525 using the LONZA Human Stem Cell Nucleofector® Kit. For each nucleofection, 500,000 i^3N iPSCs
526 were transfected with Cas9:sgRNA RNP complex (12 µg Cas9, 10 µg sgRNA 1, 10 µg sgRNA 2) using
527 program "A-023" on the LONZA 4D-Nucleofector. The cells were seeded onto Matrigel-coated six well
528 plates containing Essential 8™ Medium (ThermoFisher #A15169-01) with Y-27632 for recovery
529 following nucleofection. After 48 hours, the cells were split into new six well plates at a concentration
530 of approximately 50 cells per well for picking single colonies. Deletion clones picked from the six well
531 plates were confirmed by genotyping PCR, and the i^3N iPSC CRISPR deletion lines were differentiated
532 into cortical excitatory neurons for quantifying the expression of genes targeted by the putative enhancer
533 loci after two weeks by qPCR. Clones with WT genotypes were used as controls. To perform the
534 quantification, total RNA from the cortical excitatory neurons was extracted using the Qiagen AllPrep
535 DNA/RNA Mini Kit, and cDNA was synthesized using the Bio-RAD iScript™ cDNA Synthesis Kit.
536 qPCR was performed using FastStart Essential DNA Green Master (Roche) on the LightCycler® 96
537 System (Roche). For DRD2 PIR, additional primers were designed to flank a SNP region in targeted
538 region in order to determine the genotypes of deleted clones by Sanger sequencing. Primers flank the
539 SNP containing exon were also used to determine the relative expression levels of the two alleles by

direct Sanger Sequencing and TOPO cloning and sequencing. Detailed information about the sgRNA oligos, genotyping primers, and qPCR primers can be found in **Supplementary Table 7.**

Validation of pcHi-C interaction loci using CRISPRi

i3N iPSC induced neurons were infected with lentivirus carrying dCas9-KRAB-blast (Addgene #89567), and colonies with low integration but high dCas9 expression levels were picked out. The CROP-seq-opti vector (Addgene #106280) was used for sgRNA expression. sgRNA were designed (ordered from Integrated DNA Technologies) and cloned into vector as described previously. *[ref:Pooled CRISPR screening with single-cell transcriptome readout]* Validated assembly products from mini-prep (QIAGEN #27104) are co-transfected with lentivirus packaging plasmids pMD2.G (Addgene #12259) and psPAX (Addgene #12260) into 293T cells using PolyJet (SignaGen Laboratories #SL100688) according to manufacture's instructions. Virus containing media was collected for 72 hours, filtered by Millex-HV 0.45um (Millipore #SLHV033RS) and concentrated by Amicon Ultra (Millipore #UFC801024). Virus was stored in -80°C before infection. The virus was titrated by qPCR. Briefly, i3Neuron of 2 weeks are incubated with different volumes of virus for 12 hr. Genomic DNA was extracted 72 hr later. The internal control for qPCR is targeting an intronic region (Forward: TCCTCCGGAGTTATTCTTGGCA, Reverse: CCCCCCATCTGATCTGTTTCAC). Integration of WPRE fragment was quantified compared to a control cell line with a known copy number of WPRE. For CRISPRi validation on putative enhancers, 2-week-neurons was treated with lentivirus (MOI~3) containing sgRNA. After 12hr 2/3 of the media was changed for 3 times. 7 days after transfection, cells are collected for mRNA extraction. The expression of target genes are detected by qPCR.

Calling significant chromatin interactions

Paired-end sequencing reads were mapped to hg19, filtered, and deduplicated using the HiCUP pipeline (v0.71) running bowtie2 and filtering for ditags between 100 and 1200 bp[45]. Significant interactions were called using CHiCAGO, retaining baited fragments supported by at least 250 reads (minNPerBaits=250). Interactions between HindIII fragments with a score (defined as the negative log of the p-value) of 5 or greater in each cell type were defined to be significant chromatin interactions and used for additional downstream processing. All data processing metrics are reported in **Supplementary Table 1**.

571 To call overlaps between our sets of significant chromatin interactions and various genomic and
572 epigenomic features including promoters, chromatin signatures, and regulatory variants, interacting bins
573 were expanded to a minimum width of 5 kb, or kept as the size of the original fragments if their widths
574 exceeded 5 kb. Interactions involving fragments larger than 100 kb were omitted from our analysis due
575 to low resolution. Interactions were classified as promoter to promoter or promoter to other based on
576 overlap between their interacting ends and the set of all protein coding and noncoding RNA promoters
577 annotated in GENCODE 19. Finally, an interaction was considered to be shared between cell types if
578 both its ends intersected the corresponding ends of a significant interaction from another cell type. An
579 interaction was considered to be cell type-specific otherwise. To generate the interaction specificity
580 heatmap, each unique pair of significantly interacting loci was overlapped with the full set of
581 interactions for each cell type, and the maximum scoring overlapping interaction in each cell type was
582 used to populate the values in each column.
583
584 **Chromatin state enrichment analysis**
585 The annotations for the publicly available 15 state ChromHMM model were downloaded for the
586 dorsolateral prefrontal cortex, hippocampus, and normal human astrocytes. The states were merged as
587 follows: TssA and TssAFlnk were merged as TSS, TxFlnk, Tx, and TxWk were merged as Tx, EnhG
588 and EnhBiv were merged as other enhancer, and ReprPC and ReprPCWk were merged as ReprPC. All
589 other states (Enh, ZNF/Rpts, Het, TssBiv, and BivFlnk) were used as is. Enrichment analysis was
590 performed for each set of significant interactions by comparing the actual number of chromatin states
591 interacting with baited fragments against the mean number of chromatin states interacting with baited
592 fragments across 100 sets of bootstrapped interactions. To generate the bootstrapped interactions, the
593 baited fragments were left anchored while the interacting other ends were shuffled such that the
594 distribution of distances between the baited and other ends matched that of the original set.
595
596 **Motif enrichment analysis at distal ATAC-seq peaks**
597 We took the sets of all cell type-specific distal ATAC-seq peaks at PIRs and involved in cell promoter to
598 non-promoter interactions for each cell type, and used the sequences in 200 bp windows around the peak
599 summits to perform motif enrichment analysis using Homer. The entire genome was used as the
600 background. Enrichment significance values and expression values for each motif and its corresponding

601 TFs were reported for each cell type, with the union of the top 30 known motifs for each cell type
602 represented in the analysis.
603
604 **VISTA Enhancer Browser correspondence analysis**
605 Human enhancer regions and mouse enhancer regions with orthologous human sequences associated
606 with positive readouts in the VISTA Enhancer Browser were mined and analyzed for overlap with our
607 sets of significant interactions for each cell type. Of the 2,892 total tested elements reported in their
608 database, 2,845 were either originally from human or had orthologous human sequences, and 1,547 were
609 tested to be positive. Genes whose promoters were found to be interacting with positively tested
610 elements are reported in **Supplementary Table 3**. In addition, we determined the proportion of
611 positively tested elements that were found to interact with promoters of their nearest neighboring genes
612 as opposed to promoters of more distal genes, excluding cases where it was not feasible to resolve
613 interactions between the element and its nearest neighboring genes due to the features residing in the
614 same bin.
615
616 **GWAS SNP interaction analysis**
617 Tag SNPs for a total of eight neurological disorders including Alzheimer's disease (AD), attention
618 deficit hyperactivity disorder (ADHD), amytrophic lateral sclerosis (ALS), autism spectrum disorder
619 (ASD), bipolar disorder (BD), epilepsy (EP), frontotemporal dementia (FTD), Parkinson's disease (PD),
620 schizophrenia (SCZ), and unipolar depression (UD) were mined from the GWAS Catalog (December
621 2018) using a p-value threshold of 10^{-6}. Studies reporting variants associated with traits matching the
622 disorders as well as traits biologically associated with the disorders were included in our analysis. The
623 tag SNPs were imputed using HaploReg v4.1 at an LD threshold of 0.8 according to the reported study
624 population(s) for each SNP. The imputed SNPs were lifted over to hg19 and filtered for SNPs in
625 noncoding regions. See **Supplementary Table 4** for a detailed summary of the imputation process and
626 the list of studies used. Disease-specific enrichment for noncoding SNPs was calculated by taking the
627 ratio of the number of noncoding SNPs interacting with baited fragments in our sets of significant
628 interactions and the mean number of noncoding SNPs interacting with baited fragments in sets of
629 randomly shuffled interactions for each cell type, controlling for interaction distance.
630

631 To determine whether a tag SNP potentially interacted with a target gene, we determined whether any of
632 its promoters participated in a significant interaction with the tag SNP or any imputed SNPs in linkage
633 with the tag SNP on the other interacting end. Furthermore, we determined the proportion of tag SNPs
634 interacting with their nearest neighboring genes in the same way, except when the promoters for the
635 nearest gene and the SNPs of interest fell within the same bin and interactions could not be resolved.
636 Finally, we derived a list of interactions between imputed SNPs and target gene promoters for which the
637 imputed SNP was located within 1 kb of an ATAC-seq peak, indicating strengthened evidence for a
638 regulatory variant at that interacting locus.
639
640 **Phasing of the WTC11 genome**
641 The raw WTC11 genome sequence can be downloaded from http://genome.ucsc.edu/cgi-
642 bin/hgTracks?db=hg38&hubClear=https://s3-us-west-
643 2.amazonaws.com/downloads.allencell.org/genome-sequence/ucsc_hubs/WTC_genome_hub/hub.txt.
644 Phasing of the WTC11 genome was performed as previously described[39]. Briefly, WTC11 variants were
645 first split by chromosome and phase-informative pcHi-C reads extracted using extractHAIRS with the
646 minimum mapping quality set to 10 and the maximum insert size set to 30000000 bp[46]. Phasing was
647 performed using Hapcut with a maximum of 101 iterations. We next extracted the maximum variants
648 phased (MVP) haplotype block from the output of Hapcut to use as the seed haplotype. We modified the
649 "neighborhood correction" aspect of phasing by filtering phased variants whose predicted phase would
650 have a marginal probability below 0.99 using an in-house implementation of a phasing hidden Markov
651 model as described previously[47,48] using a reference panel of haplotypes from the 1000 Genomes
652 Project. Finally, missing variants were imputed using the same HMM using the reference haplotype set
653 from the 1000 Genomes Project.
654
655 The chromosome-wide SNP phasing data is available at the Gene Expression Omnibus under the
656 following accessible number: GSE113483
657
658 **Allelic bias analysis**
659 We utilized the phasing data for the WTC11 genome along with the allele specific mapping capabilities
660 of HiC-Pro to quantify genome-wide allelic bias between significantly interacting loci in the cortical
661 excitatory neurons. Briefly, reads were mapped using bowtie to a version of the hg19 reference genome

where all sites containing heterozygous phased SNPs were first N-masked. Next, nucleotides at the masked polymorphic sites were used to assign the reads to either allele, with reads containing conflicting allele assignments or unexpected bases omitted from further analysis. Read pairs with at least one allele specific mate were used to construct allele specific Hi-C contact maps at 10 kb resolution, for which interacting bins overlapping with the set of significant pcHi-C interactions with score >= 5 in WTC11 cell types (cortical excitatory neurons, hippocampal DG-like neurons, and lower motor neurons) was used to detect bias. The binomial test was performed to assess allelic bias for each set of 10 kb interacting bins, and the resulting p-values were adjusted using the BH method to filter out significantly biased loci with an FDR cutoff of 0.05.

