Differential Chromatin Architecture and Risk Variants in Deep Layer Excitatory Neurons and Grey Matter Microglia Contribute to Major Depressive Disorder

Anjali Chawla[1,2], Doruk Cakmakci[3], Wenmin Zhang[4], Malosree Maitra[1,2], Reza Rahimian[1], Haruka Mitsuhashi[1,2], MA Davoli[1], Jenny Yang[1], Gary Gang Chen[1], Ryan Denniston[1], Deborah Mash[5], Naguib Mechawar[1], Matthew Suderman[6,7], Yue Li[3,4]*, Corina Nagy[1,2]*, Gustavo Turecki[1,2]*#

1. McGill Group for Suicide Studies, Douglas Institute, Department of Psychiatry, McGill University, Montreal, QC, Canada
2. Integrated Program in Neuroscience, McGill University, Montreal, QC, Canada
3. School of Computer Science, McGill University, Montreal, QC, Canada
4. Quantitative Life Sciences, McGill University, Montreal, QC, Canada
5. Dr. Kiran C. Patel College of Allopathic Medicine, Nova Southeastern University, Ft. Lauderdale, FL, USA
6. Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK
7. MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK

* Shared senior authorship

Corresponding authorship

Abstract

Major depressive disorder (MDD) associated genetic variants reside primarily in the non-coding, regulatory genome. Here we investigate genome-wide regulatory differences and putative gene-regulatory effects of disease risk-variants by examining chromatin accessibility combined with single-cell gene-expression profiles in over 200,000 cells from the dorsolateral prefrontal cortex (DLPFC) of 84 individuals with MDD and neurotypical controls. MDD-associated accessibility alterations were prominent in deep-layer excitatory neurons characterized by transcription factor (TF) motif accessibility and binding of nuclear receptor (NR)4A2, an activity-dependent TF responsive to pathological stress. The same neurons were significantly enriched for MDD-associated genetic variation disrupting cis-regulatory sites and TF binding associated with genes involved in synaptic communication. Furthermore, a grey matter microglial cluster exhibited differentially closed chromatin in MDD affecting binding sites bound by TFs known to regulate immune homeostasis. In summary, our study points to specific cell types and regulatory mechanisms whereby genetic variation may increase predisposition to MDD.

Introduction

Major depressive disorder (MDD) is a debilitating and life-threatening psychiatric disorder affecting almost 5% of the world's population [1,2]. Many neurobiological factors and cellular alterations have been previously associated with MDD including changes in monoaminergic and glutamatergic systems [3-5], and abnormalities in astrocytes [6,7], oligodendrocytes [8], and immune cells [9,10]. In spite of progress, the precise molecular mechanisms mediating risk for MDD remain unknown.

Genome-wide associations studies (GWAS) have identified over 200 MDD risk loci [11-13]; however, similarly to other psychiatric phenotypes, characterizing the functional impact of such variants has remained challenging as they mostly map outside protein-coding regions. In addition to genetic variation, the adverse effects of environmental factors, such as early-life stress, on increasing the risk for depression [14,15], are mediated by regulatory changes[16]. Single-cell atlases

of open chromatin cis-regulatory elements (cCREs) holds promise to uncover mechanisms whereby disease-associated risk-variants may impact cell-specific gene-regulatory targets [17,18].

Previously, using single-nucleus RNA-seq (snRNA-seq), our group identified gene-expression changes mainly in the deep-layer excitatory neurons, and certain glial subtypes, including microglia in individuals with MDD [19,20.] Here, we bridge current knowledge gaps by employing single-nucleus assay for transposase-accessible chromatin with sequencing (snATAC-seq)[21], which provides unprecedented opportunities for interrogating epigenetic and genetic variants through the lens of chromatin accessibility [17,18,22]. By measuring chromatin accessibility in more than 200,000 cells from the DLPFC of 84 individuals and integrating snRNA-seq obtained from the same samples [19,20], we generated a large cell-type specific accessibility and expression atlas of cortical cells in MDD.

Our results revealed that MDD-associated chromatin accessibility alterations were most prominent in NR4A2+ deep-layer excitatory neurons (ExN1) and in a grey matter (GM) microglial cluster (Mic2). ExN1 neurons were also most significantly enriched for the heritability of MDD GWAS SNPs. Examining allele-specific effects of MDD-associated genetic variation in these neurons identified TFs and genes likely affecting synaptic communication and plasticity. Together, our single-nucleus multi-modal dataset provides an in-depth examination of MDD-associated genetic variation and downstream gene regulatory mechanisms with cell type specificity.

Results

Chromatin accessibility in human dorsolateral pre-frontal cortical (DLPFC) cells.

We performed snATAC-seq in 44 individuals who died during an episode of MDD and 40 age- and sex-matched neurotypical controls (Fig 1A, schematic; demographic and sample characteristics; Supplementary Table 1). Male and female nuclei were combined prior to microfluidic-capture, common variants, and chromatin accessibility measured in this dataset permitted accurate demultiplexing of pooled subjects (Supplementary Fig 1; Supplementary note). Filtering of doublets and low-quality barcodes revealed 201,456 high-quality nuclei having overall high TSS enrichment score (median: 7) and uniquely mapped fragments (median:

15,833) with equal contributions from both sexes (51% female cells) and conditions (53% case cells; Supplementary Figure 2-4).

Unsupervised iterative clustering of nuclei by genome-wide chromatin accessibility measured at 500bp resolution [23] revealed 7 broad cell-types (ExN: 35.2%, InN: 13.4%, Ast: 6.8%, Oli: 36.8%, OPC: 3.8%, Mic: 2.4%, End: 1.6%) consisting of 38 distinct clusters (Fig 1B; Extended Fig 1D). Multiple approaches for assigning cell annotations confirmed snATAC-seq cell-type and cluster identities (Fig 1C-E, Extended Fig 1-2), including label-transfer from sample-matched snRNA-seq [20] (median: 0.94; Extended Fig 1A-E) and published PFC snATAC-seq [24] (median: 0.99; Extended Fig 2A-B).

The clusters were characterized by cluster-specific accessibility across 239,824 cCREs, (hereafter called marker cCREs; Fig 1G; Wilcoxon test, FDR<0.05; Extended Fig 1I (top); Supplementary Table 2) and cluster-specific TF motif accessibility and binding (Wilcoxon test, FDR<0.05, Fig 1H-I, Supplementary Table 3). To identify gene-regulatory mechanisms in cortical cells, we examined correlations between cCRE accessibility and gene-expression (peak-to-gene linkages), identifying 117,328 unique cCREs highly correlated (r>0.45) with 11,575 expressed genes (Supplementary Table 4). As expected, almost all (>90%) of these cCREs overlapped histone markers of transcriptionally active chromatin from cortical cells [25] (Extended Fig 1I (bottom)) and captured gene-regulatory variations from each cell-type (Extended Fig 1C).

MDD-associated chromatin differences found in microglia and deep-layer neurons.

We investigated MDD-associated accessibility within clusters and cell-types using per-subject pseudo-bulked accessibility estimates (Method: Differential accessibility analysis; Supplementary Table 5). Of 38 characterized clusters, 25 clusters were associated with a total of 4751 MDD-associated differentially accessible regions (hereafter called DARs; Fig 2A), whereas only 297 differential chromatin regions were found at broad level (Supplementary figure 5A-C). Most DARs were less accessible in individuals with MDD (76%; Fig 2B) and were primarily observed in clusters of microglia Mic2 (58.5%) and excitatory neurons ExN1 (21.2%). The number of cells in these clusters were comparable to all others (except Oli4; interquartile range

(IQR)) and depicted uniform contributions from potential covariates (Fig 1F; Supplementary Fig 4).

To understand the likely regulatory impact of DARs, we asked how they overlapped with PsychEncode ChIP-seq histone marks from cortical cells [25]. Nearly all of the DARs less accessible in MDD (98%) coincided with histone marks that are typically found in gene promoters (defined by H3K4me3+H3K27ac; 74%) and enhancers (H3K27ac, 23%). By contrast, most DARs more accessible in MDD (85%) overlapped with enhancers (H3K27ac; 56%) more than promoters (28%; Fig 2B).

