Title: Open chromatin profiling of human postmortem brain infers functional roles for non-coding schizophrenia loci

John F. Fullard[1], Claudia Giambartolomei[1], Mads E. Hauberg[1,6,7,8], Ke Xu[2], Georgios Voloudakis[1], Zhiping Shao[1,4,5], Christopher Bare[3], Joel T. Dudley[2], Manuel Mattheisen[6,7,8], Nikolaos K. Robakis[1,4,5], Vahram Haroutunian[1,4,9], Panos Roussos[1,2,9]*

[1]Department of Psychiatry; [2]Department of Genetics and Genomic Science and Institute for Multiscale Biology, [3]Flow Cytometry Center of Research Excellence, [4]Department of Neuroscience,

[5]Center for Molecular Biology and Genetics of Neurodegeneration, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

[6]Department of Biomedicine, [7]Centre for Integrative Sequencing (iSEQ), Aarhus University, Aarhus, Denmark

[8]The Lundbeck Foundation Initiative of Integrative Psychiatric Research (iPSYCH), Denmark

[9]Mental Illness Research, Education, and Clinical Center (VISN 2 South), James J. Peters VA Medical Center, Bronx, NY, USA

*** Corresponding author:**
Dr. Panos Roussos
Icahn School of Medicine at Mount Sinai
Department of Psychiatry and Department of Genetics and Genomic Science and Institute for Multiscale Biology
One Gustave L. Levy Place,
New York, NY, 10029, USA
Phone: +1 (212) 824-8982
Fax: +1 (646) 537-9583
email: Panagiotis.roussos@mssm.edu

ABSTRACT

Open chromatin provides access to DNA binding proteins for the correct spatiotemporal regulation of gene expression. Mapping chromatin accessibility has been widely used to identify the location of *cis* regulatory elements (CREs) including promoters and enhancers. CREs show tissue- and cell-type specificity and disease-associated variants are often enriched for CREs in the tissues and cells that pertain to a given disease. To better understand the role of CREs in neuropsychiatric disorders we applied the Assay for Transposase Accessible Chromatin followed by sequencing (ATAC-seq) to neuronal and non-neuronal nuclei isolated from frozen postmortem human brain by fluorescence-activated nuclear sorting (FANS). Most of the identified open chromatin regions (OCRs) are differentially accessible between neurons and non-neurons, and show enrichment with known cell type markers, promoters and enhancers. Relative to those of non-neurons, neuronal OCRs are more evolutionarily conserved and are enriched in distal regulatory elements. Transcription factor (TF) footprinting analysis identifies differences in the regulome between neuronal and non-neuronal cells and ascribes putative functional roles to a number of non-coding schizophrenia (SCZ) risk variants. Among the identified variants is a Single Nucleotide Polymorphism (SNP) proximal to the gene encoding SNX19. *In vitro* experiments reveal that this SNP leads to an increase in transcriptional activity. As elevated expression of *SNX19* has been associated with SCZ, our data provides evidence that the identified SNP contributes to disease. These results represent the first analysis of OCRs and TF binding sites in distinct populations of postmortem human brain cells and further our understanding of the regulome and the impact of neuropsychiatric disease-associated genetic risk variants.

INTRODUCTION

The 3-dimensional structure of chromatin plays a central role in the regulation of gene expression and is essential for the maintenance of cell identity and function (1). Specific patterns of open chromatin expose the repertoire of *cis* regulatory elements (CREs) required to regulate transcription in a given cell or cell-type. A large proportion of neuropsychiatric disease-associated loci are non-coding and exert their effects by disrupting CREs required for the correct spatiotemporal expression of genes (2-5). Importantly, *cis* regulation of gene expression is often specific for tissue and even cell type (3, 4). Correspondingly, disease-associated loci have been shown to be enriched for CREs in tissues and cells relevant to the pathophysiology of disease (2, 4-6).

Employing existing epigenome data to further our understanding of neuropsychiatric diseases is hindered by the fact that much of the epigenomic landscape remains unexplored in the relevant cells: The Encyclopedia of DNA Elements (ENCODE) consortium (4, 7), Roadmap Epigenomics Mapping Consortium (REMC) (3, 8), and FANTOM5 (9) all focused on actively dividing cells or used only homogenate brain tissue. A limitation to the latter approach is that the resultant data was derived from a mixture of markedly different cells such as neurons, microglia, oligodendrocytes, and astrocytes. Because CRE-mediated epigenetic regulation shows cell-type specificity (3, 4), the study of mixed cell populations can fail to detect cell-type-specific signals. Furthermore, in studies using homogenized brain tissue, the proportion of each cell type contributing to the assay is undetermined, further increasing sample-to-sample variability.

Here we present, to our knowledge, the first cell-type-specific map of open chromatin regions (OCRs) of postmortem human frontopolar prefrontal cortex. We identify thousands of cell-type specific OCRs from 2 broad cell types (neuronal (NeuN+) and non-neuronal (NeuN-)) isolated from frozen postmortem tissue by fluorescence-activated nuclear sorting (FANS). In addition, we examine transcription factor (TF) binding to interrogate cell-type differences in the regulation of gene expression. Finally, a number of the identified OCRs and TF binding sites coincide with known schizophrenia (SCZ) risk loci, allowing us to pin-point the likely

causative SNP therein and providing evidence for a putative mechanism by which these variants contribute to the etiology of SCZ.

RESULTS

Profiling chromatin accessibility in Neuronal and non-Neuronal Cells

Our aim was to catalog the OCRs in human brain tissue and examine differences among neuronal and non-neuronal cells. We assessed the landscape of OCRs in neuronal and non-neuronal nuclei extracted from 8 controls samples (**Table S1**) using ATAC-seq (**Figure 1A**). We examined the effects of technical variability on ATAC-seq profiles and found high concordance of read coverage among technical replicates (Pearson's r=0.97) (**Figure S1A**). The average number of uniquely mapped and non-duplicated paired-end reads per sample was 54.8 million, with low mitochondrial DNA contamination (**Table S2**). After exploratory analysis (**Figure S1B**), a total of 72,033 and 68,606 peaks were defined in neuronal and non-neuronal samples, respectively (**Figure S1C**). For quantitative analysis of differences among neuronal and non-neuronal samples, we generated a matrix of mapped reads using a final consensus of 115,021 peaks. Unsupervised hierarchical clustering identified a clear distinction between neuronal and non-neuronal samples. (**Figure 1B**). Comparison of neuronal vs. non-neuronal peaks identified more than 50% of OCRs that were significant after multiple testing corrections (60,653 differentially modified OCRs at FDR ≤ 0.01) (**Figure 1C, Table S3**). Among these, 33,054 were neuronal and 27,599 were non-neuronal, with a moderate to large average fold change (FC) (median FC 2.19; range 1.46-15.86, **Figure 1D**). For example, a neuron-specific regulatory region is positioned ~29kb upstream of *CADM3* (Cell adhesion molecule 3), a gene that is highly expressed in cortical pyramidal cells (10) (**Figure 1E**). A non-neuronal region is positioned within the transcription start site of the shorter *TGIF1* (TGFB-Induced Factor Homeobox 1) isoform, a gene with high expression in microglia (11) associated with Holoprosencephaly-4 (OMIM: 142946) (**Figure 1F**). Overall, our analysis has identified significant differences in the epigenome landscape of OCRs among neuronal and non-neuronal populations.