Data availability statement

All data sets used in this study (pcHi-C, ATAC-seq, RNA-seq, and chromosome-wide SNP phasing data) are available at the Gene Expression Omnibus under the accession number GSE113483.

Acknowledgements

We thank Anthony Schmitt and Bing Ren for sharing promoter capture probes and the pcHi-C protocol. The genome analysis of the WTC11 line provided in this publication was made possible by the whole genome sequencing data generated by the Allen Institute for Cell Science. We thank the Institute and their founder Paul G. Allen for making this work possible. We also thank G. Hon for providing reagent, N. Ahituv, R.D. Hawkins, M. McManus, and B. Ren for providing critical feedback on the manuscript. This work was made possible in part, by NIH-NEI EY002162 - Core Grant for Vision Research and the Research to Prevent Blindness Unrestricted Grant. This work was supported by the US National Institute of Health (NIH) grants R01AG057497 to YS, LG, and HS, UM1HG009402 to YS, 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 YS. P01NS097206 and U19MH106434 to HS. R01MH105128, R35NS097370, and U19AI131130 to GLM. MS is supported by T32GM007175. FJ is supported by a NIH training grant T32GM73009

Author contribution

693 MS and YS designed the study. MS, XY, XR, LM, IJ, KJ and TWT performed the experiments. MS,
694 MT, and LW performed data analysis. JD contributed to genomic phasing using HaploSeq. SL, JY, KW,
695 BC, FJ, GLM, HS, LD, CW, and LG contributed by providing cell samples. MS and YS prepared the
696 manuscript with assistance from all authors.
697
698 **Competing interests**
699 The authors declare no competing financial interests.
700
701 **Figure legends**
702 **Figure 1. Genome-wide mapping of physical chromatin interactions in functionally distinct brain**
703 **cell types.**
704 (**a**) Schematic of the study design for generating four functionally distinct neural cell types and
705 performing integrative analysis of chromatin interactions using promoter capture Hi-C, open chromatin
706 regions using ATAC-seq, and transcriptomes using RNA-seq. (**b**) Proportions of promoters and PIRs
707 interacting within TADs for each cell type. (**c**) Histogram of chromatin interaction distances across all
708 cell types and empirical cumulative distribution function plots of chromatin interaction distances for
709 each cell type. (**d**) Counts of promoter to promoter interactions (blue) and promoter to other interactions
710 (purple) for each cell type. (**e**) Percentages of cell type-specific and shared distal ATAC-seq peaks at
711 PIRs for each cell type.
712
713 **Figure 2. Integrative analysis of epigenomic features, chromatin interactions, and gene expression.**
714 (**a**) Histograms of the number of PIRs interacting with each annotated promoter with at least one
715 interaction for cortical excitatory neurons, hippocampal DG-like neurons, and fetal astrocytes. (**b)**
716 Enrichment of active chromatin states and depletion of repressive states as inferred by ChromHMM on
717 the H3K4me1, H3K4me3, H3K9me3, H3K27me3, and H3K36me3 histone modifications in matched
718 tissue types at significant PIRs (solid bars) versus shuffled PIRs (striped bars) for each cell type. Error
719 bars show the standard deviation. (**c**) Comparative global expression analysis for promoters interacting
720 with enhancer-PIRs versus repressive-PIRs. (**d**) Boxplot of the mean normalized expression (RPKM) for
721 genes with promoters grouped according to the number of enhancer-PIRs they are interacting with.
722 Linear regression was performed on the means for each bin.
723

724 **Figure 3. Cell type-specific PIRs and TF motif enrichment analysis.**
725 (**a**) Classification of cell type-specific promoter-PIR interactions into specificity categories based on
726 their interaction scores in cortical excitatory neurons, hippocampal DG-like neurons, lower motor
727 neurons, and fetal astrocytes. Cell type-specific distal open chromatin peaks in these interactions are
728 used for TF motif enrichment analysis. The number of promoter-PIR interactions in each category is
729 displayed in **Supplementary Fig. 3a**. (**b**) Enrichment of consensus motif sequences for TFs in cell type-
730 specific open chromatin regions at PIRs and relative expression levels of the corresponding TFs. The
731 size of each circle represents the normalized expression of the TF corresponding to that motif (RPKM),
732 and the color of each circle represents the degree of enrichment for that motif in each cell type.
733
734 **Figure 4. Validation of promoter-PIR interactions.**
735 (**a**) Number of *in vivo* validated enhancers from the VISTA Enhancer Browser overlapped by promoter-
736 PIR pairs in this study. 532 or 34.4% of all functional human enhancer elements are recovered by our
737 interactions. In addition, the numbers of neural and non-neural enhancers are shown. (**b**) Numbers of *in*
738 *vivo* validated enhancers interacting with their nearest neighboring genes (center) versus those
739 interacting exclusively with more distal genes (left). The total number of non-neighboring gene targets
740 across all interacting elements is also shown (right). (**c**) Functional validation of a predicted enhancer
741 element (mm999) and its target gene by CRISPR deletion. Top panel contains a WashU Epigenome
742 Browser[49] snapshot of the genomic locus for *CDK5RAP3* and significant chromatin interactions between
743 the promoter of *CDK5RAP3* and two downstream PIRs with open chromatin regions in cortical
744 excitatory neurons. Bottom panel contains CRISPRi targeting results for an open chromatin region 30 kb
745 downstream from the *CDK5RAP3* promoter (yellow highlight) within the larger 6 kb enhancer element
746 from the VISTA Enhancer Browser (blue highlight). Significant down-regulation of *CDK5RAP5* was
747 observed using two independent sgRNAs in cortical excitatory neurons. Expression levels for
748 neighboring genes (*MRPL10*, *PNPO*, and *NFE2L1*) are also shown. Each sgRNA CRISPRi experiment
749 was performed three times, and each time with three technical replicates (* p=0.02, **** $p<10^{-4}$,
750 Student's t-test).
751
752 **Figure 5. Genetic analysis of chromatin interactions with complex neurological disorder-**
753 **associated SNPs.**

754 (**a**) Enrichment analysis for noncoding imputed SNPs from ten complex neurological diseases in
755 significant versus shuffled PIRs. Disorders included disorders including Alzheimer's disease (AD),
756 attention deficit hyperactivity disorder (ADHD), amyotrophic lateral sclerosis (ALS), autism spectrum
757 disorder (ASD), bipolar disorder (BP), epilepsy (EP), frontotemporal dementia (FTD), Parkinson's
758 disease (PD), schizophrenia (SCZ), and unipolar depression (UD). (**b**) Proportions and total counts of
759 noncoding tag SNPs with at least one SNP in LD participating in long-range chromatin interactions in
760 cortical excitatory neurons (blue), hippocampal DG-like neurons (orange), motor neuron (yellow), fetal
761 astrocytes (green), or more than one cell type (grey). (**c**) Numbers of noncoding tag SNPs with at least
762 one SNP in LD interacting with their nearest neighboring genes (center) versus those interacting
763 exclusively with more distal genes (left). The total number of non-neighboring gene targets across all
764 noncoding tag SNPs and diseases is also shown (right). (**d**) A PIR containing a region of open chromatin
765 and a cluster of SNPS associated with BP in an intron for *PTPRO* interacts with the promoter of the
766 *STRAP* gene over 300 kb away (highlighted in yellow, red arch highlight the interactions between
767 *STRAP* promoter and CRISPR deleted region). Homozygous deletion of this region in three independent
768 clones in cortical excitatory neurons resulting in significant down-regulation of *STRAP* (right panel).
769 Each bar represents the mean ± s.e.m. (**e**) A PIR containing open chromatin regions and clusters of SNPs
770 associated with SCZ interact with the promoter of *DRD2* in cortical excitatory neurons (left panel).
771 Mono-allelic deletion of this region using CRISPR results in significant down-regulation of *DRD2* in
772 cortical excitatory neurons (right panel). Each bar represents the mean ± s.e.m.
773
774 **Figure 6. Genetics variants contribute to chromatin interaction bias and alterations in gene**
775 **expression.**
776 (**a**) Quantile-quantile plot of allelic bias for chromatin interactions in cortical excitatory neurons. (**b**) An
777 example of an allelic biased interaction between a PIR containing a cluster of SNPs associated with BD
778 (orange highlight) and the promoter of *SYT17* (green highlight). Heterozygous WTC11 SNPs at the PIR
779 are shown, along with a bar graph of their detected read counts in our pcHi-C datasets. (**c**) Enrichment of
780 eQTL loci at significant versus shuffled PIRs in matched tissue types for cortical excitatory neurons,
781 hippocampal DG-like neurons, and motor neurons. Error bars show the standard deviation.
782
783 **Supplementary Figure 1. Identifying PIRs in human brain cells.** (**a**) Immunofluorescence staining
784 showing the expression of cortical markers CUX1 and MAP2 in cortical excitatory neurons. The yield of