To understand the overall functional impact of DARs, we identified the most likely affected genes (n= 2254) using peak-to-gene linkages (r>0.45) and investigated their known functions. Of these genes, the highest psychiatric disease association was found for depression (Fig 2C), whereby 80% of these associations showed consistent evidence in PsyGeNet [26] and 64% for schizophrenia. In fact, there was an overrepresentation among DAR-linked genes (FDR < 0.05; one-sided Fisher's exact test; Fig 2D) for PsyGeNet genes linked to depression, MDD-associated genetic variation [11,13] and genes previously observed to be differentially expressed in MDD [20]. We did not observe similar enrichments for genes associated with cell-type DARs observed at broad level.

Cell type-specific roles of proximal and distal gene-regulation have been previously reported. For example, activity-dependent regulation in neurons involves many distal regulatory elements [27] while microglial-activation involves mostly proximal regulation [28]. Here we similarly observed that whereas DARs in excitatory neurons tended to be distal to any TSS (e.g. 77% of DARs in ExN1 were > 2kbp from a TSS; Fig 2A), DARs in microglia tended to be proximal to a TSS (e.g. 73% of DARs in Mic2 were within 2kbp of a TSS).

Given that ExN1 and Mic2 were the most affected clusters in MDD, we focused on these to examine the relationship between altered chromatin and variants associated with gene regulation. Both ExN1 and Mic2 DARs showed significant enrichment for splicing (sQTLs) and expression quantitative trait loci (eQTLs; Fig2E-F inset), implicating potential impact of DARs on DLPFC gene expression. Moreover, using snRNA-seq clusters[20], we identified significantly increased

module scores in MDD for expression of genes linked to ExN1 DARs showing specifically increased accessibility in MDD ($p = 0.0009$, Wilcoxon rank-sum test; Extended figure 2F), and we observed tendency for the opposite in Mic2 ($p=0.1$).

Finally, to elucidate functional pathways likely affected by MDD-associated chromatin, we examined cluster-specific DAR-linked gene ontologies (GO; Supplementary Table 6). ExN1 DARs were largely more accessible in cases (73% of DARs) and linked to genes involved in modulating synaptic and neurotransmitter release functions (FDR<0.05; Fig 2H). Whereas Mic2 DARs with significantly decreased accessibility in MDD linked to genes enriched in phagocytosis- and immune-signaling pathways (FDR<0.05; Fig 2I; Supplementary figure 5D).

TF binding sites overrepresented in MDD-associated chromatin.

TF binding can both regulate and be affected by chromatin accessibility leading to gene-expression [29]. In order to determine the possible interplay between TFs and transcriptional effects, we first identified TF binding motifs enriched in cluster-specific DARs compared to GC-matched background peaks (Homer [30]; $p<0.05$, Fig 3A, Supplementary table 7). TF binding motifs highly enriched (FDR<0.05) in ExN1 DARs belonged to basic leucine zipper (bZIP), basic helix–loop–helix (bHLH), Zinc finger (Zf), and helix-turn-helix (HTH) families, which are known to regulate activity-dependent [31] and neurodevelopment-associated functions [32,33].

On the other hand, TF motifs enriched (FDR<0.05) in Mic2 DARs included many canonical microglial factors, such as ETS domain (SPI1) and interferon regulatory factors (IRFs), some of which are pioneering and lineage-specifying TFs that can regulate local chromatin accessibility [34,35] and can affect microglial development [36], proliferation [37], and immune-activation[34,38].

To confirm these observations, we assessed GC and accessibility bias-corrected TF motif deviations across motif-matching cCREs in each cluster (ChromVar [39]; Wilcoxon test; FDR<0.05). These results suggested an overall agreement between MDD-associated TF motifs enriched in differentially more or less accessible DARs among MDD individuals and those

identified comparing motif accessibility signals between MDD and controls at single-cell resolution (Supplementary Table 8).

TFs mediate effects of MDD-associated chromatin on gene expression.

We hypothesized that reduced chromatin accessibility in MDD at TF binding sites would reduce TF binding and result in reduced expression of target genes, and vice versa. We investigated this hypothesis by identifying a set of TFs whose binding was most likely to be affected by MDD-associated accessibility, compiling a list of their most likely target genes, and then examining if the expression of these genes was associated with MDD.

To identify high-priority TFs affected by differential accessibility, we selected TFs with motifs enriched in DARs that were also highly expressed in integrated snRNA-seq. Then, among these, we further restricted selection to those with the greatest MDD-associated accessibility differences at their binding sites in cases versus controls (Fig 3B; Supplementary Figure 5E).

We identified the likely targets of these prioritized TFs by constructing gene regulatory networks using igraph (Fig 3C; Method: Gene regulatory networks) in ExN1 (BACH2, ASCL1, CTCF, RFX3) and Mic2 (SP1, SPI1, IRF1, ATF7). The nodes in green represent non-coding DARs with individual TF binding motifs connected to their putative target-genes colored according to MDD DEGs or MDD-associated genetic variation and PsyGeNET depressive disorder gene status.

Finally, we calculated MDD-associated module scores for these networks using snRNA-seq clusters[20]. In general, these scores confirmed that increased binding site accessibility was associated with increased module expression levels, and vice versa (Fig 3D-F).

Strong enrichment for MDD heritability in deep-layer excitatory neurons.

To prioritize cell type clusters that are likely to mediate the heritable risk for MDD, we calculated GWAS SNP heritability enrichment within cluster-specific chromatin using stratified

linkage-disequilibrium score regression (S-LDSC; Method: Heritability enrichment) [40,41]. For this, we used GWAS summary statistics available for MDD and related psychiatric phenotypes and non-brain related control traits. The heritable risk for psychiatric disorders and traits that are known to be genetically correlated with MDD [42] were significantly enriched in marker cCREs of ExN1 and ExN2 clusters (Fig 4A). Similarly to ExN1, ExN2 showed similarities to deep-layer (layer 5-6) excitatory neurons (Extended Fig 2A, 3C). In fact, enrichments calculated from data obtained from both MDD GWAS used in our study[11,13] were coherently highest in ExN1 cluster (Fig 4A; p=0.003 (Howard), p=0.0005 (Als), Supplementary Table 9). We also calculated heritability enrichments in MDD-associated chromatin in each cluster (candidate DARs), showing that the enrichment was likewise highest in ExN1 for MDD (Fig 4B; p=0.01; nominal) and insomnia (p=0.02).

To determine if these enrichments were driven by linkage disequilibrium (LD), we first estimated the heritability of MDD-associated variants in heritable LD blocks defined by 1000 Genome European (EUR) population using heritability estimator from summary statistics (HESS) [43]. This resulted in about 5% of LD blocks significantly enriched for MDD-associated genetic variants ($p<0.05$, Supplementary table 10). We then examined if MDD-associated blocks contained an overrepresentation of candidate DARs within each cluster compared to chromosome-matched cCREs in any LD block (Permutation testing; nPerm=10,000). Consistent with previous observations, the most significant enrichment was observed in ExN1 (p=0.007; Fig 4C).

Finally, to identify regulatory TFs associated with chromatin enriched for MDD heritability, we examined TF motifs enriched in ExN1-specific cCREs located within MDD-associated LD blocks compared to GC-matched background peaks. TF families that were significantly enriched ($q<0.05$; Supplementary table 11) in both ExN1 marker and differential chromatin included those associated with activity-dependent functions (bZIP; Fig 4D) [31] and neurodevelopment (bHLH, HTH, Zf) [32,33].

MDD-associated genetic variation shows cell type specific effects on gene regulatory sites.

To examine the functional impact of MDD-associated genetic variation within each cluster, we assessed allelic-effects of MDD-associated SNPs on accessibility of cCREs and TF binding sites and identified likely target risk genes. We used MDD-associated SNPs in GWAS [11-13,] those in high LD with them (r2 value ≥0.8) and those identified by finemapping [11,44]. To predict accessibility in each cluster, we used three complementary models; in-silico mutagenesis (ISM), delta-support vector machine (delta-SVM) [45] and gapped-kmer (gkm-explain) [46] based variant-effect scores derived from cluster-specific gkmSVM classifiers [47,48] (Supplementary Fig 6-7).