Annotation of the Cell-type Specific Regulome in Neuronal and non-Neuronal Cells

We first examined the location of OCRs in regards to distance from transcription start site (TSS) and genic annotations. OCRs are in the vicinity of TSSs (**Figure 2A**). Relative to non-neuronal OCRs, the differentially accessible regions of neurons are enriched for distal regulatory elements [10kb – 100kb downstream of TSSs (OR = 0.7, P = 9.8×10^{-39}) and introns (OR =1.2, P = 6.5×10^{-130})], suggesting a more important role for long-range regulation of gene expression in neurons (**Figure 2B and 2C**). We next examined if the identified OCR regions were under evolutionary constraint as evidenced by higher Genomic Evolutionary Rate Profiling (GERP) scores (12). High GERP score was found for both neuronal and non-neuronal OCRs (**Figure 2D**). Interestingly, neuronal OCRs exhibit markedly higher GERP scores than non-neuronal OCRs (0.26 vs 0.14, $p = 7.7 \times 10^{-14}$ in a two-sided t test), which is more prominent in intergenic and promoter regions (two-sided t test $p < 2.2 \times 10^{-16}$ for both comparisons), respectively (**Figure S2**). Overall, compared to non-neuronal, the neuronal OCRs tend to be located distally from TSSs and are evolutionary more conserved.

OCRs are enriched for active and poised promoters and active and repressed enhancers derived from published histone modification data from homogenates of human prefrontal cortex (3) (**Figure 2E**). Systematic comparison of cell-type specific OCRs with published regulatory sequences across multiple tissues showed a strong enrichment in neuronal OCRs for brain-related enhancers, promoters and OCRs (**Figure S3**). The non-neuronal OCRs show lower specificity for brain related CREs. To further explore the identified OCRs, we conducted pathway analysis by annotating peaks to proximal genes. In this analysis, neuronal peaks showed a significant enrichment for the terms synaptic transmission, cell-cell signaling, and cellular morphogenesis (**Figure 2F**), whereas non-neuronal peaks were enriched for pathways related to protein kinase binding and telencephalon development. Next, we examined the overlap between genes proximal to the differentially accessible elements and cell type-specific markers (10, 11). We detected a highly significant overlap between the set of the predicted human neuronal genes and the set of cell type-specific genes for neurons, including pyramidal cells and interneurons (**Figure 2G**). For non-neuronal OCRs, we observed a strong enrichment for

oligodendrocytes and astrocytes. Overall, neuronal OCRs show stronger enrichment with published brain-related enhancers and are enriched for molecular pathways related to neuronal function, including synaptic transmission and cell-cell signaling.

Cell-type Specific Transcription factor binding and gene regulation

We performed TF binding analysis based on a set of 432 motifs (13). Out of these motifs, we focus on 179 as they are enriched in the neuronal or non-neuronal OCRs compared to the genomic background. In addition to being enriched in motif binding sites, the shape of the reads within ATAC-peaks can also be used to infer TF binding, where the actual binding site shows a "footprint" as the bound TF sterically protects the DNA from transposase integration (14, 15) (**Figure 3A**). In order to improve TF binding prediction, we applied Protein Interaction Quantification (PIQ) (16), which, for each motif, learns the pattern of transposition insertion in the vicinity of potential binding sites. To measure the relative importance of each TF in the neuronal and non-neuronal samples, we used the output of PIQ to calculate motif scores and compared them against each other (**Figure 3B, Table S4 and S5**). We note that similar results to the one described below could be obtained by comparing the enrichment of motifs in the ATAC-seq peaks regardless of footprinting (**Figure S4**).

The most prominent cell-type-specific TF differences were observed for promoter-depleted TFs. The TFs most specific for neurons were the FOS/JUN families (which together form the AP1-complex), showing a 2.7-fold higher motif score in neuronal cells. These TFs also showed greater than two-fold enrichment in neuronal peaks with, conversely, a slight depletion in peaks of non-neurons (**Table S5**). For non-neuronal cells, the two most specific motifs were ONECUT1/2/3 and PAX3/7, of which the latter has the most established role in brain function (17, 18). We further note that the TF motifs specific to either the neuronal or non-neuronal cells appear to fall into distinct families. For instance, the Basic helix-loop-helix (bHLH) family seems to be primarily active in neuronal cells whereas the homeodomain/sox TF families seem to be predominantly non-neuronal (**Table S4**). On the other hand, several motifs of general TFs are identified mostly in promoter regions and show a high motif score and enrichment in both cell types (**Figure 3B**). We provide an illustrative example of cell type-

specific TF binding sites for *TRPM3* (Transient Receptor Potential Cation Channel, Subfamily M, Member 3), where two CTCF sites are predicted only in neuronal samples (**Figure 3C**). Overall, we identified a distinct repertoire of TFs for neuronal and non-neuronal cells, including AP1-complex and bHLH family which has an important role in neuronal function (19, 20) and shows preferential activity for neurons in our data.

We next examined the likelihood that the TFs were regulating the same or different genes in the neuronal and non-neuronal samples by calculating the regulatory divergence (21). The two most divergent motifs jointly represented the neurogenin and neuroD TF families (**Figure 3D**), which have an opposite effect on neuronal and non-neuronal function as they promote neurogenesis and inhibit gliogenesis (22). At the other end of the spectrum, the least divergent TFs consisted of many general TFs, including NFYB and NFYA (**Table S5**). As an alternative approach, we applied the concept of regulatory divergence to identify genes that showed either a similar or a dissimilar pattern of TF regulation in the two cell types (**Figure 3E).** The most divergently regulated gene was *SYT1*, which encodes Synaptotagmin 1, a neuronally expressed protein known to mediate synaptic vesicle exocytosis (23). Pathway analysis of the top 1000 most divergently regulated genes showed significant enrichment for biological processes related to G-protein coupled receptor (GPCR) activity (**Figure 3F**). This is consistent with recent findings demonstrating cell type specific expression of GPCR genes in neuronal and non-neuronal cells (10).

Enrichment of SCZ risk loci for OCRs and TF binding sites

As an initial analysis, we tested if the identified OCRs were enriched in risk loci for SCZ (24), Alzheimer's disease (AD) (25) and non-neuropsychiatric diseases (26-29), using an empirical Bayes approach (30). This statistical approach leverages Genome Wide Association Study (GWAS) summary statistics and identifies functional annotations that are significantly enriched for risk loci. We found enrichment for SCZ in neuronal [log2 enrichment (95% Confidence Interval or 95CI) = 2.11 (1.21 - 2.73)] and non-neuronal [log2 enrichment (95CI) = 1.57 (0.01 - 2.38] OCRs (**Figure S5A**). This is consistent with the tissue specific enrichment of OCRs

with relevant diseases (2, 4). In addition, compared to non-neuronal OCRs, we found a stronger enrichment in neuronal OCRs, thereby providing additional support for neurons as the functional unit affected by SCZ susceptibility loci.

Given the significant enrichment of OCRs with SCZ loci, we next tested whether OCRs within specific genic annotations are more enriched in SCZ. The most enriched annotations were neuronal introns [log2 enrichment (95CI) = 2.95 (2.10 - 3.55)] and non-neuronal promoters [log2 enrichment (95CI) = 3.33 (2.38 - 3.98)] (**Figure S5B**). These two annotations were used to build a combined model to identify the functional variant underlying each disease-associated locus. This approach incorporates functional data and reweights GWAS SNPs, allowing to identify variants with higher likelihood of being functional compared to other SNPs in the same locus. This yielded 29 SNPs that localize within 20 out of the 108 GWAS SCZ loci as the most likely candidates to be the causal polymorphisms in each region (**Table S6 OCR model**). For 19 out of the 20 GWAS SCZ loci, the functional SNP is not the GWAS index SNP. We found a substantial increase of ~10 folds for the likelihood (estimated based on the fitted empirical prior probability) of a functional SNP (average prior = 0.38%) to be the causal polymorphism in this region compared to the index GWAS SNP (average prior = 0.036%). For 13 out of the 20 SCZ risk loci, we also identified an effect of the putative functional SNP on gene expression using expression quantitative trait analysis from the CommonMind Consortium (31).