785 mature cortical neurons is calculated as the percentage of CUX1 and MAP2 double positive cells in total
786 live cells. Immunofluorescence staining demonstrating that lower motor neurons express HB9, a motor
787 neuron identity marker, and pan-neuronal neurofilament marker SMI32. The yield of mature motor
788 neurons is calculated as the percentage of HB9 and SMI32 double positive cells in total live cells.
789 Immunofluorescence staining showing that hippocampal DG-like neurons express PROX1, a
790 transcription factor constitutively expressed in the hippocampal DG cell lineage. The yield of mature
791 hippocampal neurons is calculated as the percentage of PROX1 and MAP2 double positive cells in total
792 live cells. Immunofluorescence staining showing that astrocytes express GFAP. The yield is calculated
793 as the percentage of GFAP positive cells in total live cells. Data are mean ± sem (N and total cell
794 number are indicated). (**c**) The fetal astrocytes used in this study exhibit an expression profile consistent
795 with their identity as astrocyte progenitor cells. Heatmap displays normalized gene expression values for
796 astrocyte progenitor markers (highly expressed) versus mature astrocytes markers (lowly expressed). (**c**)
797 Numbers of coding and noncoding promoters participating in chromatin interactions for each cell type.
798
799 **Supplementary Figure 2. Correlation between chromatin states at PIRs and gene targets**
800 **expression.** (**a**) Histograms of the number of significant chromatin interactions per gene. The average
801 number of significant interactions for each cell type are marked with a red line. (**b**) Comparative global
802 expression analysis for promoters interacting with enhancer-PIRs versus repressive-PIRs for each cell
803 type. (**c**) Boxplot of the mean normalized expression (RPKM) for promoters grouped according to the
804 number of enhancer-PIRs they are interacting with for each cell type.
805
806 **Supplementary Figure 3. Promoter long-range chromatin interactions are cell type-specific.** (**a**)
807 Venn diagram displaying numbers of overlapping promoter-PIR interaction loci that are significant in
808 cortical excitatory neurons, hippocampal DG-like neurons, lower motor neurons, and fetal astrocytes
809 (groups 1-15 in **Fig. 3a**). (**b**) Examples of cell type-specific interactions between PIRs and the promoters
810 of *GRIN2B* and *CHAT* (orange highlight). (**c**) Significant down-regulation of *CDK5RAP5* was observed
811 in three independent homozygous deletion clones in cortical excitatory neurons. Each bar represents the
812 mean ± s.e.m.
813
814 **Supplementary Figure 4. Using chromatin interaction to elucidate the functions of GWAS SNPs.**
815 (**a**) Numbers of noncoding tag SNPs with at least one SNP in LD interacting with their nearest

816 neighboring genes (center) versus those interacting exclusively with more distal genes (left) for each
817 disease. The total number of non-neighboring gene targets is also shown (right) for each disease. (**b**)
818 Significant interactions in cortical excitatory neurons and hippocampal DG-like neurons detected via
819 pcHi-C recapitulate a previously reported interaction between the *FOXG1* promoter and a distal ATAC-
820 seq peak containing a schizophrenia SNP[6]. (**c**) qPCR results for *STRAP* and its neighboring genes
821 *PTPRO*, *MGST1*, *WBP11* with CRISPRi targeting the PIR interacting with the *STRAP* promoter. Each
822 CRISPRi experiment was performed three times, and each time with three technical replicates (***$p<10^{-}$
823 ^{3}, **** $p<10^{-4}$, Student's t-test). (**d**) Schematic of detected genotypes in the *DRD2* gene and its deleted
824 regulatory PIR in mono-allelic deletion clones. Genotyping and RT-PCR sequencing of WTC11 variants
825 in *DRD2* reveal biased expression between the two alleles for the *DRD2* gene, corresponding to the
826 deletion of the PIR in one of the alleles. Results for the two control clones and three deletion clones are
827 shown for comparison. Each bar represents the mean ± s.e.m.
828
829 **Tables**
830 **Supplementary Table 1. pcHi-C, ATAC-seq, and RNA-seq data processing metrics.**
831 **Supplementary Table 2. Significant interactions called by CHiCAGO.**
832 **Supplementary Table 3. Putative target genes for in vivo validated enhancer elements.**
833 **Supplementary Table 4. GWAS SNP imputation summary.**
834 **Supplementary Table 5. Putative target genes for neurological disorder-associated SNPs.**
835 **Supplementary Table 6. Sequence and primer information for the CRISPR experiments**
836
837 **Reference:**
838 1 Mumbach, M. R. *et al.* Enhancer connectome in primary human cells identifies target genes
839 of disease-associated DNA elements. *Nat Genet* **49**, 1602-1612, doi:10.1038/ng.3963
840 (2017).
841 2 Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex
842 trait gene targets. *Nat Genet* **48**, 481-487, doi:10.1038/ng.3538 (2016).
843 3 Smemo, S. *et al.* Obesity-associated variants within FTO form long-range functional
844 connections with IRX3. *Nature* **507**, 371-375, doi:10.1038/nature13138 (2014).
845 4 Claussnitzer, M. *et al.* FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. *N*
846 *Engl J Med* **373**, 895-907, doi:10.1056/NEJMoa1502214 (2015).
847 5 de la Torre-Ubieta, L. *et al.* The Dynamic Landscape of Open Chromatin during Human
848 Cortical Neurogenesis. *Cell* **172**, 289-304 e218, doi:10.1016/j.cell.2017.12.014 (2018).
849 6 Won, H. *et al.* Chromosome conformation elucidates regulatory relationships in developing
850 human brain. *Nature* **538**, 523-527, doi:10.1038/nature19847 (2016).

851 7 Wang, C. *et al.* Scalable Production of iPSC-Derived Human Neurons to Identify Tau-
852 Lowering Compounds by High-Content Screening. *Stem Cell Reports* **9**, 1221-1233,
853 doi:10.1016/j.stemcr.2017.08.019 (2017).
854 8 Mertens, J. *et al.* Differential responses to lithium in hyperexcitable neurons from patients
855 with bipolar disorder. *Nature* **527**, 95-99, doi:10.1038/nature15526 (2015).
856 9 Visel, A., Minovitsky, S., Dubchak, I. & Pennacchio, L. A. VISTA Enhancer Browser--a
857 database of tissue-specific human enhancers. *Nucleic Acids Res* **35**, D88-92,
858 doi:10.1093/nar/gkl822 (2007).
859 10 Zhang, Y. *et al.* Rapid single-step induction of functional neurons from human pluripotent
860 stem cells. *Neuron* **78**, 785-798, doi:10.1016/j.neuron.2013.05.029 (2013).
861 11 Yu, D. X. *et al.* Modeling hippocampal neurogenesis using human pluripotent stem cells.
862 *Stem Cell Reports* **2**, 295-310, doi:10.1016/j.stemcr.2014.01.009 (2014).
863 12 Fernandopulle, M. S. *et al.* Transcription Factor-Mediated Differentiation of Human iPSCs
864 into Neurons. *Curr Protoc Cell Biol* **79**, e51, doi:10.1002/cpcb.51 (2018).
865 13 Zhang, Y. *et al.* Purification and Characterization of Progenitor and Mature Human
866 Astrocytes Reveals Transcriptional and Functional Differences with Mouse. *Neuron* **89**, 37-
867 53, doi:10.1016/j.neuron.2015.11.013 (2016).
868 14 Cairns, J. *et al.* CHiCAGO: robust detection of DNA looping interactions in Capture Hi-C data.
869 *Genome Biol* **17**, 127, doi:10.1186/s13059-016-0992-2 (2016).
870 15 Jackson, D. A., Hassan, A. B., Errington, R. J. & Cook, P. R. Visualization of focal sites of
871 transcription within human nuclei. *EMBO J* **12**, 1059-1065 (1993).
872 16 Zhang, Y. *et al.* Chromatin connectivity maps reveal dynamic promoter-enhancer long-
873 range associations. *Nature* **504**, 306-310, doi:10.1038/nature12716 (2013).
874 17 Diao, Y. *et al.* A tiling-deletion-based genetic screen for cis-regulatory element identification
875 in mammalian cells. *Nat Methods* **14**, 629-635, doi:10.1038/nmeth.4264 (2017).
876 18 Engreitz, J. M. *et al.* Local regulation of gene expression by lncRNA promoters, transcription
877 and splicing. *Nature* **539**, 452-455, doi:10.1038/nature20149 (2016).
878 19 Consortium, E. P. An integrated encyclopedia of DNA elements in the human genome.
879 *Nature* **489**, 57-74, doi:10.1038/nature11247 (2012).
880 20 Shen, Y. *et al.* A map of the cis-regulatory sequences in the mouse genome. *Nature* **488**,
881 116-120, doi:10.1038/nature11243 (2012).
882 21 Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and
883 characterization. *Nat Methods* **9**, 215-216, doi:10.1038/nmeth.1906 (2012).
884 22 Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human epigenomes.
885 *Nature* **518**, 317-330, doi:10.1038/nature14248 (2015).
886 23 Endele, S. *et al.* Mutations in GRIN2A and GRIN2B encoding regulatory subunits of NMDA
887 receptors cause variable neurodevelopmental phenotypes. *Nat Genet* **42**, 1021-1026,
888 doi:10.1038/ng.677 (2010).
889 24 Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-
890 regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-589,
891 doi:10.1016/j.molcel.2010.05.004 (2010).
892 25 Ren, G. *et al.* CTCF-Mediated Enhancer-Promoter Interaction Is a Critical Regulator of Cell-
893 to-Cell Variation of Gene Expression. *Mol Cell* **67**, 1049-1058 e1046,
894 doi:10.1016/j.molcel.2017.08.026 (2017).
895 26 Guo, Y. *et al.* CRISPR Inversion of CTCF Sites Alters Genome Topology and
896 Enhancer/Promoter Function. *Cell* **162**, 900-910, doi:10.1016/j.cell.2015.07.038 (2015).