In order to determine cell-type and cluster-specific effects of MDD-associated genetic variation, we focused on MDD-associated SNPs overlapping cell-type- or cluster-specific candidate cCREs (referred to as candidate SNPs (cdSNPs) with 14% cdSNPs in candidate DARs and remaining in marker cCREs; Extended Fig 3A-B; Supplementary Table 12). Compared to MDD-associated SNPs, cdSNPs showed significant enrichment for expression quantitative trait loci (eQTLs; $p<1x10^{-3}$; Hypergeometric test) and splicing QTLs (sQTLs; $p<5x10^{-8}$), suggesting an over-representation of MDD SNPs with known effects on DLPFC gene-expression using our candidate cCREs (Extended Fig 3C-D).

Overall, we identified 97 non-coding statistically significant cdSNPs (sSNPs; 15% in candidate DARs; Extended Fig 3B) that were significant across all three gkmSVM models. Notably, 97.9% of MDD sSNPs were exclusively identified in clusters of a single broad cell-type (Extended Fig 3E), suggesting largely cell-type specific impact of genetic variation on chromatin accessibility.

Further, GWAS SNPs often alter expression of distally-located genes [17]. Putatively long-range gene-regulatory effect of genetic variation was determined by identifying gene-expression targets most significantly correlated (up to 500kbps; peak-to-gene linkages) with accessibility of cCREs associated with MDD sSNPs. Both nearest and linked target genes of MDD sSNPs enriched for pathways related to synapse organization and communication (Extended Fig 4A-B; Supplementary Table 13).

Functional impact of MDD-associated genetic variation in ExN1 and Mic2 clusters.

To examine functional effects of MDD sSNPs in clusters that were strongly associated with MDD, we focused on sSNPs showing significant allelic-impact in ExN1 and Mic2 and overlapping cell-type- or cluster-specific candidate cCREs. We highlighted sSNPs that strongly disrupted accessibility in these cluster (medium- to high-confidence) and integrated PsychEncode DLPFC histone modifications [25] and 3D chromatin loops [49] to interpret long-range gene-regulatory targets (Method: Multi-modal Visualization).

Of all MDD sSNPs, 25% (52% s/eQTLs; Fig 5A) mapped to ExN1 marker cCREs or candidate DARs. Together, the target genes of ExN1 sSNPs (nearest, linked, and eQTLs) enriched for pathways related to synaptic communication and plasticity (Extended Fig 4C). This included a lead MDD sSNPs (rs2276138; Fig 5B) in GWAS meta-analysis [50] overlapping ExN1 marker cCRE in intronic region of *SPTBN2* gene and disrupting TF binding site for RFX3, whose motif accessibility was also altered in ExN1 DARs and at single-cell level in MDD (Fig 3B). Moreover, RFX3 motifs enriched in ExN1 cCREs located within heritable LD regions for MDD (Fig 4D) and have been previously associated with neuropsychiatric disorders [33,51]. This sSNP is an eQTL and sQTL for a few genes (Fig 5A) including a GTEx eQTL, *ZDHHC24,* with major regulatory effects on synaptic plasticity [52] and associated through peak-to-gene linkages (r=0.42) with expression of *SYT12*, a synaptic vesicle protein. Additionally, MDD sSNP in *SHANK2* gene (rs11601579; Fig 5C) disrupted binding site accessibility for GLI2 TF, associated with neurodevelopment [53]. This sSNP overlapped ExN1 marker cCRE and significantly correlated with the anti-sense RNA expression of *SHANK2* (r=0.6), a postsynaptic density scaffolding protein linked to neuropsychiatric disorders [54,55]. Finally, about 40% (10/25) of ExN1 sSNPs were found to alter accessibility exclusively in this cluster (Extended Fig 5A-B).

In Mic2, MDD-associated variants with allelic-effects on chromatin accessibility (Fig 6A) included a sSNP mapping to microglia cell-type marker cCRE (rs5995992; Fig 6B) and showing Hi-C linkages to *ARL4C* gene, which regulates lipid homeostasis in immune cells [56]. This sSNP disrupted binding site for FLI1 TF, which is known to be involved in microglia development and homeostasis [36,57] and whose motifs were enriched in Mic2 DARs (Fig 3A). Additionally, a lead MDD GWAS SNP [11,12], which is an e/sQTL for several genes (rs5995992; Fig 6C) mapped to

candidate Mic2 DAR in promoter region of *EP300*, a gene encoding histone acetyl-transferases and is co-activator of NF-κB mediated immune-activation [58].

Molecular characterization of ExN1 and Mic2 clusters in depression.

To relate molecular profiles of clusters with their likely functional roles, we compared promoter-accessibility with gene-expression from the Allen brain institute (ABI) snRNA-seq [59]. ExN1 neurons had the highest similarity to layer 6 (L6) intratelencephalic (IT) Car3 neurons (AUC=0.98; Extended Fig 6A), which are transcriptionally similar to NR4A2-marked neurons [60] implicated in regulating stress-response [61]. Likewise, compared to other excitatory neurons, ExN1 showed the highest TF motif accessibility and binding for NR4A2 (Fig 1H; Extended Fig 6B-C), also known as NURR1, which has been consistently linked to stress-response [62-70], and depression [64,66,71,72]. Moreover, ExN1 marker TF, NR4A2 showed evidence for interactions with MDD-associated ExN1 TFs (Method: StringDB; direct interactions: bZIP: FOS, FOSB, bHLH: ASCL1; indirect: Zf: EGR1; Extended Fig 6D). Together, this presents an intriguing possibility for the role of coordinated TF-activities in NR4A2+ ExN1 neurons in regulating stress-related molecular alterations in depression.

On the other hand, between the two microglial clusters, Mic2 showed a significantly higher enrichment for GM than white matter (WM) microglial genes [73], (Extended Fig 6E-F). Likewise, Mic2 localized mostly to the GM when integrated with DLPFC spatial gene-expression dataset [74] (Extended Fig 6G). GM and WM microglia tend to be phenotypically different [73], and in a recent postmortem study on freshly-isolated microglia, only GM microglia showed differential gene-expression in MDD with downregulation of immune-related genes (referred to as, disease-associated microglia in depression (depDAM)) [75]. Compared to Mic1, promoter-accessibility and gene-expression profiles of Mic2 reflected molecular status of depDAM (Extended Fig 6H-I), with significantly lower module scores for downregulated and higher for upregulated genes in MDD. While these patterns were not statistically different between groups, genes linked to downregulated Mic2 DARs in MDD significantly captured downregulated genes in depDAM (p-value=8.1e-05; one-sided Fisher's exact test).

Discussion

We examined chromatin accessibility with integrated gene-expression in ~200,000 cells from the DLPFC. Our results indicate that nearly 75% of DARs identified across cell-type clusters were differentially closed in MDD individuals. These differences were found primarily in GM microglia (Mic2) and NR4A2+ deep-layer excitatory neuronal cluster (ExN1), which was also significantly enriched for the heritability of MDD GWAS SNPs.

In line with a previous cell-type-specific study examining PFC maturation [76], our study identified MDD-associated differentially accessible chromatin in deep-layer (L5/6) ExN1 neurons enriched for activity-dependent (bZIP) motifs. These neurons were also enriched for neurodevelopment-associated (bHLH/HTH) TF binding sites. Moreover, MDD-associated TFs from these families in ExN1 showed interactions with ExN1-specific marker TF, NR4A2. NR4A2 is an immediate-early factor [62] which responds to glucocorticoids [67,77], acute and chronic stress [63-70], and has been previously associated with depression [64,66,71,72]. NR4A2 gene and protein levels were found to be decreased in bulk-tissue DLPFC from MDD individuals [78]. Moreover, NR4A2 induction can have neuroprotective effects after exposure to neuropathological stress [69] and chronic inflammation-induced depression-like behaviors [71]. The roles of NR4A2+ neurons in stress-response are consistent with observations in Nr4a2-null heterozygous mice showing significant reduction in cortical and limbic dopamine levels [79] and greater behavioral impairment associated with neuropsychiatric traits after exposure to developmental stress [80]. This suggests that NR4A2+ ExN1 neurons may be more sensitive to stress-related perturbations in depression, possibly mediated by coordinated TF activities.

Furthermore, NR4A2+ ExN1 neurons were found to be most similar to L6 IT Car3 neurons [71], which share developmental and transcriptional similarities with NR4A2-marked Claustrum neurons [60,81]. Chemogenetic silencing of these neurons attenuates stress-induced anxiety- and depression-like behaviors in rodents [61]. Moreover, activity-dependent induction of NR4A2 can modulate AMPA-receptor mediated synaptic plasticity [82] whose alterations are strongly associated with depression [83].