SCZ causal variants might affect the binding sites for distinct classes of transcription factors in neuronal and non-neuronal cells. To examine this hypothesis, we performed enrichment analysis of SCZ GWAS summary statistics with TF binding sites detected in neuronal and non-neuronal cells. This analysis was run by considering each TF separately (single TF model). We identified 15 TFs derived mostly from neuronal (11 TFs) compared to non-neuronal (4 TFs) cells that were enriched with SCZ genetic variants (**Figure 4A**). In order to examine the combinational effect of the significant TFs in SCZ, we combined the 15 TFs in a joint model (combined TF model), using cross-validation to overcome overfitting in the model. Our best-fitting model included 4 TFs: ZSCAN10, NANOG/NANOGP1, CEBPZ and ZNF354C, all of which were specific to

neuronal cells (**Figure 4A**). The single and combined TF models were used to reweigh the SCZ GWAS and identified 7 SNPs that localize within 6 out of the 108 GWAS SCZ loci as the candidates most likely to be the causal polymorphisms in each region (**Table S6 TF model**). In this analysis, none of the functional SNPs were the GWAS index SNP. We found a substantial increase of ~100 folds for likelihood (estimated based on the fitted empirical prior probability) of a functional SNP (average prior = 1.91%) to be the causal polymorphism in this region compared to the index GWAS SNP (average prior = 0.019%). For 3 out of the 6 SCZ risk loci, we also identified an effect of the putative functional SNP on gene expression of 4 transcripts (*MAN2A1*, *SNAP91*, *MRAP2* and *SNX19*) using expression quantitative trait analysis from the CommonMind Consortium (31) (**Table S6 TF model**).

Figure S6 shows the distribution of the prior probability of functional SNPs across the 108 GWAS SCZ loci. Compared to the neuronal introns and non-neuronal promoters combined model, the TF models (single and combined) include multiple functional SNPs with high likelihood (priors > 1%) in SCZ loci (**Figure S6**). A subset of variants has additional support for a functional role, affecting gene expression abundance of specific transcripts. **Figure 4B** shows an illustrative example for a locus near the *SNX19* gene, where the combined TF model identified rs10750450 as the most likely causal variant in this region. This SNP has a P value of 2.19×10^{-12}, which is close to the level of significance of the index SNP of that locus (rs10791097; $P = 1.56 \times 10^{-12}$). However, this SNP falls in the proximity of 2 TF binding sites (ZSCAN10 and ZNF354C), leading the model to assign a prior (14.58%) that is almost three orders of magnitude higher than the prior of the index SNP (0.02%).

Functional validation of rs10750450 on transcriptional activity

To validate this finding, we examined the effect of rs10750450 risk-T and reference-G alleles and the binding motifs of ZNF354C and ZSCAN10 on transcriptional activity in *in vitro* experiments (**Table S7**). We found a significant effect of the different constructs on luciferase activity (ANCOVA: $F = 42.73$, $df = 6$, $P < 2e\text{-}16$) (**Figure 4C**). All constructs resulted in increased luciferase activity ranging from 47% to 128% (all adjusted Ps < 0.0005) compared to empty pGL4.24 vector. Compared to the reference G allele, the risk T allele results in a

34% increase in luciferase activity (adjusted P = $1.4x10^{-4}$). rs10750450 is in proximity to binding sites for ZNF354C and ZSCAN10. Our data supports a role for both TFs as transcriptional repressors, as excision of ZNF354C (34% increase, adjusted P = $1.2x10^{-4}$) or ZSCAN10 (36% increase, adjusted P = $6.4x10^{-5}$) binding sites result in increased luciferase activity compared to the reference G allele. We note that excision of both ZNF354C and ZSCAN10 binding sites resulted in a further increase of luciferase activity (47% increase, adjusted P = $1.5x10^{-7}$) and a similar effect was observed when we removed the rs10750450 nucleotide (56% increase, adjusted P = $1.0x10^{-9}$). Overall, our results indicate that, compared to the reference sequence, the schizophrenia risk allele leads to increased luciferase activity which is consistent with the association of the risk allele with increased *SNX19* gene expression (32). Based on our data, this effect appears to be mediated through an allele specific effect of the rs10750450-T allele on ZNF354C and ZSCAN10 binding, leading to de-repression of transcriptional activity.

DISCUSSION

Recent genetic studies have implicated numerous common risk variants in SCZ (24). One of the next challenges is to further understand the biological mechanisms of the large number, and diversity, of genes that are associated with SCZ. To that end, we need to generate additional data capturing putative molecular processes that are relevant to the development of the disease. The majority of SCZ risk variants is found within non-coding regions of the genome and is predicted to disrupt the function of CREs. We explored SCZ genetic architecture by leveraging, for the first time to our knowledge, cell type-specific OCRs mapped in human brain tissue. While, SCZ-associated abnormalities have been demonstrated in neuronal and non-neuronal cells (2, 33), here we demonstrate that SCZ risk variants show a higher enrichment in neuronal OCRs. This is consistent with genetic findings implicating genes that participate in neuronal function and synaptic transmission in the etiology of SCZ (24, 34, 35). It is worth noting that we did not find a significant association between risk loci for AD and OCRs of the brain. This is consistent with recent findings that suggest AD may have an immune cell component, as AD disease variants are enriched in enhancers of peripheral blood cells (3, 36, 37).

Comparison of neuronal and non-neuronal OCRs demonstrated the following interesting observations: First, although we observed an increased frequency of accessible elements for both cell-types in proximity to transcription start sites (TSSs), distal regulatory regions appear to play a more critical role in neurons when compared to non-neurons. This finding compliments a previous study, using an independent cohort of brain samples, which reported that, compared to non-neurons, neuron specific methylated regions tend to be located distally from TSSs and are enriched within predicted enhancer elements (38). Second, while both neuronal and non-neuronal OCRs are evolutionary conserved, those of neurons appear to be under stronger functional constraint than other brain cells. Third, we also observe cell-type differences in the regulation of gene expression between the two cell types with promoter-depleted TFs showing the most marked differences between cells. Based on our analysis, the TFs with higher binding affinity for neurons were among the FOS/JUN (39) and bHLH families (40) whereas PAX3/7 (41) and the homeodomain/sox (42) TF families display non-neuronal preference.

To further refine the OCRs, we performed TF digital footprinting analysis, which provides higher resolution of the functional genomic regions (from ~1kb average OCR size to ~10bp for TF binding site) and assigns a potential functional role based on known TFs. Comparing TF to OCR data, we found a substantial increase of ~5 folds for likelihood of a functional SNP being the causal polymorphism in certain SCZ risk regions. In addition, we identified a subset of TFs that are highly enriched in SCZ loci, including ZSCAN10, NANOG/NANOGP1, CEBPZ and ZNF354C, all of which were specific to neuronal cells. While little is known about the function of CEBPZ and ZNF354C, both NANOG and ZSCAN10 have been shown to play a role in the maintenance of pluripotency in embryonic stem cells (ESCs). NANOG is a downstream target of ZSCAN10 transcriptional activity (43) and NANOG expression is thought to be restricted to ESCs, becoming progressively down-regulated during differentiation and embryonic development (44, 45). This raises the question as to why we detect TF footprints for developmental genes in neurons of the adult brain? While it is possible that NANOG and ZSCAN10 have heretofore unknown roles in postmitotic neurons, recent evidence suggests that pro-neural TF footprints can be maintained over time to ensure proper neuronal and glial

differentiation (46). As such, we may be detecting the impression left by a protein on the structure of DNA, several decades after the fact. To that end, there is evidence supporting a role for ZSCAN10 in maintaining a multipotent progenitor cell population in mid-gestation embryos and adult organs (47). This is also supported by our data, where a large proportion of ZSCAN binding sites (~27%, see supplement) is present in active enhancers (defined based on H3K27ac ChIP-seq studies (48)) during adulthood.