897 27 Handoko, L. *et al.* CTCF-mediated functional chromatin interactome in pluripotent cells. *Nat*
898 *Genet* **43**, 630-638, doi:10.1038/ng.857 (2011).
899 28 Hou, C., Dale, R. & Dean, A. Cell type specificity of chromatin organization mediated by CTCF
900 and cohesin. *Proc Natl Acad Sci U S A* **107**, 3651-3656, doi:10.1073/pnas.0912087107
901 (2010).
902 29 Liddell, J. R. Are Astrocytes the Predominant Cell Type for Activation of Nrf2 in Aging and
903 Neurodegeneration? *Antioxidants (Basel)* **6**, doi:10.3390/antiox6030065 (2017).
904 30 Yin, X., Warner, D. R., Roberts, E. A., Pisano, M. M. & Greene, R. M. Novel interaction between
905 nuclear co-activator CBP and the CDK5 activator binding protein - C53. *Int J Mol Med* **16**,
906 251-256 (2005).
907 31 Xie, Y. H. *et al.* Cloning and characterization of human IC53-2, a novel CDK5 activator
908 binding protein. *Cell Res* **13**, 83-91, doi:10.1038/sj.cr.7290153 (2003).
909 32 MacArthur, J. *et al.* The new NHGRI-EBI Catalog of published genome-wide association
910 studies (GWAS Catalog). *Nucleic Acids Res* **45**, D896-D901, doi:10.1093/nar/gkw1133
911 (2017).
912 33 Ward, L. D. & Kellis, M. HaploReg: a resource for exploring chromatin states, conservation,
913 and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res*
914 **40**, D930-934, doi:10.1093/nar/gkr917 (2012).
915 34 Booth, H. D. E., Hirst, W. D. & Wade-Martins, R. The Role of Astrocyte Dysfunction in
916 Parkinson's Disease Pathogenesis. *Trends Neurosci* **40**, 358-370,
917 doi:10.1016/j.tins.2017.04.001 (2017).
918 35 Burghes, A. H. & Beattie, C. E. Spinal muscular atrophy: why do low levels of survival motor
919 neuron protein make motor neurons sick? *Nat Rev Neurosci* **10**, 597-609,
920 doi:10.1038/nrn2670 (2009).
921 36 Zhang, J. P. *et al.* Association of a Schizophrenia Risk Variant at the DRD2 Locus With
922 Antipsychotic Treatment Response in First-Episode Psychosis. *Schizophr Bull* **41**, 1248-
923 1255, doi:10.1093/schbul/sbv116 (2015).
924 37 Pearlson, G. D. *et al.* In vivo D2 dopamine receptor density in psychotic and nonpsychotic
925 patients with bipolar disorder. *Arch Gen Psychiatry* **52**, 471-477 (1995).
926 38 Miyaoka, Y. *et al.* Isolation of single-base genome-edited human iPS cells without antibiotic
927 selection. *Nat Methods* **11**, 291-293, doi:10.1038/nmeth.2840 (2014).
928 39 Selvaraj, S., J, R. D., Bansal, V. & Ren, B. Whole-genome haplotype reconstruction using
929 proximity-ligation and shotgun sequencing. *Nat Biotechnol* **31**, 1111-1118,
930 doi:10.1038/nbt.2728 (2013).
931 40 Sudhof, T. C. Calcium control of neurotransmitter release. *Cold Spring Harb Perspect Biol* **4**,
932 a011353, doi:10.1101/cshperspect.a011353 (2012).
933 41 Consortium, G. T. *et al.* Genetic effects on gene expression across human tissues. *Nature*
934 **550**, 204-213, doi:10.1038/nature24277 (2017).
935 42 Straws in a haystack. *Nat Genet* **50**, 631, doi:10.1038/s41588-018-0125-9 (2018).
936 43 Rao, S. S. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of
937 chromatin looping. *Cell* **159**, 1665-1680, doi:10.1016/j.cell.2014.11.021 (2014).
938 44 Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of
939 native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-
940 binding proteins and nucleosome position. *Nat Methods* **10**, 1213-1218,
941 doi:10.1038/nmeth.2688 (2013).

45 Wingett, S. *et al.* HiCUP: pipeline for mapping and processing Hi-C data. *F1000Res* **4**, 1310, doi:10.12688/f1000research.7334.1 (2015).

46 Bansal, V. & Bafna, V. HapCUT: an efficient and accurate algorithm for the haplotype assembly problem. *Bioinformatics* **24**, i153-159, doi:10.1093/bioinformatics/btn298 (2008).

47 Li, N. & Stephens, M. Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics* **165**, 2213-2233 (2003).

48 Delaneau, O., Zagury, J. F. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods* **10**, 5-6, doi:10.1038/nmeth.2307 (2013).

49 Zhou, X. *et al.* Exploring long-range genome interactions using the WashU Epigenome Browser. *Nat Methods* **10**, 375-376, doi:10.1038/nmeth.2440 (2013).

Figure 1

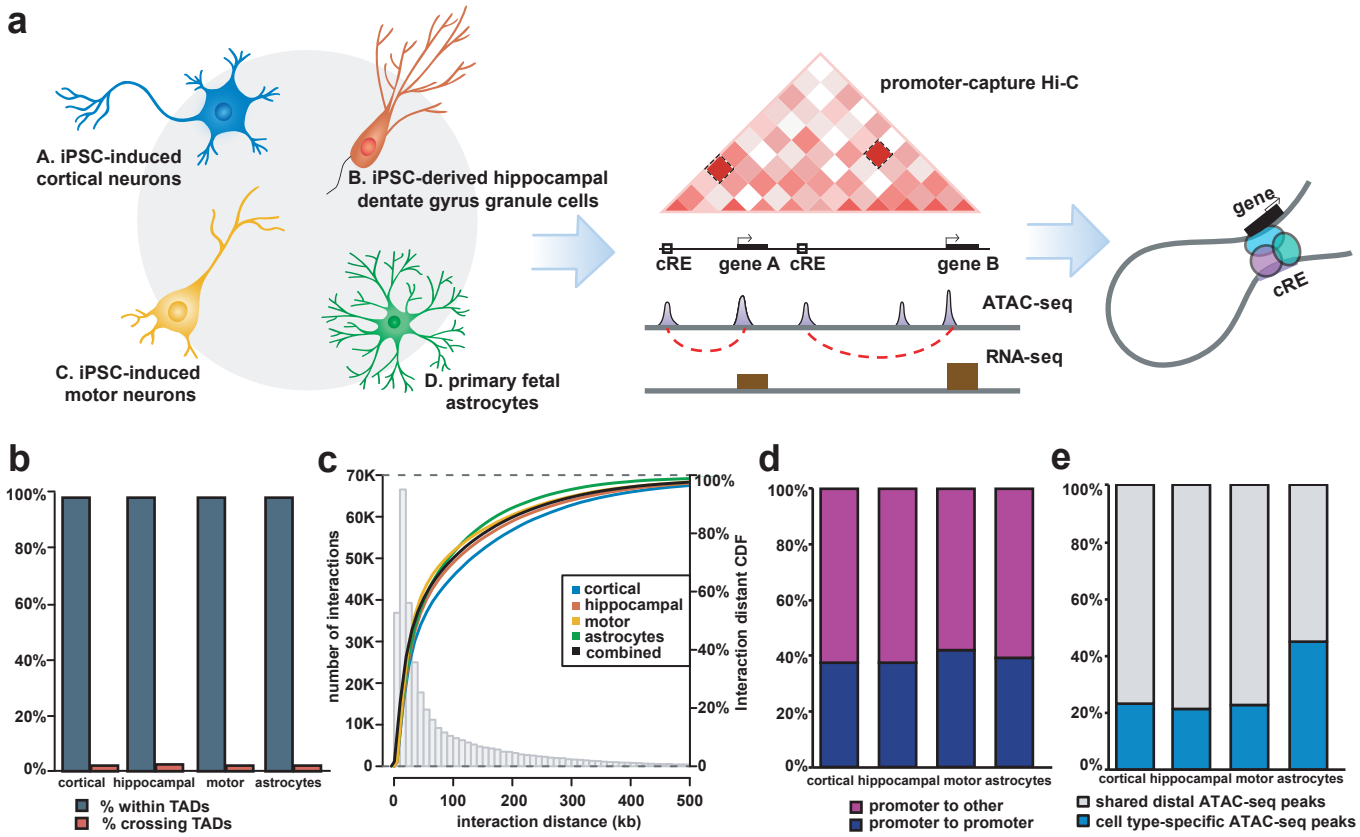

a
A. iPSC-induced cortical neurons
B. iPSC-derived hippocampal dentate gyrus granule cells
C. iPSC-induced motor neurons
D. primary fetal astrocytes
promoter-capture Hi-C
cRE
gene A
cRE
gene B
ATAC-seq
RNA-seq
gene
cRE
b
100%
80%
60%
40%
20%
0%
cortical
hippocampal
motor
astrocytes
% within TADs
% crossing TADs
c
70K
60K
50K
40K
30K
20K
10K
0
number of interactions
100%
80%
60%
40%
20%
0
Interaction distant CDF
cortical
hippocampal
motor
astrocytes
combined
0
100
200
300
400
500
interaction distance (kb)
d
100%
80%
60%
40%
20%
0%
cortical
hippocampal
motor
astrocytes
promoter to other
promoter to promoter
e
100%
80%
60%
40%
20%
0%
cortical
hippocampal
motor
astrocytes
shared distal ATAC-seq peaks
cell type-specific ATAC-seq peaks

Figure 2

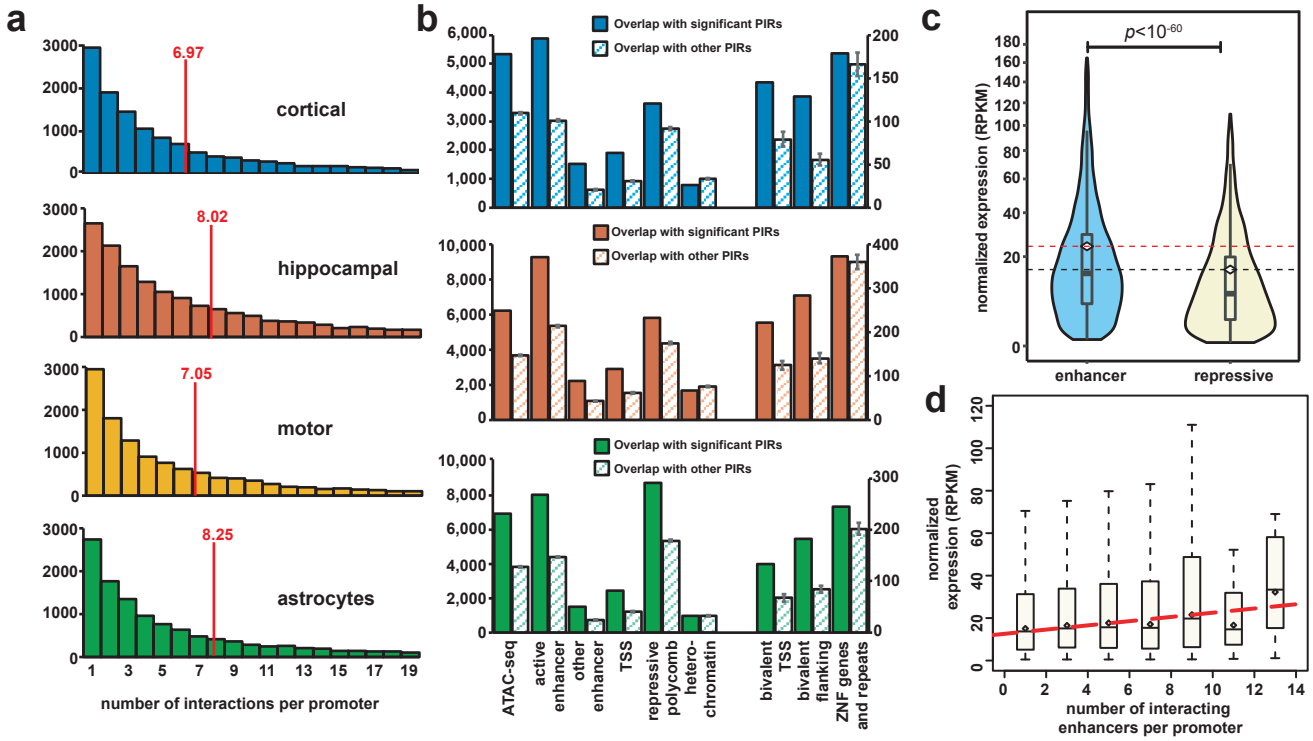

a
cortical
6.97
hippocampal
8.02
motor
7.05
astrocytes
8.25
number of interactions per promoter
b
Overlap with significant PIRs
Overlap with other PIRs
ATAC-seq
active enhancer
other enhancer
TSS
repressive
polycomb
hetero-chromatin
bivalent TSS
bivalent flanking
ZNF genes and repeats
c
p<10-60
normalized expression (RPKM)
enhancer
repressive
d
normalized expression (RPKM)
number of interacting enhancers per promoter