In our study, NR4A2+ ExN1 neurons were also most significantly enriched for the heritability of MDD GWAS SNPs, especially in the marker cCREs, suggesting cell-intrinsic programs may be involved in regulating differential accessibility alterations. MDD-associated genetic variation in

these neurons affected regulatory sites linked to genes involved in synaptic plasticity and communication, possibly impacting the sensitivity of ExN1 neurons to stress and depression.

In contrast to ExN1, a microglial cluster Mic2 in GM showed differentially closed chromatin associated with genes involved in phagocytosis- and immune-related pathways. Mic2 cluster also showed transcriptional similarities with a cortical GM microglia (depDAM), depicting down-regulation of genes involved in immune-response and phagocytosis, suggesting an overall immune-suppressed phenotype in MDD[75]. MDD-associated downregulation in immune-signaling pathways may be contributed by decreased binding activities of immune-regulatory TFs identified in this study, such as ETS domain and IRFs [34,38]. Likewise, chronic stress-induced depression-like behavior in rodents has been associated with decreased interferon-response TF motif accessibility and transcriptional-repression in microglia [84]. MDD-associated TFs identified in Mic2 were also among those depicting motif accessibility changes in response to disruption in microglia micro-environment and associated with neurodegenerative and psychiatric disorder genes [28]. This suggests possible roles of surrounding tissue-environment [28] or microglia-neuron communication [20,75] in shaping regulatory landscape of microglia.

Previous work investigating tissue inflammation in depression produced contradictory findings, suggesting both increased [85-87] and decreased immune-activation [10,75,88,89], either peripherally or centrally. Consistent with our findings, bulk-tissue [9] and single-nucleus [20] RNA-seq in the DLPFC revealed downregulation of genes associated with inflammatory cytokines, microglia–activation, and immune-signaling pathways in MDD. Some studies in rodents [84,90] and humans [91] have likewise suggested suppression of microglial immune-response in chronic stress-induced disorders. Furthermore, recent reports provide evidence at gene-expression [9,10,20,75] and protein levels [10,88,92] for a non-inflammatory microglial phenotype in MDD.

It is possible that inflammatory changes taking place over time may eventually lead to widespread changes in chromatin accessibility of microglia, dysregulating TF binding and gene-expression and altering overall immune-homeostasis in MDD individuals. Additionally, MDD has been associated with alterations in microglial homeostatic functions [10,88], such as microglia-neuron communication [20,75]. For example, early-life stress in mice induced disruption in microglia-associated pruning of excitatory synapses leading to aberrant adult stress-response [93]. Whereas both microglia-suppressing and -stimulating properties associate with context-specific

anti-depressant response [94]. These results emphasize the need to further determine the influence of microglia-associated inflammatory as well as synapse-regulatory functions in context of depression.

Unlike excitatory neurons, marker cCREs specific to glial clusters were not significantly enriched for the heritability of MDD GWAS risk loci; however, examining variant-effects of individual MDD sSNPs pointed to disruption of specific gene-regulatory sites. Together, these results delineated regulatory mechanisms by which MDD-associated genetic and environmental factors may impact distinct cortical clusters (Extended Fig 7).

Limitations

Sex-specific transcriptomic changes have been identified in depression [20,95], Here, we interrogated MDD-associated accessibility by accounting for sex of individuals but did not assess moderation effects of sex on phenotypic differences. Future studies hold promise to uncover chromatin regions possibly mediating sex differences in MDD.

The downstream interpretation of MDD-associated chromatin could be limited by the number of subjects or cells. Notably, the cohort-size of this study was comparable, if not larger, to other publicly-available postmortem single-cell brain studies [18,24,76,96]. Due to cohort-size and overrepresentation of EUR population, we opted to use reference LD panels instead of in-sample LD. Consequently, our findings may not be applicable to other populations.

Finally, we adopted gkm-SVM models to predict allelic-impact of genetic variants on accessibility. Although the performance of our models is comparable to other studies [18], these models do not directly evaluate read-count differences between individuals [97].

Conclusions

To our knowledge, our study is the first single-nucleus measurement of chromatin accessibility and integrated gene-expression in individuals affected by a psychiatric phenotype. Our study revealed MDD-associated accessibility alterations in the deep-layer excitatory neurons

characterized by TF activity of NR4A2, which is known to regulate stress-response. NR4A2+ excitatory neurons were also most significantly enriched for the heritable risk for MDD, whereby the associated genetic variation disrupted regulatory sites linked to genes involved in synaptic communication. Additionally, a grey matter microglial cluster showed significant chromatin accessibility disruption in MDD, impacting binding of TFs known to mediate immune homeostasis. This rich resource will prompt future investigations into narrowing down the functional effects of MDD-associated regulatory sites and risk-variants using more targeted approaches, in attempts to further tease-apart the interplay of genetic and environmental factors contributing to MDD.

Methods

Post-mortem brain samples for snATAC-seq

We investigated a total of eighty-four human post-mortem dorsolateral prefrontal cortex samples obtained from the Douglas Bell Canada Brain Bank (www.douglasbrainbank.ca) and University of Miami Miller School of Medicine Brain Endowment Bank (https://med.miami.edu/programs/brain-endowment-bank). Clinical information was obtained using psychological autopsies, which were performed using proxy-based, structured and standardized interviews [98]. This study was approved by the Douglas Hospital Research Ethics Board, and written informed consent from next of kin was obtained. The frozen histological grade samples of grey and white matter were dissected from the dlPFC (Brodmann Area 9) by expert neuroanatomists and stored at –80 °C. Cases met criteria for MDD and died by suicide during an episode of major depression, whereas controls were individuals who died of natural causes or in accidents and did not meet criteria for major axis I disorders. No differences were observed for the age of individuals (p-value=0.81, Case=44.6, Control=44.15). Although there was a difference between MDD and control subjects in the mean values for post-mortem interval

(PMI; p-value=0.0037, Case=44.2, Control=33.4), no significant differences were observed for the pH values, which is a better correlate of tissue integrity (p-value=0.16, Case=6.5, Control=6.4; Supplementary Figure 2).

Nuclei extraction, library multiplexing, and sequencing

We extracted nuclei using thick coronal sections (~350um on average) as they were found to produce higher nuclei-to-debris ratio compared to thinner sections (data not shown). This also allowed for obtaining nuclei across cross-sections of GM and WM. Nuclei extraction from the homogenized brain tissue was performed as previously described [99] with some modifications (Supplementary note: Nuclei Extraction). We combined nuclei extracted from a male and female pair (either control or MDD) into one well of 10x microfluidic device (Supplementary note: Multiplexing and library construction) and sequenced the combined libraries together using Illumina NovaSeq 6000. These libraries were demultiplexed using 1000 Genomes based common variants and sex-specific chromatin accessibility (Supplementary note: Demultiplexing and assignment of sex). For each library, we obtained on an average 505M raw sequencing read-pairs with on an average 89.1% sequenced read pairs with mapping quality > 30 and having on an average 141.34 bp insert-size. Quality-metrics for each subject are provided in (Supplementary Figure 1, Supplementary Table 1).

Doublet removal

10x fragment files were used as an input to ArchR [release_1.0.1]. High-quality cells were defined as those passing threshold criteria for the number of unique fragments per cell > 1000, TSSEnrichmentScore > 3.5 and fraction in mitochondrial reads < 10%. Cells detected as 10x gel-bead doublets or barcode multiplets originally identified in [100] and cells with possibly mixed genotype profiles detected using Vireo [101] were filtered out.

Unsupervised iterative clustering

Genome-wide tile-matrix was generated by binning the genomic accessibility (GRCh38) into 500bp bins. Dimensionality reduction was then performed on the tile-matrix using the estimated iterative LSI procedure in ArchR and batch-effects were corrected with Harmony [102]. Unsupervised graph-based clustering with a smart local moving algorithm was applied to addClusters(reducedDims="Harmony",dimsToUse=1:25,method=3,resolution=0.4,filterBias=TRUE). This first round of iterative clustering produced a total of 11 clusters, including 1 cluster of endothelial (C4), microglial (C5), astrocytic (C10), OPCs (C11), inhibitory neurons (C6), 3 oligodendrocyte clusters (C1-3), and 3 excitatory neuronal clusters distinguished by upper (Cluster 7), middle (C8), and deep cortical layers (C9), which were then combined for subsequent clustering within broad cell-types. For each broad cell-type, we re-computed LSI (15,000 varFeatures and 1:20 dimsToUse) and retained clusters with at least 50 cells. The second round of clustering was performed at multiple resolutions (0.2,0.4,0.6), retaining the clustering resolution that produced the highest median of mean Silhouette scores across clusters. For larger cell-types (oligodendrocytes and excitatory neuronal cell-types) for which Silhouette scores could not be computed (>70,000 cells), we used clustering resolution of 0.6.