SCZ risk loci are frequently large and often contain multiple implicated SNPs due to local linkage disequilibrium patterns. Below, we describe a number of genes identified in our analysis with particular emphasis on genes previously associated with increased risk for SCZ from the CommonMind Consortium (CMC) analysis (31).

We applied our TF footprinting models to reweigh the SCZ GWAS and identified the variants that have a putative functional role for 6 out of the 108 GWAS SCZ risk loci, none of which is the index SNP identified by GWAS. Of the 6 loci, 3 were identified as having an effect on gene expression using expression quantitative trait analysis from the CMC (31). One example is a locus adjacent to the *SNX19* gene where the index SNP was identified as rs10791097. Our combined TF model identified a different SNP, rs10750450, as the most likely causal variant in this region, due to its proximity to binding sites for ZSCAN10 and ZNF354C. *In vitro* experiments confirmed that this SNP leads to an increase in transcriptional activity. In addition, rs10750450 is an expression quantitative trait locus (eQTL) for the *SNX19* transcript in the human brain (31), but also in multiple tissues based on the GTEx data (http://www.gtexportal.org/). Furthermore, *SNX19* was recently identified as a gene (out of a total of two genes) whose expression level was associated with SCZ (32). By locating the putative functional regulatory region for this gene within neurons, we add to the growing evidence that up-regulation of *SNX19* increases the risk of SCZ.

Our approach has allowed us to identify additional, potentially functional, SNPs among regulatory regions proximal to a number of genes including; *MAP3K*, which has previously been implicated in SCZ, major

depressive disorder (MDD) and Bipolar (BP) disorder (49, 50), *TOM1L2*, a lipid pathway gene that resides in a locus associated with increased risk of dementia (51) and *GATAD2A*, a zinc-finger domain containing transcriptional repressor involved in methylation-dependent gene silencing (52). We also identify a locus adjacent to the gene encoding *SNAP91*. The GWAS index SNP in this instance is chr6_84280274_D, whereas our TF model identifies rs2022265 and rs2023569 as more likely causal SNPs given their proximity to MEF2A, MEF2B, MEF2C and MEF2D binding sites. Both MEF2A and MEF2C are thought to play roles in neuronal differentiation (53). *SNAP91* encodes a 91 kDa synaptosomal-associated protein, also known as Clathrin Assembly Protein 180 (AP180), that is highly expressed in neuronal presynaptic termini (54, 55) where it is required for the formation and function of clathrin coated vesicles (56). *SNAP91* has also been shown to control the growth of postmitotic neurons in embryonic hippocampus (57) and has proposed roles in calcium signaling and the Wnt pathway (58). In addition to its association with SCZ, *SNAP91* has been identified in a GWAS for BP disorder risk loci (59).

Another SNP identified in our studies is rs4318227, adjacent to the gene encoding Double C2 Domain Alpha (*DOC2A*). *DOC2A* has been identified as an activity-dependent Ca(2+) sensor, active during the asynchronous phase of neurotransmitter release during synaptic transmission (60, 61). In addition, our OCR model identified another SNP (rs6553440) adjacent to *CLCN3*, a voltage-gated chloride channel enriched in the glutamatergic synapses of the hippocampus. *Clcn3*-/- knockout mice have altered GABAergic function and display postnatal degeneration of both the hippocampus and the retina, indicating the critical role played by *CLCN3* in normal neurotransmission in central neurons (62-64). More recently, *CLCN3* has been shown to play a role in controlling glutamatergic synaptic strength by regulating neurotransmitter levels and synaptic vesicle (SV) release in cultured mouse hippocampal *Clcn3*-/- neurons (65).

Our study is not without limitations, including a relatively small sample size consisting of elderly non-schizophrenic controls. As such, our findings warrant further validation in future studies. As is the case with most postmortem studies, the current observations could be partially attributed to technical and clinical

covariates, including postmortem interval and agonal state. In addition, we chose to focus on a discreet region of the brain and, due to the limitations of working with frozen postmortem tissue (in particular, the loss of cytoplasmic markers upon thawing), our analysis was restricted to the study of two broad populations of cells – neurons and non-neurons. However, both populations are themselves very heterogeneous, consisting of a variety of neural and non-neuronal sub-types. Because isolation of non-neuronal nuclei relies on negative selection (NeuN negative), this population is more heterogeneous and includes several diverse cell types such as oligodendrocytes, astrocytes and microglia. This heterogeneity in the non-neuronal population will lead to less power to detect cell type specific OCRs (which are located more distally to TSS) and might explain some of the differences we observed in comparison to neurons. Studying different regions of the brain, coupled with the application of additional cell-type specific nuclear markers, e.g. (66) would broaden the scope of our approach towards a more thorough understanding of the means by which CREs influence brain function and disease.

We have generated the first open chromatin maps of distinct populations of cells in human postmortem brain. This has allowed us to identify specific patterns of gene regulation in neuronal and non-neuronal cells. In addition, our approach provides a means to pin-point non-coding SCZ risk variants and to assign functional roles to these SNPs by identifying the genes and transcription factor pathways they disrupt.

METHODS AND MATERIALS

Brain tissue specimens from the frontopolar prefrontal cortex (Brodmann area 10) of 8 controls were obtained from the NIH Brain and Tissue Repository (Table S1). Neuronal (NeuN+) and non-neuronal (NeuN-) nuclei were sorted using a FACSAria flow cytometer.

The Assay for Transposase Accessible Chromatin followed by sequencing (ATAC-seq) was performed using an established protocol (14) and sequenced on Hi-Seq2500 (Illumina) obtaining 2x50 paired-end reads. We used the edgeR package (67) to model the normalized read counts using negative binomial distributions including cell type, gender, age of death and PMI as covariates. p-values were adjusted for multiple hypothesis testing

using false discovery rate (FDR) ≤ 1%. The protein interaction quantitation PIQ framework (16) was used to predict transcription factor binding sites from the genome sequence. To integrate functional annotations and GWAS results, we used the fGWAS software (30) and performed two different models that considered OCRs (OCR model) or TF binding sites (TF model). Functional SNPs were further explored using expression quantitative trait loci (eQTLs) from prefrontal cortex (31). More details are provided in the **Supplement**.

ACKNOWLEDGEMENTS

This work was supported by the National Institutes of Health [R01AG050986 to P.R., R01MH109677 to P.R.]; Brain Behavior Research Foundation [20540 to P.R.]; Alzheimer's Association [NIRG-340998 to P.R.]; and the Veterans Affairs [Merit grant BX002395 to P.R.].

CONFLICT OF INTEREST STATEMENT

The authors report no biomedical financial interests or potential conflicts of interest.