Figure 3

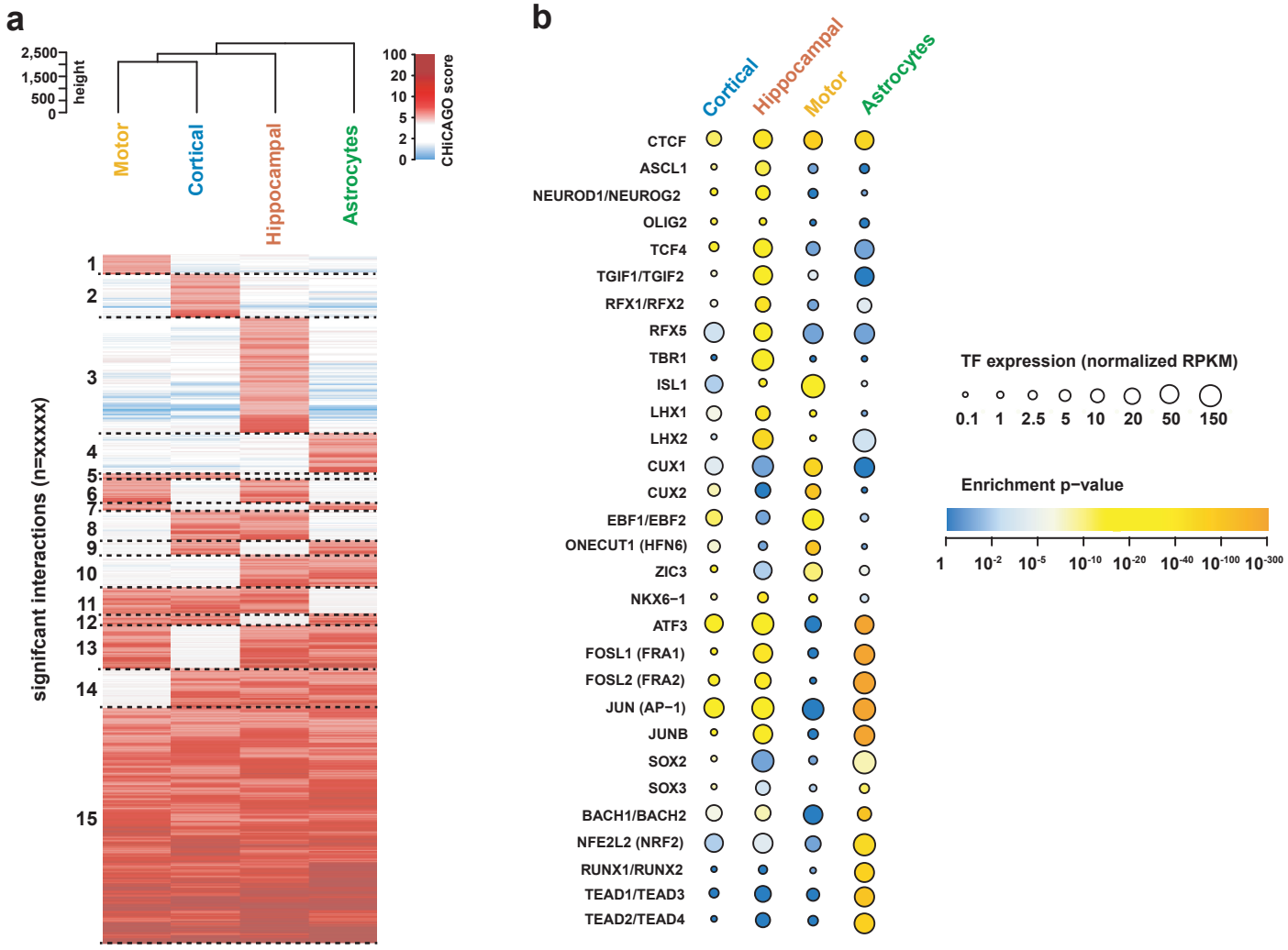

a
2,500
1,500
500
0
height
Motor
Cortical
Hippocampal
Astrocytes
100
20
10
5
2
0
CHiCAGO score
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
signifcant interactions (n=xxxxx)
b
Cortical
Hippocampal
Motor
Astrocytes
CTCF
ASCL1
NEUROD1/NEUROG2
OLIG2
TCF4
TGIF1/TGIF2
RFX1/RFX2
RFX5
TBR1
ISL1
LHX1
LHX2
CUX1
CUX2
EBF1/EBF2
ONECUT1 (HFN6)
ZIC3
NKX6−1
ATF3
FOSL1 (FRA1)
FOSL2 (FRA2)
JUN (AP−1)
JUNB
SOX2
SOX3
BACH1/BACH2
NFE2L2 (NRF2)
RUNX1/RUNX2
TEAD1/TEAD3
TEAD2/TEAD4
TF expression (normalized RPKM)
0.1
1
2.5
5
10
20
50
150
Enrichment p−value
1
10-2
10-5
10-10
10-20
10-40
10-100
10-300

Figure 4

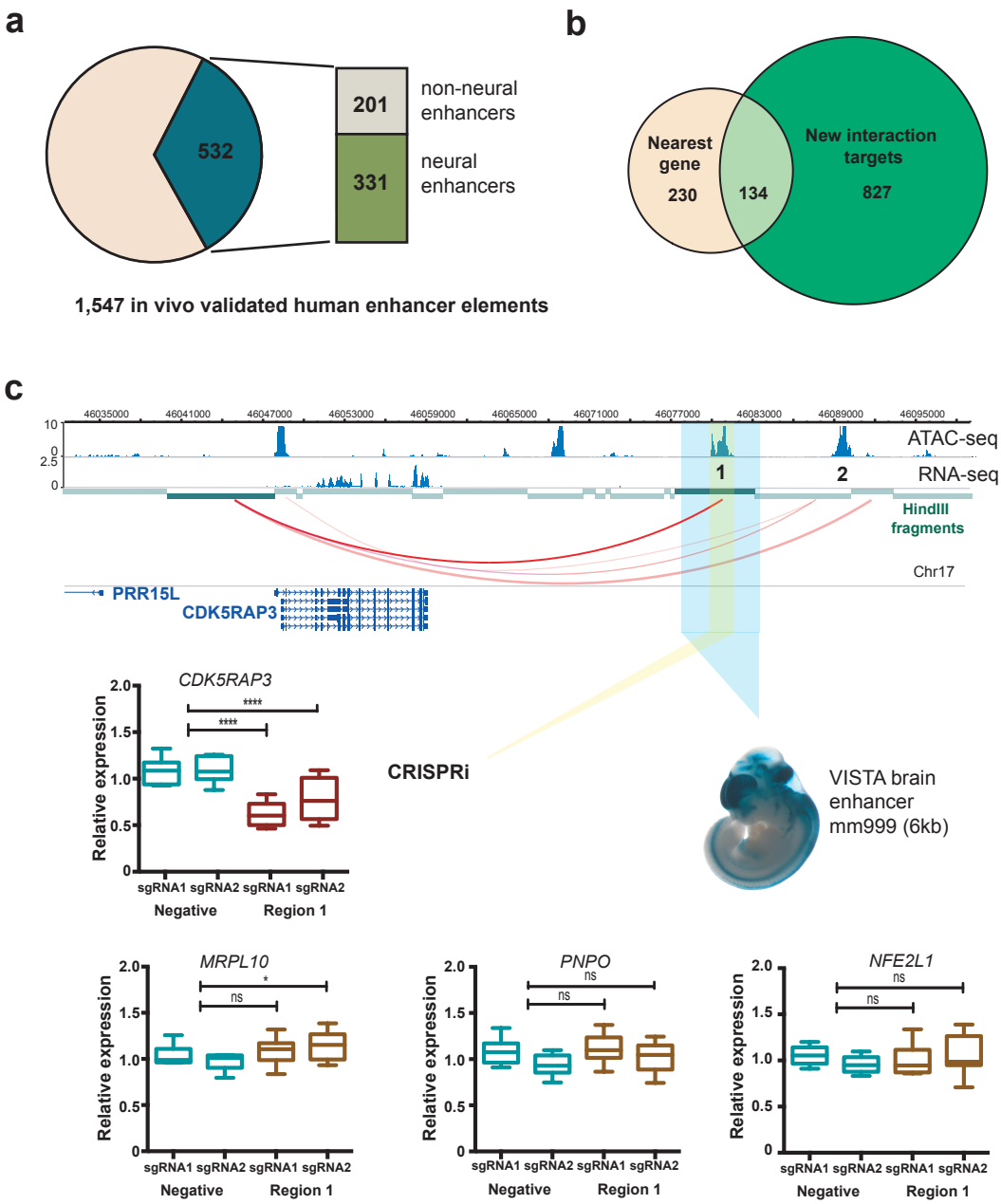

a
532
201
non-neural enhancers
331
neural enhancers
1,547 in vivo validated human enhancer elements
b
Nearest gene
230
134
New interaction targets
827
c
ATAC-seq
RNA-seq
1
2
HindIII fragments
Chr17
PRR15L
CDK5RAP3
CDK5RAP3
Relative expression
sgRNA1 sgRNA2 sgRNA1 sgRNA2
Negative
Region 1
CRISPRi
VISTA brain enhancer mm999 (6kb)
MRPL10
PNPO
NFE2L1
ns
*
****