Quality-assessment of clusters

The quality of identified clusters produced after the first and second rounds of clustering was assessed based on: a) uniform percentage contribution from subjects, batches, sexes, and conditions in each cluster, b) having at least 10 marker genes based on ArchR GeneScore matrix (Wilcoxon test, FDR < 0.05 & Log2FC > 0.5). This retained 11 out of 15 clusters after the first round and 38 out of 40 clusters after second round of clustering for downstream analysis.

Identification of chromatin cis-regulatory elements (cCREs)

To identify reproducible sets of cCREs capturing biological variability across cell-types and clusters, cells belonging to same cell-type and cluster were first partitioned into four non-overlapping pseudobulk aggregates whereby each replicate had uniform contributions from subjects belonging to each sex and condition. Open chromatin regions (OCRs) were called using MACS2 [103] on each pseudobulk replicate available for each cell-type and cluster using addReproduciblePeakSet(excludeChr="chrM", maxPeaks=300000, cutOff=0.1) in ArchR. Finally, OCRs reproducible in at least two of the four replicates in each broad cell-type and cluster were retained. Next, we generated a non-overlapping and fixed-with peak-set (501bp) separately at cell-type and cluster levels using iterative overlap peak merging and removal (IPR) procedure in ArchR. This retains the most significant OCR (based on MACS2 significance) among those overlapping across clusters. OCRs accessible in at least 1% of cells resulted in approximately 600 and 800 thousand non-overlapping cCREs at broad cell-type and cluster resolution, respectively. Majority of broad cell-type cCREs (86%) overlapped with cluster-resolved cCREs. Conversely, a much smaller percentage of cluster cCREs (64%) overlapped

with broad level cCREs. Notably, most cluster cCREs (95%) were located within 5000bp neighborhood of broad cCRE coordinates. Finally, we computed per-cell Tn5 insertion counts in each cell with respect to the above-mentioned non-overlapping cCREs from broad cell-type and cluster resolution, which were then used for downstream analyses.

Marker cCREs

Cell-type- and cluster-specific marker cCREs were identified using the respective Tn5 insertion count matrices. For this, getMarkerFeatures() in ArchR was used with Wilcoxon test (FDR < 0.05 & Log2FC > 0.5). Majority of broad cell-type-specific marker cCREs (77%) overlapped with cluster-specific marker cCREs. Conversely, a much smaller number of cluster marker cCREs (36%) overlapped with the broad marker cCREs. About 67% of cluster marker cCREs were within 5000bp of broad marker cCRE coordinates. Three (ExN7, InN3, and Oli6) of the 38 clusters were potentially low-quality as they had very few marker cCREs (<5) compared to all others.

Cluster annotation

Clusters were annotated based using several approaches: a) pseudo-bulk accessibility profiles (normalized by reads in TSS) at cell-type marker genes plotted using ArchR plotBrowserTrack function, b) assessing gene regulatory activity of cell-type marker genes in BrainInABlender [104] (Species=="Human") for every cluster using ArchR addModuleScores() function, and c)

measuring gene activity profiles with ArchR GeneScore matrix (Wilcoxon test; FDR < 0.05 & Log2FC > 0.5) and manually assessing the enrichment of following marker genes:

Neurons: *SNAP25*, *RBFOX3*; Excitatory neurons: *SATB2*, *SLC17A7*, *SLC17A6*; Excitatory neuronal cortical layers (upper to lower): *CUX2*, *RORB*, *PCP4*, *BCL11B*, *FEZF2*; Inhibitory neurons: *GAD1*, *GAD2*, *SLC32A1*; Inhibitory neuronal subtypes: *LHX6*, *PVALB*, *SST*, *ADARB2*, *VIP*, *LAMP5* Macrophage/microglia: *MRC1*, *CSF1R*, *CX3CR1*; Endothelial: *CLDN5*, *SRGN*, *FLT1*; Astrocytes: *AQP4*, *SLC1A2*, *GFAP*; OPCs: *HAS2*, *CSPG4*, *OLIG1*; OPCs: *HAS2*, *CSPG4*, and Oligodendrocytes: *MBP*, *OPALIN*, *PLP1*, *MAG*, *MOG*.

Validation of cell type clusters

Enrichment of Fluorescence-activated nuclei (FAN)-sorted cell-type marker peaks

We thoroughly validated snATAC-seq cell-type and cluster profiles using various strategies. First, we examined enrichment overlap of marker cCREs from FAN-sorted bulk ATAC-seq cell-types [105] in cluster-specific marker cCREs defined by our snATAC-seq data using ArchR peakAnnoEnrichment(cutOff ="FDR < 0.05 & Log2FC > 0.5") and plotted FANS-sorted cell-type cCREs with the highest enrichment for each cluster using plotEnrichHeatmap(enrichRegions, n = 1).

Comparison with external snATAC-seq dataset

Second, using Signac [106], cell-types identified in our data were compared to the cell-types available in PFC snATAC-seq [24]. Specifically, for each cell in our cohort, fragment counts were computed using peaks identified in the external PFC snATAC-seq dataset. Next, cell-type labels from the external snATAC-seq were transferred onto our data using cell-by-peak fragment count matrix and using FindTransferAnchors(reduction="lsiproject", dims=2:20) function in Seurat [109].

MetaNeighbor

Finally, we assessed the concordance between cell-type and clusters in this dataset with multiple snRNA-seq datasets. For this, we used the top 3000 most variable features to create gene-expression classifiers for each of the cluster defined in these snRNA-seq datasets [20 59 107]. Using MetaNeighbor [108], we predicted their similarities (AUCROC plots; Extended Figure 2C-D) to snATAC-seq clusters based on promoter-accessibility (<2kbp to gene TSS) computed using GeneActivity() function in Signac [106]

Integration with snRNA-seq dataset

Prior to integration, cells with uncertain ("Mixed") annotation in MDD snRNA-seq data [20] were removed. First, an unconstrained integration was performed using snRNA-seq gene-expression as reference and snATAC-seq GeneScore matrix as query to derive prediction scores and assess concordance between cell-types. As we observed high similarity between cell-type labels between two datasets (Extended Fig 1A-B), we individually applied constrained integration within broad cell-types across two datasets and subsequently imputed gene expression from

snRNA-seq to snATAC-seq cells using addGeneIntegrationMatrix(reducedDims="Harmony", addToArrow = TRUE, dimsToUse=1:20) function in ArchR.

Integration with spatial gene-expression dataset

snATAC-seq microglia clusters (Mic1 and Mic2) were mapped onto 10x Visium spatial gene-expression dataset comprising of male and female DLPFC tissue sections [74] using Seurat [109]. Briefly, SCTransform() was used on promoter-accessibility and integrated snRNA-seq gene-expression for snATAC-seq cells mapping to microglial clusters. Microglia cluster labels were then transferred using the top 30 principal components (PC) with FindTransferAnchors() function and predicted spatial positions were visualized using SpatialFeaturePlot() in Seurat.

Differential accessibility analysis

Differential accessibility analysis was performed in every broad cell-type and cluster using limma-voom [110] implemented in muscat [111]. First, pseudo-bulk accessibility profiles were generated by aggregating counts across cells for each cCRE for each subject within the broad cell-types and clusters. One female was omitted prior to differential analysis due to discordance between the sex identified and reported (Supplementary note). Batch, sex, age, PMI, and the total number of cells contributed by each subject were added as covariates in the limma-voom differential accessibility model. A series of quality-control measures for differential accessibility analyses were taken: a) Filtering out subjects with very low cell contribution (less than 10 cells in broad cell-types or 5 cells in clusters), b) Filtering out potential outlier subjects based on peak-

count matrix using isOutlier(type = "lower", nmads = 3) function in muscat, c) Filtering out cCREs with consistently low counts across multiple subjects using filterByExpression() function from edgeR [112]. Finally, for each broad cell-type and cluster, we retained cCREs accessible in more than 1% and 3% of cells, respectively, on an average in either cases or controls. Differentially accessible regions (DARs) with abs(logFC) > log2(1.1) and FDR-adjusted local p-value < 0.05 was used for all the analysis, except GWAS based analysis for which we defined candidate DARs (abs(logFC) > log2(1.1) and FDR < 0.2).