REFERENCES

1. Fullard, J.F., Halene, T.B., Giambartolomei, C., Haroutunian, V., Akbarian, S., Roussos, P. Understanding the genetic liability to schizophrenia through the neuroepigenome. *Schizophrenia Research*,
2. Roussos, P., Mitchell, A.C., Voloudakis, G., Fullard, J.F., Pothula, V.M., Tsang, J., Stahl, E.A., Georgakopoulos, A., Ruderfer, D.M., Charney, A.*, et al.* (2014) A role for noncoding variation in schizophrenia. *Cell Rep*, **9**, 1417-1429.
3. Roadmap Epigenomics, C., Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J.*, et al.* (2015) Integrative analysis of 111 reference human epigenomes. *Nature*, **518**, 317-330.
4. Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J.*, et al.* (2012) Systematic localization of common disease-associated variation in regulatory DNA. *Science*, **337**, 1190-1195.
5. Trynka, G., Sandor, C., Han, B., Xu, H., Stranger, B.E., Liu, X.S., Raychaudhuri, S. (2013) Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nature genetics*, **45**, 124-130.
6. Franzen, O., Ermel, R., Cohain, A., Akers, N.K., Di Narzo, A., Talukdar, H.A., Foroughi-Asl, H., Giambartolomei, C., Fullard, J.F., Sukhavasi, K.*, et al.* (2016) Cardiometabolic risk loci share downstream cis- and trans-gene regulation across tissues and diseases. *Science*, **353**, 827-830.
7. Bernstein, B.E., Birney, E., Dunham, I., Green, E.D., Gunter, C., Snyder, M. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57-74.
8. Zhu, J., Adli, M., Zou, J.Y., Verstappen, G., Coyne, M., Zhang, X., Durham, T., Miri, M., Deshpande, V., De Jager, P.L.*, et al.* (2013) Genome-wide chromatin state transitions associated with developmental and environmental cues. *Cell*, **152**, 642-654.
9. Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T.*, et al.* (2014) An atlas of active enhancers across human cell types and tissues. *Nature*, **507**, 455-461.
10. Zeisel, A., Munoz-Manchado, A.B., Codeluppi, S., Lonnerberg, P., La Manno, G., Jureus, A., Marques, S., Munguba, H., He, L., Betsholtz, C.*, et al.* (2015) Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science*, **347**, 1138-1142.
11. Zhang, Y., Chen, K., Sloan, S.A., Bennett, M.L., Scholze, A.R., O'Keeffe, S., Phatnani, H.P., Guarnieri, P., Caneda, C., Ruderisch, N.*, et al.* (2014) An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J Neurosci*, **34**, 11929-11947.
12. Cooper, G.M., Stone, E.A., Asimenos, G., Program, N.C.S., Green, E.D., Batzoglou, S., Sidow, A. (2005) Distribution and intensity of constraint in mammalian genomic sequence. *Genome research*, **15**, 901-913.
13. Weirauch, M.T., Yang, A., Albu, M., Cote, A.G., Montenegro-Montero, A., Drewe, P., Najafabadi, H.S., Lambert, S.A., Mann, I., Cook, K. (2014) Determination and inference of eukaryotic transcription factor sequence specificity. *Cell*, **158**, 1431-1443.
14. Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., Greenleaf, W.J. (2013) Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature methods*, **10**, 1213-1218.
15. Neph, S., Vierstra, J., Stergachis, A.B., Reynolds, A.P., Haugen, E., Vernot, B., Thurman, R.E., John, S., Sandstrom, R., Johnson, A.K. (2012) An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature*, **489**, 83-90.
16. Sherwood, R.I., Hashimoto, T., O'Donnell, C.W., Lewis, S., Barkal, A.A., van Hoff, J.P., Karun, V., Jaakkola, T., Gifford, D.K. (2014) Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape. *Nat Biotechnol*, **32**, 171-178.
17. Seo, H.-C., Sætre, B.O., Håvik, B., Ellingsen, S., Fjose, A. (1998) The zebrafish Pax3 and Pax7 homologues are highly conserved, encode multiple isoforms and show dynamic segment-like expression in the developing brain. *Mechanisms of development*, **70**, 49-63.

18. Baldwin, C.T., Hoth, C.F., Macina, R.A., Milunsky, A. (1995) Mutations in PAX3 that cause Waardenburg syndrome type I: ten new mutations and review of the literature. *American journal of medical genetics*, **58**, 115-122.
19. Su, Y., Shin, J., Zhong, C., Wang, S., Roychowdhury, P., Lim, J., Kim, D., Ming, G.L., Song, H. (2017) Neuronal activity modifies the chromatin accessibility landscape in the adult brain. *Nat Neurosci*,
20. Lai, H.C., Meredith, D.M., Johnson, J.E. (2013) Chapter 18 - bHLH Factors in Neurogenesis and Neuronal Subtype Specification A2 - Rubenstein, John L.R. In: Rakic P, editor. *Patterning and Cell Type Specification in the Developing CNS and PNS*. Oxford: Academic Press, pp 333-354.
21. Qu, K., Zaba, L.C., Giresi, P.G., Li, R., Longmire, M., Kim, Y.H., Greenleaf, W.J., Chang, H.Y. (2015) Individuality and variation of personal regulomes in primary human T cells. *Cell Systems*, **1**, 51-61.
22. Sun, Y., Nadal-Vicens, M., Misono, S., Lin, M.Z., Zubiaga, A., Hua, X., Fan, G., Greenberg, M.E. (2001) Neurogenin promotes neurogenesis and inhibits glial differentiation by independent mechanisms. *Cell*, **104**, 365-376.
23. Lee, H.K., Yang, Y., Su, Z., Hyeon, C., Lee, T.S., Lee, H.W., Kweon, D.H., Shin, Y.K., Yoon, T.Y. (2010) Dynamic Ca2+-dependent stimulation of vesicle fusion by membrane-anchored synaptotagmin 1. *Science*, **328**, 760-763.
24. PGC-SCZ (2014) Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, **511**, 421-427.
25. Lambert, J.C., Ibrahim-Verbaas, C.A., Harold, D., Naj, A.C., Sims, R., Bellenguez, C., DeStafano, A.L., Bis, J.C., Beecham, G.W., Grenier-Boley, B.*, et al.* (2013) Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nature genetics*, **45**, 1452-1458.
26. Liu, J.Z., van Sommeren, S., Huang, H., Ng, S.C., Alberts, R., Takahashi, A., Ripke, S., Lee, J.C., Jostins, L., Shah, T.*, et al.* (2015) Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nature genetics*, **47**, 979-986.
27. Okada, Y., Wu, D., Trynka, G., Raj, T., Terao, C., Ikari, K., Kochi, Y., Ohmura, K., Suzuki, A., Yoshida, S.*, et al.* (2014) Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature*, **506**, 376-381.
28. Consortium, C.D. (2015) A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nature genetics*, **47**, 1121-1130.
29. Global Lipids Genetics, C., Willer, C.J., Schmidt, E.M., Sengupta, S., Peloso, G.M., Gustafsson, S., Kanoni, S., Ganna, A., Chen, J., Buchkovich, M.L.*, et al.* (2013) Discovery and refinement of loci associated with lipid levels. *Nature genetics*, **45**, 1274-1283.
30. Pickrell, J.K. (2014) Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *American journal of human genetics*, **94**, 559-573.
31. Fromer, M., Roussos, P., Sieberts, S.K., Johnson, J.S., Kavanagh, D.H., Perumal, T.M., Ruderfer, D.M., Oh, E.C., Topol, A., Shah, H.R.*, et al.* (2016) Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat Neurosci*, **19**, 1442-1453.
32. Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M.R., Powell, J.E., Montgomery, G.W., Goddard, M.E., Wray, N.R., Visscher, P.M.*, et al.* (2016) Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nature genetics*, **48**, 481-487.
33. Insel, T.R. (2010) Rethinking schizophrenia. *Nature*, **468**, 187-193.
34. Fromer, M., Pocklington, A., Kavanagh, D., Williams, H., Dwyer, S., Gormley, P., Georgieva, L., Rees, E., Palta, P., Ruderfer, D.*, et al.* (2014) De novo mutations in schizophrenia implicate synaptic networks. *Nature*, **506**, 179-184.
35. Purcell, S.M., Moran, J.L., Fromer, M., Ruderfer, D., Solovieff, N., Roussos, P., O/'Dushlaine, C., Chambert, K., Bergen, S.E., Kahler, A.*, et al.* (2014) A polygenic burden of rare disruptive mutations in schizophrenia. *Nature*, **506**, 185-190.
36. Gjoneska, E., Pfenning, A.R., Mathys, H., Quon, G., Kundaje, A., Tsai, L.H., Kellis, M. (2015) Conserved epigenomic signals in mice and humans reveal immune basis of Alzheimer's disease. *Nature*, **518**, 365-369.