Figure 5

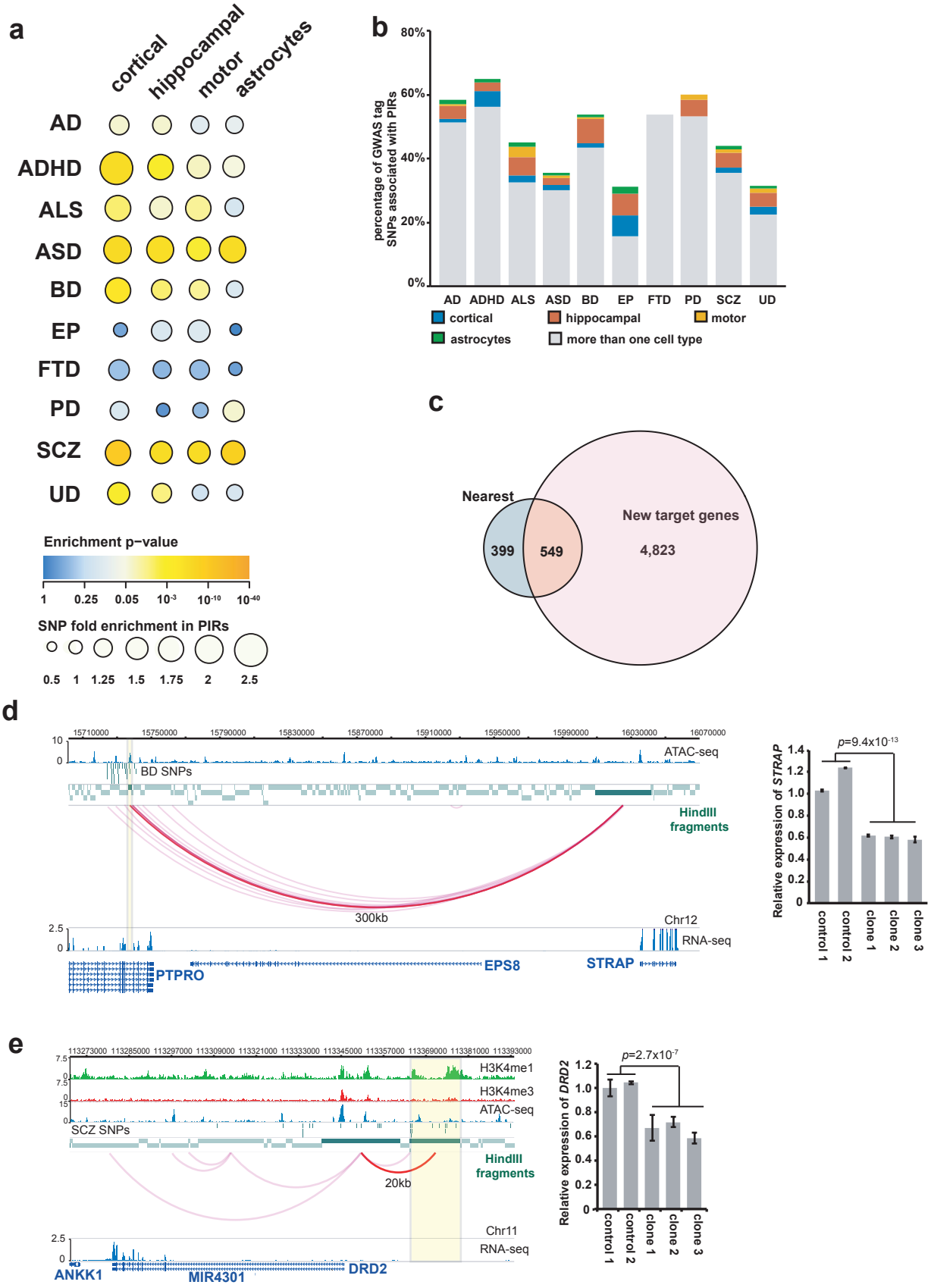

a
cortical
hippocampal
motor
astrocytes
AD
ADHD
ALS
ASD
BD
EP
FTD
PD
SCZ
UD
Enrichment p−value
1
0.25
0.05
10-3
10-10
10-40
SNP fold enrichment in PIRs
0.5
1
1.25
1.5
1.75
2
2.5
b
percentage of GWAS tag SNPs associated with PIRs
80%
60%
40%
20%
0%
AD
ADHD
ALS
ASD
BD
EP
FTD
PD
SCZ
UD
cortical
hippocampal
motor
astrocytes
more than one cell type
c
Nearest
399
549
New target genes
4,823
d
ATAC-seq
BD SNPs
HindIII fragments
300kb
Chr12
RNA-seq
PTPRO
EPS8
STRAP
p=9.4x10-13
Relative expression of STRAP
control 1
control 2
clone 1
clone 2
clone 3
e
H3K4me1
H3K4me3
ATAC-seq
SCZ SNPs
HindIII fragments
20kb
Chr11
RNA-seq
ANKK1
MIR4301
DRD2
p=2.7x10-7
Relative expression of DRD2
control 1
control 2
clone 1
clone 2
clone 3

Figure 6

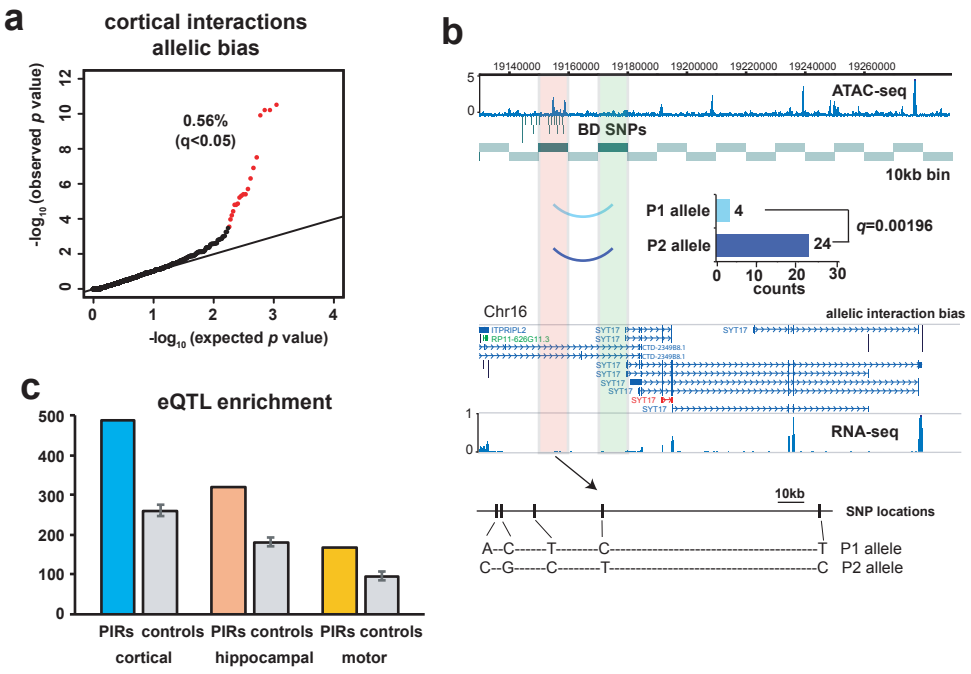

a
cortical interactions
allelic bias
0.56%
(q<0.05)
-log10 (observed p value)
-log10 (expected p value)
b
ATAC-seq
BD SNPs
10kb bin
P1 allele
P2 allele
4
24
q=0.00196
counts
Chr16
allelic interaction bias
RNA-seq
10kb
SNP locations
A--C------T--------C----------T P1 allele
C--G------C--------T----------C P2 allele
c
eQTL enrichment
PIRs controls PIRs controls PIRs controls
cortical
hippocampal
motor