Differential motif analysis

DARs were first partitioned by the direction of their association with MDD and were then provided as inputs to Homer. GC-content matched background regions were automatically selected by Homer and motifs were scanned within 200bp region of the peak center using findMotifsGenome.pl function with default parameters. We only assessed the “known” motifs for downstream results. Additionally, ChromVar [39] based GC content and mean accessibility biases-corrected motif accessibility deviations z-scores were computed at single-cell level with Cisbp motif database using addDeviationsMatrix() function in ArchR. Maker TFs for each cell-type and cluster were defined as those with significantly higher motif accessibility in a cluster compared to all others using Wilcoxon test (FDR < 0.05 & MeanDiff > 1.5) with getMarkerFeatures() default function in ArchR. MDD-associated TFs were also identified using getMarkerFeatures(useGroups = "Case", bgdGroups = "Control") after adding PMI, age, and total number of fragments as bias variables.

TF motif footprinting

We generated TF binding profiles by aggregating Tn5 insertion counts across all the motif matching peaks using Cisbp motif database and subtracting Tn5 insertion sequence bias before plotting. Specifically, for cell-type and cluster-specific marker TFs, we generated a maximum of 5 pseudo-bulk aggregates (default parameter) representing each cluster. For MDD-associated TFs, we created 10 pseudo-bulk replicate profiles from each group using addGroupCoverages(groupBy="Condition", maxReplicate=10) in ArchR.

Gene regulatory networks

We prioritized MDD-associated TFs for constructing gene regulatory networks (Supplementary Figure 5E). Briefly, among the topmost significantly enriched TF motifs (FDR<0.05) in cluster-specific DARs, we selected the top two TFs with the highest median gene-expression using snRNA-seq integrated expression. Between these two TFs, we prioritized the one with the most significant difference (FDR<0.05) between case and control cells using ChromVAR. For ExN1 gene-regulatory network, following DARs (FDR<0.05, abs(logFC) >log2(1.1)) were used, but for Mic2 network, we selected topmost significant DARs (FDR<0.01, abs(logFC) >log2(1.1)). To improve readability of graphs, we only plotted DARs mapping to ArchR-defined non-coding elements, including promoters, introns and distal regions and excluded exons. The edges in these networks are directed from DAR nodes to gene nodes and are drawn whenever the corresponding DAR contained a TF binding site and: (a) connects any DAR (promoter, introns, or distal) to corresponding target-genes through peak-to-gene linkages (r>0.45); and (b) connects promoter

DARs to their nearest gene. Directed gene-regulatory networks were visualized using the Fruchterman-Reingold layout algorithm with the R package igraph (v 1.2.6). The nodes in green represent DARs enriched for individual TF binding sites directed towards putative target-genes colored according to MDD associated gene or MDD DEG status. For MDD-associated genes, we compiled list of potential risk-genes in MDD GWAS [11-13,72,113,114], Hi-C based analysis (H-MAGMA; FDR<0.1) [115] and PsyGeNET depressive disorders genes [26]. For MDD DEGs, we used bulk (FDR<0.1) [116] and snRNA-seq DEGs [20] obtained from the meta-analysis of both sexes (FDR<0.1, abs(log2FC) > log2(1.1), signs > 0).

Functional interpretation of differential results

- **Co-accessible peaks and peak-to-gene linkages**

Co-accessibility of all pairs of cCREs at cell-type and cluster levels were computed using via addCoAccessibility(maxDist=5e+05) function in ArchR. Moreover, using MDD snRNA-seq integrated expression, for all pairs of cCREs and genes, we computed the correlations between cCRE-accessibility and gene-expression, using addPeak2GeneLinks(maxDist = 5e+05) in ArchR. We pruned the resulting peak-to-peak co-accessibility and peak-to-gene linkage matrices to retain all significantly correlated pairs (r>0.45), as defined in ArchR. Additionally, we retained genes that were most significantly correlated (lowest FDR values) with cCREs associated with MDD sSNPs using peak-to-gene linkages (Supplementary Table 12).

- **Gene ontology (GO) pathways**

Gene ontology pathways were identified for DAR-linked genes (r>0.45) in each cluster using enrichGO(minGSSize=5, pAdjustMethod="BH", pvalueCutoff=0.05, ont="ALL") with clusterProfiler [117]. We also split genes according to differentially more or less accessible DARs and performed pathway enrichment analysis using more relaxed parameters: enrichGO(pAdjustMethod="BH", pvalueCutoff=0.2).

- **PsyGeNET**

For examining the association of DAR-linked genes with psychiatric phenotypes, we used psygenet2r package with database="ALL" and default parameters associated with PsyGeNET

database [26]. The total number of gene-disease associations and evidence index for each psychiatric disorder was computed using geneAttrPlot() with default parameters.

- **StringDB**

Protein-protein (PP) interactions were extracted from “full String network” of organism= ”Homo Sapiens.” The minimum required interaction score was set to “high” (i.e., 0.7) and the maximum number of interactions allowed in first and second shells of the network were limited to 20 and 5 interactions, respectively. The fully-connected PP interaction network was generated by removing the disconnected nodes.

- **cCRE overlap analysis**

To compute overlap of PsychEncode H3K27ac and H3K4me3 histone modification peaks consolidated from cortical cells [25] with marker cCREs, cCREs in peak-to-gene linkages (r>0.45), and MDD DARs, we used subsetByOverlaps() in IRanges R package with default parameters.

- **Gene overlap analysis**

To assess enrichment of snRNA-seq DEGs, GWAS risk-genes, PsyGeNET depressive disorder genes in DAR-linked genes at broad and cluster levels, we used one-sided Fisher’s exact test in newGOM() GeneOverlap R package [v.1.36.0; http://shenlab-sinai.github.io/shenlab-sinai/] and plotted FDR-corrected p-values using drawHeatmap(go.obj, adj.p = T,log.scale = T, ncolused = 9).

- **Module scores**

Gene-expression module scores (average weighted expression of genes at single-cell level) were calculated using Seurat AddModuleScore() function [109] for DAR-linked genes and TF target-genes. These scores were compared between MDD versus control cells (Wilcoxon test) using the most similar snRNA-seq defined clusters [20] to ExN1 and Mic2 snATAC-seq clusters. Both ExN16_L56 and ExN20_L56 snRNA-seq clusters showed correspondence to ExN1 (Extended Figure 2C); however, similarly to ExN1, ExN16_L56 also showed the highest similarity [20] with ExN L6 IT Car3 neurons in the ABI snRNA-seq [59], and had the highest mean expression for *NR4A2* (Extended Figure 2E), a cell-defining TF marker for ExN1. On the other hand, snRNA-seq data defined only one microglial cluster, which was used for the assessment of Mic2 gene-expression module scores. In addition, we examined module scores using promoter-accessibility and snRNA-seq integrated gene-expression for molecular characterization of microglia. For GM versus WM comparison, we used top 25 differential genes between GM and WM microglia (sorted by fold change) [73]. For comparison with MDD-associated microglia (depDAM), we used all the genes that were upregulated or downregulated in MDD [75].

Heritability enrichment

To calculate heritability enrichment of GWAS SNPs within cluster-specific marker cCREs and MDD-associated chromatin, we applied stratified LDSC regression model (v 1.0.1) to GWAS summary statistics for MDD [11 13] and other complex traits [118 119 120 121 122 123 124]. Cluster-specific marker cCREs were retained at FDR<0.05 & Log2FC>0.5. For MDD-associated chromatin, we used candidate DARs (abs(logFC) > log2(1.1) & FDR < 0.2) to obtain topmost dysregulated

cCREs representing most of the clusters. These cCREs were converted from hg38 to hg19 coordinates using the UCSC liftover tool (v377). Cell-type specific partitioned heritability analysis was computed using a full baseline-LD model and after adding LD scores estimated on matching cCRE set combined across all clusters as background. Finally, one-sided p-values were computed based on the enrichment z-scores for each annotation relative to the background for every trait.