37. Raj, T., Rothamel, K., Mostafavi, S., Ye, C., Lee, M.N., Replogle, J.M., Feng, T., Lee, M., Asinovski, N., Frohlich, I.*, et al.* (2014) Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. *Science*, **344**, 519-523.
38. Kozlenkov, A., Roussos, P., Timashpolsky, A., Barbu, M., Rudchenko, S., Bibikova, M., Klotzle, B., Byne, W., Lyddon, R., Di Narzo, A.F.*, et al.* (2014) Differences in DNA methylation between human neuronal and glial cells are concentrated in enhancers and non-CpG sites. *Nucleic Acids Research*, **42**, 109-127.
39. Raivich, G., Behrens, A. (2006) Role of the AP-1 transcription factor c-Jun in developing, adult and injured brain. *Progress in neurobiology*, **78**, 347-363.
40. Ross, S.E., Greenberg, M.E., Stiles, C.D. (2003) Basic Helix-Loop-Helix Factors in Cortical Development. *Neuron*, **39**, 13-25.
41. Monsoro-Burq, A.H. (2015) PAX transcription factors in neural crest development. *Seminars in Cell & Developmental Biology*, **44**, 87-96.
42. Reiprich, S., Wegner, M. (2015) From CNS stem cells to neurons and glia: Sox for everyone. *Cell and tissue research*, **359**, 111-124.
43. Wang, Z.-X., Kueh, J.L.L., Teh, C.H.-L., Rossbach, M., Lim, L., Li, P., Wong, K.-Y., Lufkin, T., Robson, P., Stanton, L.W. (2007) Zfp206 Is a Transcription Factor That Controls Pluripotency of Embryonic Stem Cells. *STEM CELLS*, **25**, 2173-2182.
44. Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., Yamanaka, S. (2003) The Homeoprotein Nanog Is Required for Maintenance of Pluripotency in Mouse Epiblast and ES Cells. *Cell*, **113**, 631-642.
45. Thomson, M., Liu, Siyuan J., Zou, L.-N., Smith, Z., Meissner, A., Ramanathan, S. (2011) Pluripotency Factors in Embryonic Stem Cells Regulate Differentiation into Germ Layers. *Cell*, **145**, 875-889.
46. Ziller, M.J., Edri, R., Yaffe, Y., Donaghey, J., Pop, R., Mallard, W., Issner, R., Gifford, C.A., Goren, A., Xing, J.*, et al.* (2015) Dissecting neural differentiation regulatory networks through epigenetic footprinting. *Nature*, **518**, 355-359.
47. Kraus, P., V, S., Yu, H.B., Xing, X., Lim, S.L., Adler, T., Pimentel, J.A.A., Becker, L., Bohla, A., Garrett, L.*, et al.* (2014) Pleiotropic Functions for Transcription Factor Zscan10. *PLoS ONE*, **9**, e104568.
48. Kundakovic, M., Jiang, Y., Kavanagh, D.H., Dincer, A., Brown, L., Pothula, V., Zharovsky, E., Park, R., Jacobov, R., Magro, I.*, et al.* (2016) Practical Guidelines for High-Resolution Epigenomic Profiling of Nucleosomal Histones in Postmortem Human Brain Tissue. *Biological psychiatry*,
49. Yuan, P., Zhou, R., Wang, Y., Li, X., Li, J., Chen, G., Guitart, X., Manji, H.K. (2010) Altered levels of extracellular signal-regulated kinase signaling proteins in postmortem frontal cortex of individuals with mood disorders and schizophrenia. *J Affect Disord*, **124**, 164-169.
50. Kyosseva, S.V., Elbein, A.D., Griffin, W.S., Mrak, R.E., Lyon, M., Karson, C.N. (1999) Mitogen-activated protein kinases in schizophrenia. *Biological psychiatry*, **46**, 689-696.
51. Reynolds, C.A., Hong, M.G., Eriksson, U.K., Blennow, K., Wiklund, F., Johansson, B., Malmberg, B., Berg, S., Alexeyenko, A., Gronberg, H.*, et al.* (2010) Analysis of lipid pathway genes indicates association of sequence variation near SREBF1/TOM1L2/ATPAF2 with dementia risk. *Hum Mol Genet*, **19**, 2068-2078.
52. Gnanapragasam, M.N., Scarsdale, J.N., Amaya, M.L., Webb, H.D., Desai, M.A., Walavalkar, N.M., Wang, S.Z., Zu Zhu, S., Ginder, G.D., Williams, D.C., Jr. (2011) p66Alpha-MBD2 coiled-coil interaction and recruitment of Mi-2 are critical for globin gene silencing by the MBD2-NuRD complex. *Proc Natl Acad Sci U S A*, **108**, 7487-7492.
53. Lin, X., Shah, S., Bulleit, R.F. (1996) The expression of MEF2 genes is implicated in CNS neuronal differentiation. *Brain Res Mol Brain Res*, **42**, 307-316.
54. Yao, P.J., Coleman, P.D., Calkins, D.J. (2002) High-resolution localization of clathrin assembly protein AP180 in the presynaptic terminals of mammalian neurons. *J Comp Neurol*, **447**, 152-162.
55. Wu, F., Mattson, M.P., Yao, P.J. (2010) Neuronal activity and the expression of clathrin-assembly protein AP180. *Biochem Biophys Res Commun*, **402**, 297-300.
56. Ryan, T.A. (2006) A pre-synaptic to-do list for coupling exocytosis to endocytosis. *Curr Opin Cell Biol*, **18**, 416-421.

57. Schwartz, C.M., Cheng, A., Mughal, M.R., Mattson, M.P., Yao, P.J. (2010) Clathrin assembly proteins AP180 and CALM in the embryonic rat brain. *J Comp Neurol*, **518**, 3803-3818.
58. Kohn, A.D., Moon, R.T. (2005) Wnt and calcium signaling: beta-catenin-independent pathways. *Cell Calcium*, **38**, 439-446.
59. Goes, F.S., Hamshere, M.L., Seifuddin, F., Pirooznia, M., Belmonte-Mahon, P., Breuer, R., Schulze, T., Nothen, M., Cichon, S., Rietschel, M.*, et al.* (2012) Genome-wide association of mood-incongruent psychotic bipolar disorder. *Transl Psychiatry*, **2**, e180.
60. Yao, J., Gaffaney, J.D., Kwon, S.E., Chapman, E.R. (2011) Doc2 is a Ca2+ sensor required for asynchronous neurotransmitter release. *Cell*, **147**, 666-677.
61. Groffen, A.J., Friedrich, R., Brian, E.C., Ashery, U., Verhage, M. (2006) DOC2A and DOC2B are sensors for neuronal activity with unique calcium-dependent and kinetic properties. *J Neurochem*, **97**, 818-833.
62. Dickerson, L.W., Bonthius, D.J., Schutte, B.C., Yang, B., Barna, T.J., Bailey, M.C., Nehrke, K., Williamson, R.A., Lamb, F.S. (2002) Altered GABAergic function accompanies hippocampal degeneration in mice lacking ClC-3 voltage-gated chloride channels. *Brain Res*, **958**, 227-250.
63. Yoshikawa, M., Uchida, S., Ezaki, J., Rai, T., Hayama, A., Kobayashi, K., Kida, Y., Noda, M., Koike, M., Uchiyama, Y.*, et al.* (2002) CLC-3 deficiency leads to phenotypes similar to human neuronal ceroid lipofuscinosis. *Genes Cells*, **7**, 597-605.
64. Riazanski, V., Deriy, L.V., Shevchenko, P.D., Le, B., Gomez, E.A., Nelson, D.J. (2011) Presynaptic CLC-3 determines quantal size of inhibitory transmission in the hippocampus. *Nat Neurosci*, **14**, 487-494.
65. Guzman, R.E., Alekov, A.K., Filippov, M., Hegermann, J., Fahlke, C. (2014) Involvement of ClC-3 chloride/proton exchangers in controlling glutamatergic synaptic strength in cultured hippocampal neurons. *Front Cell Neurosci*, **8**, 143.
66. Kozlenkov, A., Wang, M., Roussos, P., Rudchenko, S., Barbu, M., Bibikova, M., Klotzle, B., Dwork, A.J., Zhang, B., Hurd, Y.L.*, et al.* (2015) Substantial DNA methylation differences between two major neuronal subtypes in human brain. *Nucleic Acids Research*,
67. Robinson, M.D., McCarthy, D.J., Smyth, G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139-140.