Figure S1

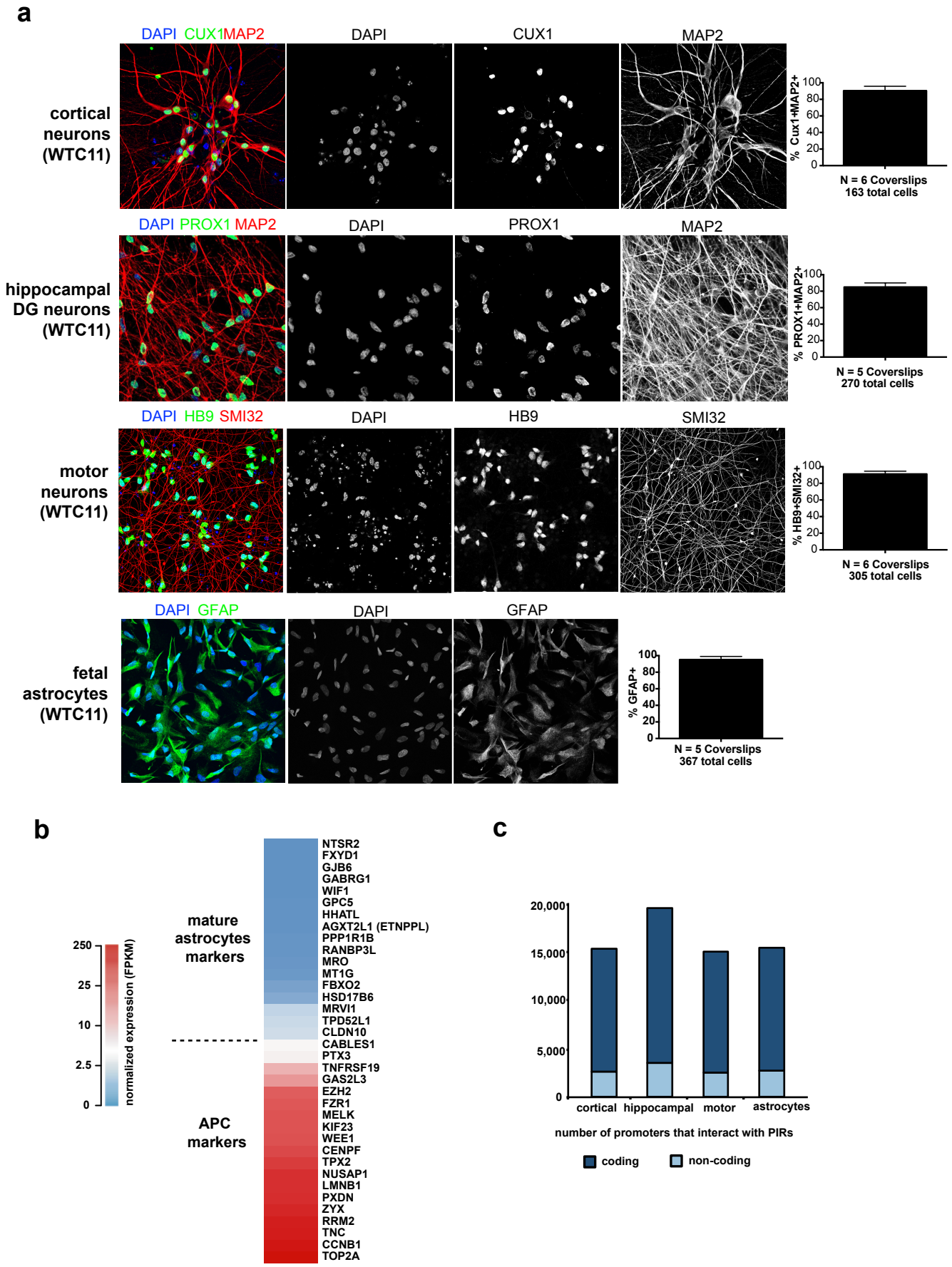

a
DAPI CUX1MAP2
DAPI
CUX1
MAP2
cortical neurons (WTC11)
% Cux1+MAP2+
N = 6 Coverslips
163 total cells
DAPI PROX1 MAP2
DAPI
PROX1
MAP2
hippocampal DG neurons (WTC11)
% PROX1+MAP2+
N = 5 Coverslips
270 total cells
DAPI HB9 SMI32
DAPI
HB9
SMI32
motor neurons (WTC11)
% HB9+SMI32+
N = 6 Coverslips
305 total cells
DAPI GFAP
DAPI
GFAP
fetal astrocytes (WTC11)
% GFAP+
N = 5 Coverslips
367 total cells
b
normalized expression (FPKM)
mature astrocytes markers
APC markers
NTSR2
FXYD1
GJB6
GABRG1
WIF1
GPC5
HHATL
AGXT2L1 (ETNPPL)
PPP1R1B
RANBP3L
MRO
MT1G
FBXO2
HSD17B6
MRVI1
TPD52L1
CLDN10
CABLES1
PTX3
TNFRSF19
GAS2L3
EZH2
FZR1
MELK
KIF23
WEE1
CENPF
TPX2
NUSAP1
LMNB1
PXDN
ZYX
RRM2
TNC
CCNB1
TOP2A
c
cortical
hippocampal
motor
astrocytes
number of promoters that interact with PIRs
coding
non-coding

Figure S2

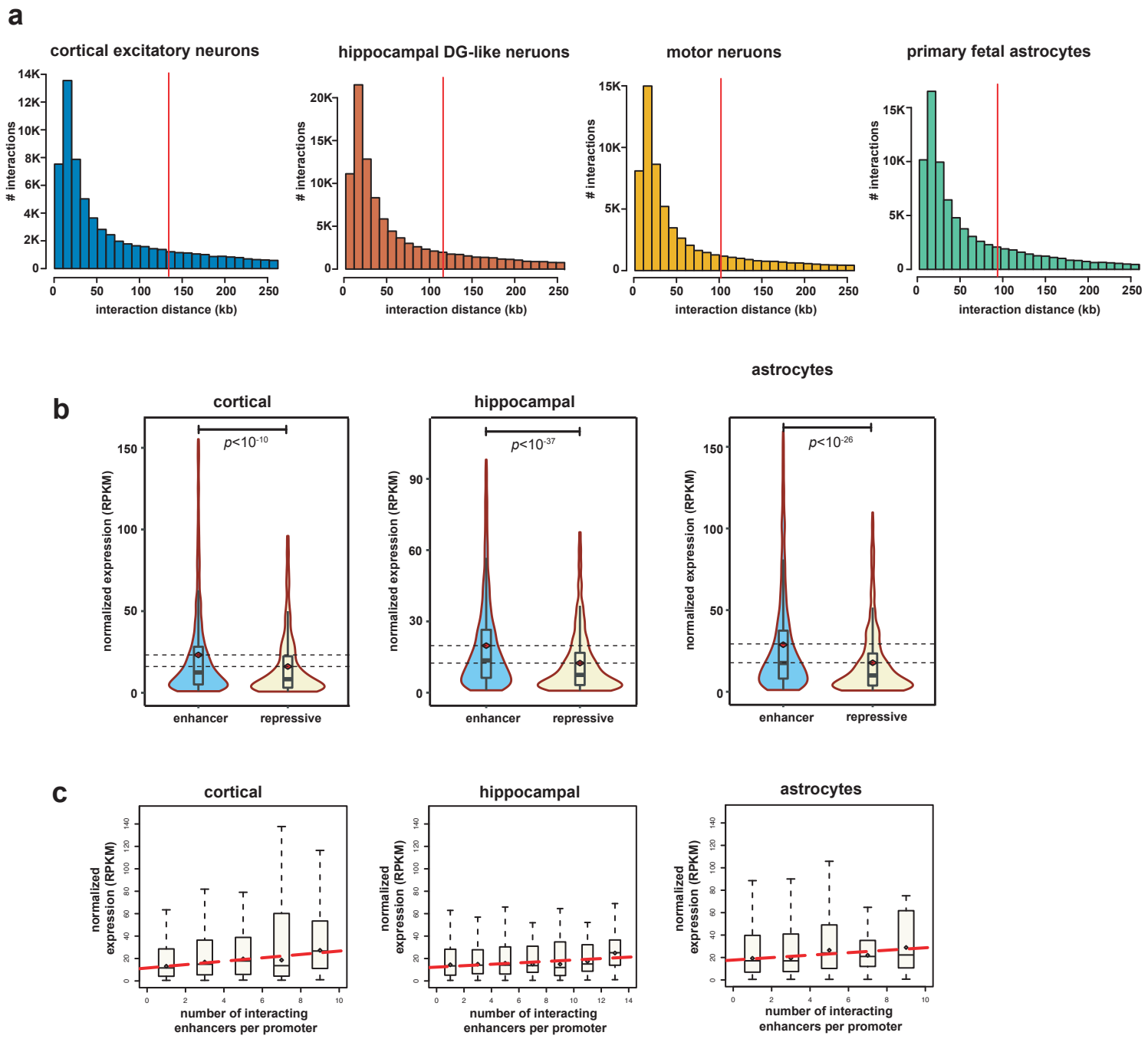
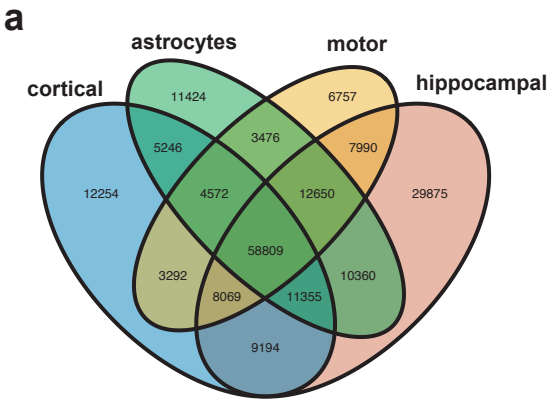
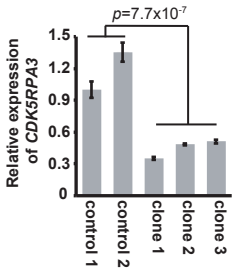
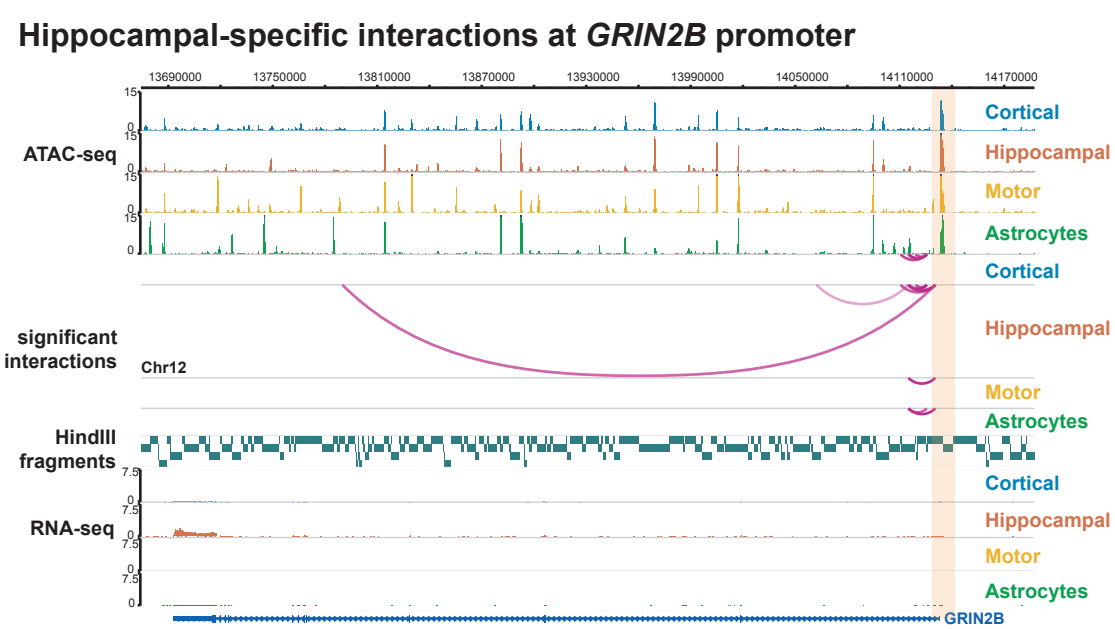

a
cortical excitatory neurons
hippocampal DG-like neruons
motor neruons
primary fetal astrocytes
interactions
interaction distance (kb)
b
cortical
hippocampal
astrocytes
normalized expression (RPKM)
p<10-10
p<10-37
p<10-26
enhancer
repressive
c
cortical
hippocampal
astrocytes
normalized expression (RPKM)
number of interacting enhancers per promoter

Figure S3


a
astrocytes
motor
cortical
hippocampal
11424
6757
5246
3476
7990
12254
4572
12650
29875
58809
3292
10360
8069
11355
9194


c
p=7.7x10-7
Relative expression of CDK5RPA3
1.5
1.2
0.9
0.6
0.3
0
control 1
control 2
clone 1
clone 2
clone 3


b
Hippocampal-specific interactions at GRIN2B promoter
13690000
13750000
13810000
13870000
13930000
13990000
14050000
14110000
14170000
ATAC-seq
Cortical
Hippocampal
Motor
Astrocytes
significant interactions
Chr12
HindIII fragments
RNA-seq
GRIN2B