Heritability Estimator from Summary Statistics (HESS)

We used HESS [43] to estimate local heritability for MDD-associated variants [11] in LD blocks defined on 1000 Genomes [125]. This identified LD blocks significantly enriched ($p<0.05$) for MDD heritability. We then tested the enrichment of cluster-specific candidate DARs within MDD-associated heritable LD blocks identified above. For this, we compared cluster-specific overlap of candidate DARs in significantly heritable LD blocks for MDD (n=90) with per-cluster null distribution of the overlap of chromosome-matched cCREs within any LD block permutated 10,000 times to derive an empirical p-value of significance ($p<0.05$). Finally, both ExN1-specific marker cCREs and candidate DARs whose peak-coordinates were located specifically within significant MDD heritable LD blocks were input to Homer motif analysis (as described above). TF families enriched ($q<0.05$) commonly in both ExN1 marker cCREs and candidate DARs were identified. The occurrence frequency of enriched TF families was plotted using worlcloud2(v. 0.2.1) based on the number of individual TF members (belonging to each family) that were significantly enriched in ExN1-specific cCREs.

Enrichment of DLPFC expression and splicing QTLs in cCREs

To perform e/sQTL enrichment in cCREs, we used binomial test to assess whether uniformly processed DLPFC eQTL and sQTLs (FDR<0.05; Supplementary method) are significantly enriched within cCREs. First, eQTL and sQTL SNPs located within 501-bp cCREs were identified. Next, the expected number of eQTL and sQTLs in cCREs were computed as the product of (1) cCRE coverage with respect to hg38 genome; and (2) total number of QTL SNPs at FDR < 0.05 (~3,000,000 eQTL and ~2,450,000 sQTL SNPs). Furthermore, binomial enrichment test p-values are calculated for each QTL type and cCREs. Here, success rate, number of trials, and number of successes for the binomial test were set to the cCRE coverage percentage, total number QTL SNPs, and the number of QTL SNPs observed in cCREs, respectively.

Functional impact of MDD-associated genetic variation in each cluster

To identify allelic-effects of MDD-associated GWAS SNPs and those associated by LD (r>0.8) and fine-mapping (Supplementary methods) on chromatin accessibility of each cluster, we adapted three complementary gapped kmer support vector machine (gkmSVM) models [46 47 48]. Briefly, for each of the snATAC cluster, gkmSVM models were trained on top 60,000 OCRs (whenever available) identified in each cluster by MACS2 (ranked by FDR) to predict whether 1001bp reference genome sequences were accessible (i.e., cCRE) or inaccessible (i.e., GC-content and chromosome-matched sequences) for the corresponding cluster. We specifically focused on MDD SNPs located in either marker cCREs or candidate DARs of each cluster (e.g., ExN1) and its respective cell-type (e.g., ExN) to assess the effects of MDD sSNPs at cluster-

refined resolution. Using cluster-specific trained models, we then identified significant MDD SNPs (sSNPs) altering chromatin accessibility of 201bp neighborhood of a cCRE. Moreover, we partitioned sSNPs identified per cluster to high, medium and low confidence sets reflecting significance of accessibility disruption in its immediate local neighborhood (<201bp). Putatively disrupted TF binding sites were determined by matching affected allele sequence with Cisbp motif database in TOMTOM [126].

Multi-modal Visualization

For interpreting functional targets of MDD sSNPs, we used snRNA-seq integrated peak-to-gene linkages (r>0.45) and snATAC-seq based peak-to-peak co-accessibility (r>0.45) using all the co-accessible loops within 10kbp region of cCREs associated with MDD sSNPs. We further overlayed PsychEncode DLPFC histone modification peaks (H3K27ac, H3K4me3) merged across subjects [25] and DLPFC NeuN+/- promoter-interacting Hi-C linkages (10kbp resolution) [49] (Supplementary note: PsychEncode data).

References

1. Cai, H. *et al.* Prevalence of Suicidality in Major Depressive Disorder: A Systematic Review and Meta-Analysis of Comparative Studies. *Front. Psychiatry* **12**, 690130 (2021).

2. James, S. L. *et al.* Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet* **392**, 1789–1858 (2018).
3. Clevenger, S. S., Malhotra, D., Dang, J., Vanle, B. & IsHak, W. W. The role of selective serotonin reuptake inhibitors in preventing relapse of major depressive disorder. *Ther. Adv. Psychopharmacol.* **8**, 49–58 (2018).
4. Abdallah, C. G., Sanacora, G., Duman, R. S. & Krystal, J. H. The neurobiology of depression, ketamine and rapid-acting antidepressants: Is it glutamate inhibition or activation? *Pharmacol. Ther.* **190**, 148–158 (2018).
5. Sequeira, A. *et al.* Global Brain Gene Expression Analysis Links Glutamatergic and GABAergic Alterations to Suicide and Major Depression. *PLoS ONE* **4**, e6585 (2009).
6. John, C. S. *et al.* Blockade of Astrocytic Glutamate Uptake in the Prefrontal Cortex Induces Anhedonia. *Neuropsychopharmacology* **37**, 2467–2475 (2012).
7. Nagy, C. *et al.* Astrocytic abnormalities and global DNA methylation patterns in depression and suicide. *Mol. Psychiatry* **20**, 320–328 (2015).
8. Edgar, N. & Sibille, E. A putative functional role for oligodendrocytes in mood regulation. *Transl. Psychiatry* **2**, e109–e109 (2012).
9. Pantazatos, S. P. *et al.* Whole-transcriptome brain expression and exon-usage profiling in major depression and suicide: evidence for altered glial, endothelial and ATPase activity. *Mol. Psychiatry* **22**, 760–773 (2017).
10. Snijders, G. J. L. J. *et al.* Distinct non-inflammatory signature of microglia in post-mortem brain tissue of patients with major depressive disorder. *Mol. Psychiatry* **26**, 3336–3349 (2021).
11. Howard, D. M. *et al.* Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat. Neurosci.* **22**, 343–352 (2019).

12. Levey, D. F. *et al.* Bi-ancestral depression GWAS in the Million Veteran Program and meta-analysis in >1.2 million individuals highlight new therapeutic directions. *Nat. Neurosci.* **24**, 954–963 (2021).
13. Als, T. D. *et al. Identification of 64 new risk loci for major depression, refinement of the genetic architecture and risk prediction of recurrence and comorbidities.* http://medrxiv.org/lookup/doi/10.1101/2022.08.24.22279149 (2022) doi:10.1101/2022.08.24.22279149.
14. Chen, Y. *et al.* The interaction of early life factors and depression-associated loci affecting the age at onset of the depression. *Transl. Psychiatry* **12**, 294 (2022).
15. Peterson, R. E. *et al.* Molecular Genetic Analysis Subdivided by Adversity Exposure Suggests Etiologic Heterogeneity in Major Depression. *Am. J. Psychiatry* **175**, 545–554 (2018).
16. Ochi, S. & Dwivedi, Y. Dissecting early life stress-induced adolescent depression through epigenomic approach. *Mol. Psychiatry* **28**, 141–153 (2023).
17. Maurano, M. T. *et al.* Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. *Science* **337**, 1190–1195 (2012).
18. Corces, M. R. *et al.* Single-cell epigenomic analyses implicate candidate causal variants at inherited risk loci for Alzheimer's and Parkinson's diseases. *Nat. Genet.* **52**, 1158–1168 (2020).
19. Nagy, C. *et al.* Single-nucleus transcriptomics of the prefrontal cortex in major depressive disorder implicates oligodendrocyte precursor cells and excitatory neurons. *Nat. Neurosci.* **23**, 771–781 (2020).
20. Maitra, M. *et al.* Cell type specific transcriptomic differences in depression show similar patterns between males and females but implicate distinct cell types and genes. *Nat. Commun.* **14**, 2912 (2023).
21. Buenrostro, J. D. *et al.* Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 (2015).

22. Chawla, A., Nagy, C. & Turecki, G. Chromatin Profiling Techniques: Exploring the Chromatin Environment and Its Contributions to Complex Traits. *Int. J. Mol. Sci.* **22**, 7612 (2021).

23. Granja, J. M. *et al.* ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat. Genet.* **53**, 403–411 (2021).

24. Morabito, S. *et al.* Single-nucleus chromatin accessibility and transcriptomic characterization of Alzheimer's disease. *Nat. Genet.* **53**, 1143–1155 (2021).