FIGURE LEGENDS

Figure 1. Differential analysis of chromatin accessibility in neuronal and non-neuronal cells

(A) Schematic outline of study design. Briefly, nuclei isolated from frozen frontopolar prefrontal cortex specimens were separated in to NeuN+ (neuronal) and NeuN- (non-neuronal) fractions by FANS. 50,000 nuclei from each cell population were then subjected to transposition reactions followed by library amplification and sequencing (50bp paired-end reads). **(B)** Unsupervised hierarchical clustering of ATAC-seq data. **(C)** Volcano plot showing the distribution of -log10 p-value and log2 fold-change of differential chromatin accessibility analysis across 115,021 OCRs. **(D)** Distribution of log2 fold-change of differential chromatin accessibility analysis for 33,054 neuronal and 27,599 non-neuronal OCRs. Averaged cell-type and gender-specific ATAC-seq signals at **(E)** *CADM3* and **(F)** *TGIF1* gene loci. Positive and negative logFC indicate neuronal and non-neuronal differential signals, respectively. Red box indicates the significant differentially accessible region in **(E)** neuronal and **(F)** non-neuronal cells.

Figure 2. Annotation of the neuronal and non-neuronal regulome

(A) Average read count frequency of OCRs in TSS regions. Confidence interval estimated by bootstrap method. **(B)** Distribution of all peaks and differential OCRs relative to TSS. **(C)** Distribution of genomic features of all and differential OCRs. **(D)** Average GERP score as a function of distance from the center of [-1000bp, 1000bp] of all peaks and differential OCRs. Curves and their 95% confidence intervals are calculated on a 50 bp sliding window. (**E)** Enrichment of all peaks and differential OCRs in various human prefrontal brain tissue chromatin states. **(F)** Gene Ontology terms for differential OCRs enriched in neuronal (dark red) and non-neuronal (dark green) samples. **(G)** Enrichment of differential OCRs for neuronal (dark red) and non-neuronal (dark green) cell type-specific markers.

Figure 3. Transcription factor and gene regulome in neuronal and non-neuronal cells

(A) The average transposase insertion probability at all predicted binding CTCF binding sites within neuronal

and non-neuronal OCRs. **(B)** Motif scores in neuronal and non-neuronal cells. Based on the ratio of the motif scores between the two samples, the cell-type specificity of the motifs is categorized as probable (≥1.25) and definite (≥1.5). Motifs with an enrichment of matches within motifs compared to the genomic background of less than 1.25 were categorized with "no enrichment". Based on the relative occurrence of motifs in promoter peaks compared to all peaks, the motifs were categorized as promoter depleted (≤1/1.25) or enriched (≥1.25). We highlight the most enriched motifs for all TFs (Left panel). A zoomed version (Gray box) is illustrated on the right panel. Here, the most cell-type specific TFs are highlighted, listing only the most significant TF of each TF-family. **(C)** Two CTCF predicted sites near the 3' and 5' ends of the *TRPM3* gene were detected only in neuronal cells. The exact location of the CTCF binding motif for the 5'-CTCF and 3'-CTCF binding sites is illustrated on the right. **(D)** The regulatory divergence of TF targets between the neuronal and non-neuronal samples. **(E)** The regulatory divergence of genes between the neuronal and non-neuronal samples. **(F)** Gene Ontology terms enriched in the top 1000 most divergently regulated genes.

Figure 4. Integration of OCRs and TF binding sites with SCZ risk variants and *in vitro* validation

(A) Enrichment of predicted TF sites within SCZ risk loci from neuronal and non-neuronal cells using a single TF model (top panel) and a joint model (bottom panel). The maximum-likelihood estimates and 95% confidence intervals of the enrichment parameter for each TF is illustrated. Annotations are ranked based on the improvement of the likelihood of the model (at the top are those that improved the likelihood the most). The number in parenthesis is the likelihood of the model for each TF (single model) or the combined model. **(B)** Regional plot surrounding the *SNX19* locus. The top panel shows a plot of the r^2 and P values for association with SCZ, including the index (rs10791097) and the functional (rs10750450) SNP. In the middle panel (Prior TF) is the fitted empirical prior probability based on the TF combined model and the positions of the TFs for this region. The overlap of the SNP with the binding sites of ZSCAN10 and ZNF354C are illustrated. In the lower panel (Prior OCR) is the fitted empirical prior probability and position of the OCR model. Note the difference in the fitted empirical prior probability among the TF and OCR model for the functional SNP (in gray box). **(C)** Examining the effect of rs10750450 risk-T and reference-G alleles and the binding motifs of

ZNF354C and ZSCAN10 on transcriptional activity in *in vitro* experiments. Top panel is a graphical representation of the rs10750450 risk locus. All constructs result in increased luciferase activity in comparison to empty pGL4.24 vector. Compared to the reference G allele, the risk T allele enhances luciferase expression, as does excision of the ZNF354C or ZSCAN10 binding site. Excision of the rs10750450 nucleotide gives rise to a further increase of luciferase expression and a similar effect is observed upon simultaneous excision of both ZNF354C and ZSCAN10 binding sites.

ABBREVIATIONS

Assay for Transposase Accessible Chromatin followed by sequencing (ATAC-seq)

Basic helix-loop-helix (bHLH)

Bayesian Information Criterion (BIC)

Bipolar (BP)

Brodmann area (BA)

CADM3 (Cell adhesion molecule 3)

chomatin immunoprecipitation sequencing (ChIPseq)

cis regulatory element (CRE)

CommonMind Consortium (CMC)

DNase hypersensitive sites (DHS)

embryonic stem cells (ESCs)

expression quantitative trait loci (eQTLs)

false discovery rate (FDR)

fluorescence-activated nuclear sorting (FANS)

fold change (FC)

fraction of reads in peaks (FRiP)

Genomic Evolutionary Rate Profiling (GERP)

Genome Wide Association Study (GWAS)

Human embryonic kidney 293 cells (HEK-293)

linkage disequilibrium (LD)

major depressive disorder (MDD)

Model-based Analysis of ChIP-seq (MACS)

negative binomial (NB)

open chromatin region (OCR)

PCR bottleneck coefficient (PBC)

Pearson Correlation Coefficient *R* (PCC)

Protein Interaction Quantification (PIQ)

relative strand cross-correlation coefficient (RSC)

Roadmap Epigenomics Mapping Consortium (REMC)

schizophrenia (SCZ)

Single Nucleotide Polymorphism (SNP)

Transcription factor (TF)

TGIF1 (TGFB-Induced Factor Homeobox 1)

transcription start site (TSS)

TRPM3 (Transient Receptor Potential Cation Channel, Subfamily M, Member 3)