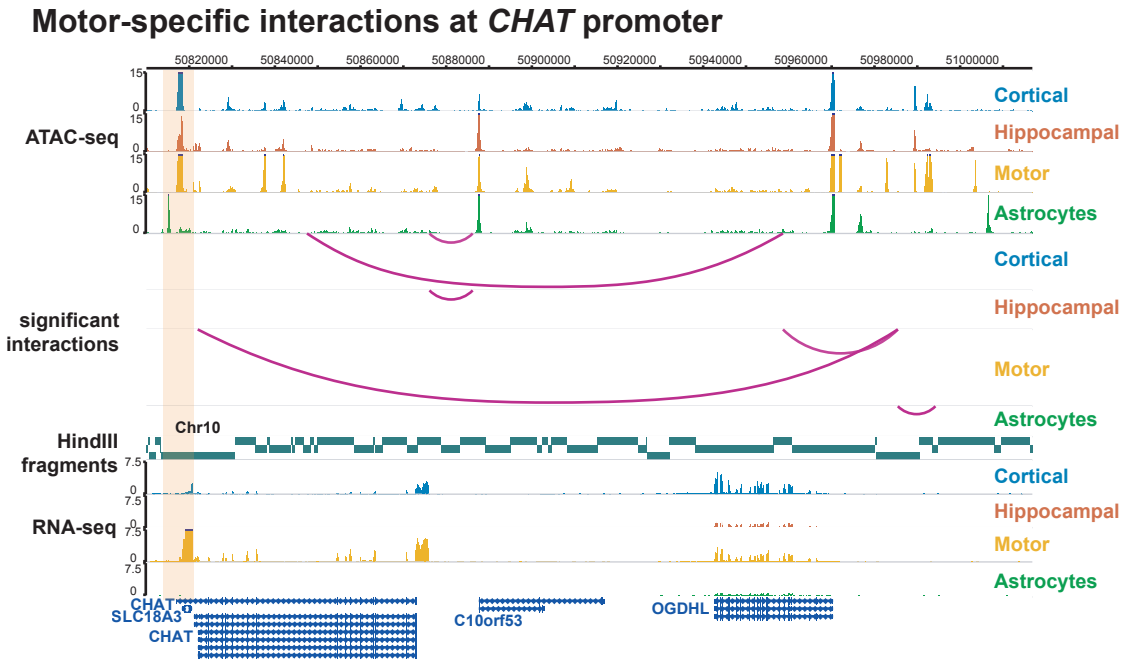
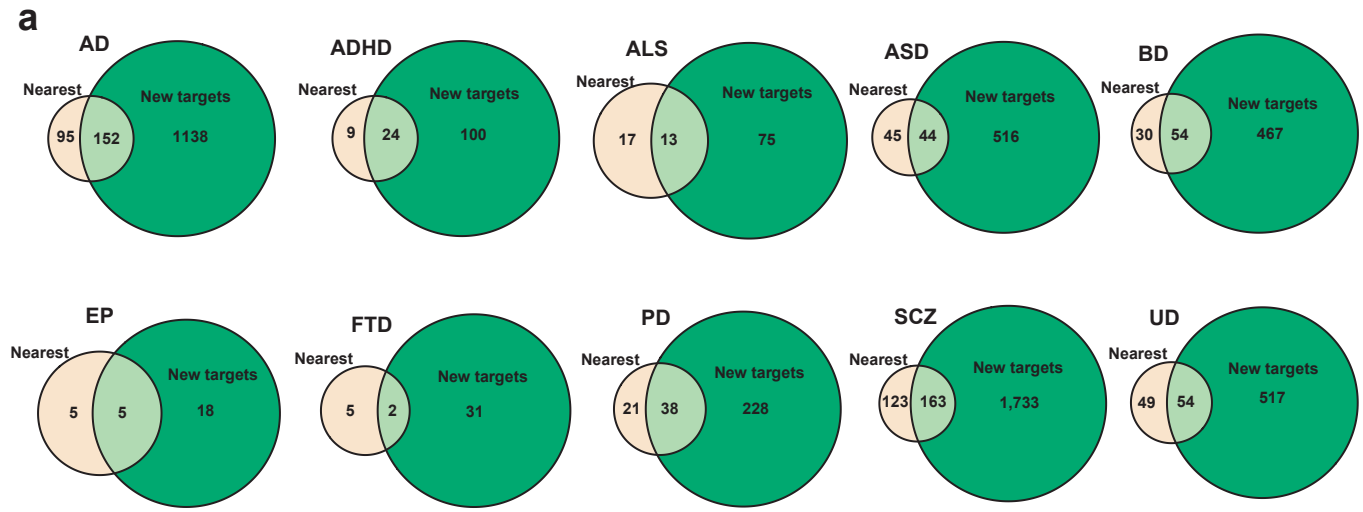
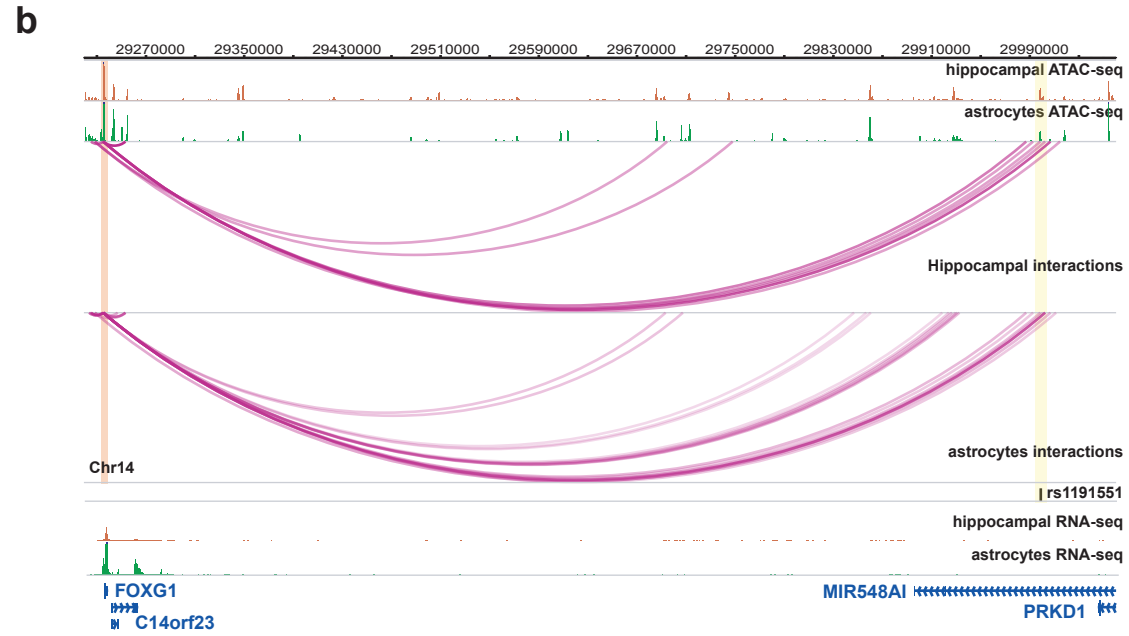
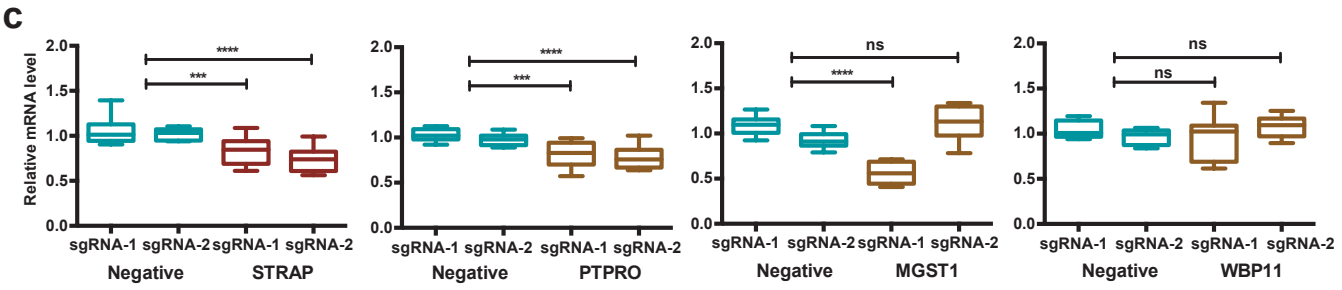
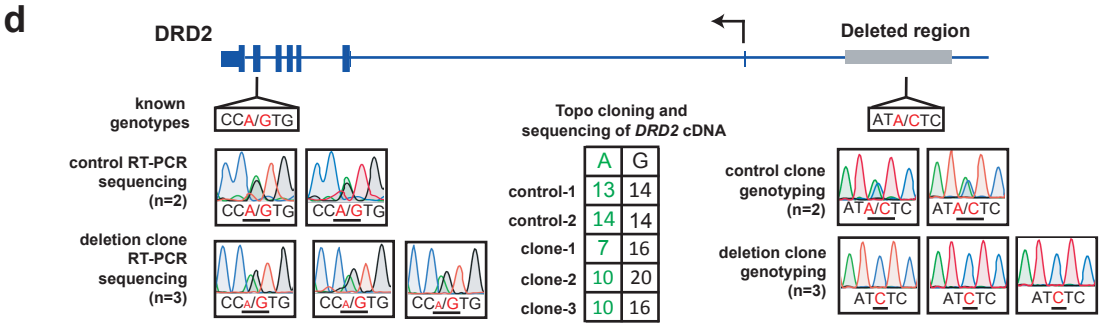

Motor-specific interactions at CHAT promoter
50820000
50840000
50860000
50880000
50900000
50920000
50940000
50960000
50980000
51000000
ATAC-seq
Cortical
Hippocampal
Motor
Astrocytes
significant interactions
Chr10
HindIII fragments
RNA-seq
CHAT
SLC18A3
C10orf53
OGDHL

Figure S4


a
AD
Nearest
New targets
95
152
1138
ADHD
9
24
100
ALS
17
13
75
ASD
45
44
516
BD
30
54
467
EP
5
5
18
FTD
5
2
31
PD
21
38
228
SCZ
123
163
1,733
UD
49
54
517


b
29270000
29350000
29430000
29510000
29590000
29670000
29750000
29830000
29910000
29990000
hippocampal ATAC-seq
astrocytes ATAC-seq
Hippocampal interactions
astrocytes interactions
Chr14
rs1191551
hippocampal RNA-seq
astrocytes RNA-seq
FOXG1
C14orf23
MIR548AI
PRKD1


c
Relative mRNA level
2.0
1.5
1.0
0.5
0.0
****
***
ns
sgRNA-1
sgRNA-2
Negative
STRAP
PTPRO
MGST1
WBP11


d
DRD2
Deleted region
known genotypes
CCA/GTG
ATA/CTC
control RT-PCR sequencing (n=2)
deletion clone RT-PCR sequencing (n=3)
Topo cloning and sequencing of DRD2 cDNA
A
G
control-1
13
14
control-2
14
14
clone-1
7
16
clone-2
10
20
clone-3
10
16
control clone genotyping (n=2)
deletion clone genotyping (n=3)
ATCTC