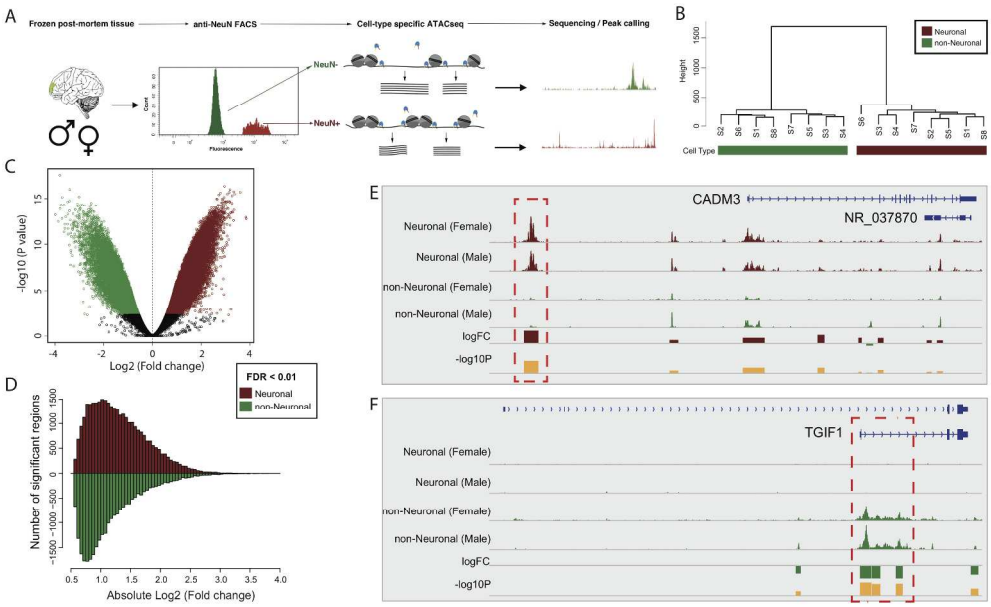

Figure 1 Differential analysis of chromatin accessibility in neuronal and non-neuronal cells

360x219mm (300 x 300 DPI)

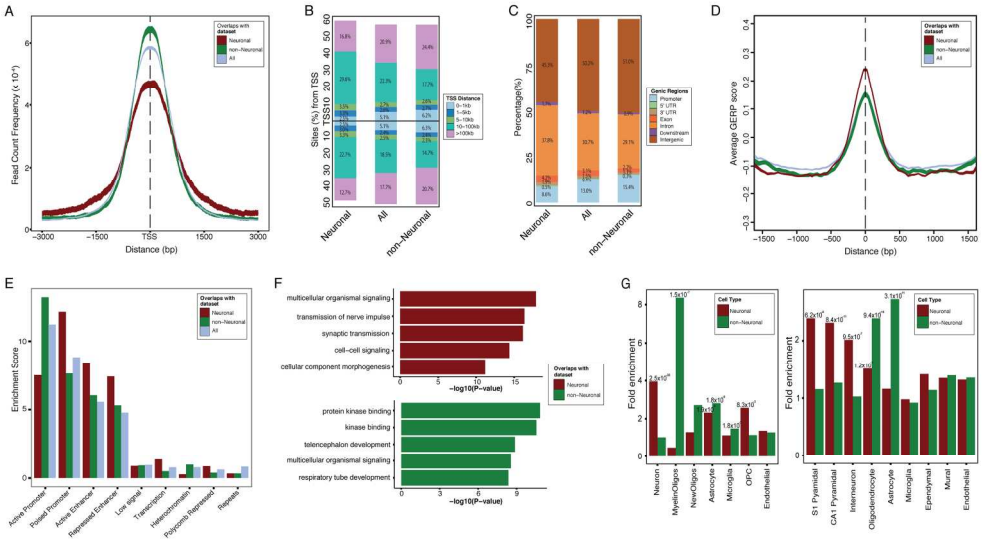

Figure 2 Annotation of the neuronal and non-neuronal regulome

215x118mm (300 x 300 DPI)

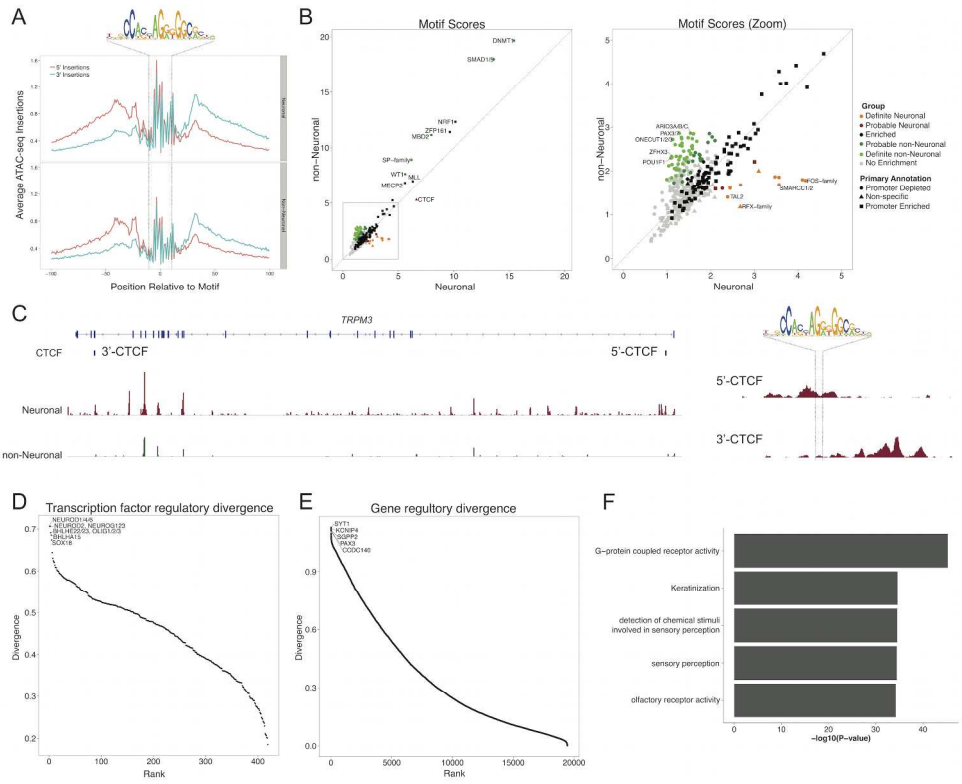

Figure 3.
Transcription factor and gene regulome in neuronal and non-neuronal cells

254x205mm (300 x 300 DPI)

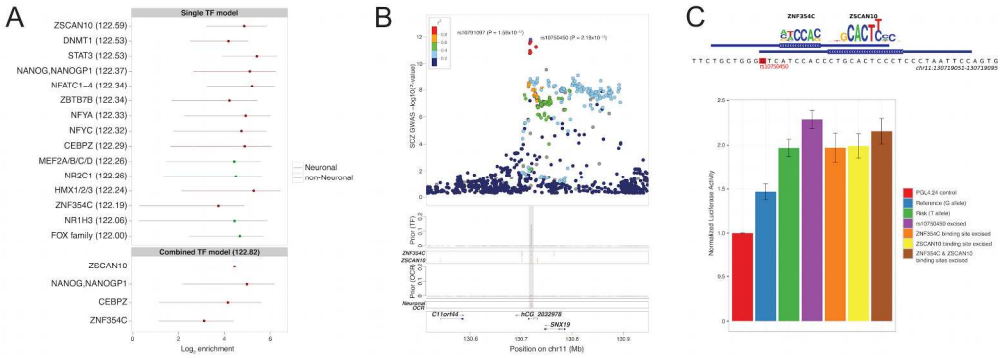

Figure 4 Integration of OCRs and TF binding sites with SCZ risk variants and in vitro validation

254x90mm (300 x 300 DPI)