25. Girdhar, K. *et al.* Cell-specific histone modification maps in the human frontal lobe link schizophrenia risk to the neuronal epigenome. *Nat. Neurosci.* **21**, 1126–1136 (2018).

26. Gutiérrez-Sacristán, A. *et al.* PsyGeNET: a knowledge platform on psychiatric disorders and their genes. *Bioinformatics* **31**, 3075–3077 (2015).

27. Boulting, G. L. *et al.* Activity-dependent regulome of human GABAergic neurons reveals new patterns of gene regulation and neurological disease heritability. *Nat. Neurosci.* **24**, 437–448 (2021).

28. Gosselin, D. *et al.* An environment-dependent transcriptional network specifies human microglia identity. *Science* **356**, eaal3222 (2017).

29. Klemm, S. L., Shipony, Z. & Greenleaf, W. J. Chromatin accessibility and the regulatory epigenome. *Nat. Rev. Genet.* **20**, 207–220 (2019).

30. Heinz, S. *et al.* Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol. Cell* **38**, 576–589 (2010).

31. West, A. E. & Greenberg, M. E. Neuronal Activity-Regulated Gene Transcription in Synapse Development and Cognitive Function. *Cold Spring Harb. Perspect. Biol.* **3**, a005744–a005744 (2011).

32. Dennis, D. J., Han, S. & Schuurmans, C. bHLH transcription factors in neural development, disease, and reprogramming. *Brain Res.* **1705**, 48–65 (2019).

33. Harris, H. K. *et al.* Disruption of RFX family transcription factors causes autism, attention-deficit/hyperactivity disorder, intellectual disability, and dysregulated behavior. *Genet. Med.* **23**, 1028–1040 (2021).

34. Zhou, N., Liu, K., Sun, Y., Cao, Y. & Yang, J. Transcriptional mechanism of IRF8 and PU.1 governs microglial activation in neurodegenerative condition. *Protein Cell* **10**, 87–103 (2019).

35. Holtman, I. R., Skola, D. & Glass, C. K. Transcriptional control of microglia phenotypes in health and disease. *J. Clin. Invest.* **127**, 3220–3229 (2017).

36. Wehrspaun, C. C., Haerty, W. & Ponting, C. P. Microglia recapitulate a hematopoietic master regulator network in the aging human frontal cortex. *Neurobiol. Aging* **36**, 2443.e9-2443.e20 (2015).

37. Belhocine, S. *et al.* Context-dependent transcriptional regulation of microglial proliferation. *Glia* **70**, 572–589 (2022).

38. Kierdorf, K. & Prinz, M. Factors regulating microglia activation. *Front. Cell. Neurosci.* **7**, (2013).

39. Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods* **14**, 975–978 (2017).

40. Schizophrenia Working Group of the Psychiatric Genomics Consortium *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).

41. ReproGen Consortium *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).

42. Nguyen, T.-D. *et al.* Genetic heterogeneity and subtypes of major depression. *Mol. Psychiatry* **27**, 1667–1675 (2022).

43. Shi, H., Kichaev, G. & Pasaniuc, B. Contrasting the Genetic Architecture of 30 Complex Traits from Summary Association Data. *Am. J. Hum. Genet.* **99**, 139–153 (2016).

44. Zhang, W., Najafabadi, H. & Li, Y. *SparsePro: an efficient fine-mapping method integrating summary statistics and functional annotations*. http://biorxiv.org/lookup/doi/10.1101/2021.10.04.463133 (2021) doi:10.1101/2021.10.04.463133.

45. Lee, D. *et al.* A method to predict the impact of regulatory variants from DNA sequence. *Nat. Genet.* **47**, 955–961 (2015).

46. Shrikumar, A., Prakash, E. & Kundaje, A. GkmExplain: fast and accurate interpretation of nonlinear gapped *k* -mer SVMs. *Bioinformatics* **35**, i173–i182 (2019).

47. Ghandi, M. *et al.* gkmSVM: an R package for gapped-kmer SVM. *Bioinformatics* **32**, 2205–2207 (2016).

48. Lee, D. LS-GKM: a new gkm-SVM for large-scale datasets. *Bioinformatics* **32**, 2196–2198 (2016).

49. Hu, B. *et al.* Neuronal and glial 3D chromatin architecture informs the cellular etiology of brain disorders. *Nat. Commun.* **12**, 3968 (2021).

50. Yao, X. *et al.* Integrative analysis of genome-wide association studies identifies novel loci associated with neuropsychiatric disorders. *Transl. Psychiatry* **11**, 69 (2021).

51. Zhang, D., Zeldin, D. C. & Blackshear, P. J. Regulatory factor X4 variant 3: A transcription factor involved in brain development and disease. *J. Neurosci. Res.* **85**, 3515–3522 (2007).

52. Matt, L., Kim, K., Chowdhury, D. & Hell, J. W. Role of Palmitoylation of Postsynaptic Proteins in Promoting Synaptic Plasticity. *Front. Mol. Neurosci.* **12**, 8 (2019).

53. Callejas-Marin, A. *et al.* Gli2-Mediated Shh Signaling Is Required for Thalamocortical Projection Guidance. *Front. Neuroanat.* **16**, 830758 (2022).

54. Pappas, A. L. *et al.* Deficiency of Shank2 causes mania-like behavior that responds to mood stabilizers. *JCI Insight* **2**, e92052 (2017).

55. Ciuculete, D. M. *et al.* meQTL and ncRNA functional analyses of 102 GWAS-SNPs associated with depression implicate HACE1 and SHANK2 genes. *Clin. Epigenetics* **12**, 99 (2020).

56. Hong, C. *et al.* Constitutive activation of LXR in macrophages regulates metabolic and inflammatory gene expression: identification of ARL7 as a direct target. *J. Lipid Res.* **52**, 531–539 (2011).

57. Suzuki, E. *et al.* The transcription factor Fli-1 regulates monocyte, macrophage and dendritic cell development in mice. *Immunology* **139**, 318–327 (2013).

58. Mukherjee, S. P. *et al.* Analysis of the RelA:CBP/p300 Interaction Reveals Its Involvement in NF-κB-Driven Transcription. *PLoS Biol.* **11**, e1001647 (2013).

59. BRAIN Initiative Cell Census Network (BICCN) *et al.* A multimodal cell census and atlas of the mammalian primary motor cortex. *Nature* **598**, 86–102 (2021).

60. Peng, H. *et al.* Morphological diversity of single neurons in molecularly defined cell types. *Nature* **598**, 174–181 (2021).

61. Niu, M. *et al.* Claustrum mediates bidirectional and reversible control of stress-induced anxiety responses. *Sci. Adv.* **8**, eabi6375 (2022).

62. Safe, S. *et al.* Nuclear receptor 4A (NR4A) family – orphans no more. *J. Steroid Biochem. Mol. Biol.* **157**, 48–60 (2016).

63. Imura, T., Kobayashi, Y., Suzutani, K., Ichikawa-Tomikawa, N. & Chiba, H. Differential expression of a stress-regulated gene Nr4a2 characterizes early- and late-born hippocampal granule cells. *Hippocampus* **29**, 539–549 (2019).

64. Jeanneteau, F. *et al.* The Stress-Induced Transcription Factor NR4A1 Adjusts Mitochondrial Function and Synapse Number in Prefrontal Cortex. *J. Neurosci.* **38**, 1335–1350 (2018).

65. Campos-Melo, D., Galleguillos, D., Sánchez, N., Gysling, K. & Andrés, M. E. Nur transcription factors in stress and addiction. *Front. Mol. Neurosci.* **6**, (2013).

66. Rojas, P., Joodmardi, E., Perlmann, T. & Ögren, S. O. Rapid increase of Nurr1 mRNA expression in limbic and cortical brain structures related to coping with depression-like behavior in mice. *J. Neurosci. Res.* **88**, 2284–2293 (2010).

67. Helbling, J.-C., Minni, A. M., Pallet, V. & Moisan, M.-P. Stress and glucocorticoid regulation of NR4A genes in mice: NR4A Receptors, Stress, and Glucocorticoids. *J. Neurosci. Res.* **92**, 825–834 (2014).

68. Bordoni, L., Petracci, I., Calleja-Agius, J., Lalor, J. G. & Gabbianelli, R. NURR1 Alterations in Perinatal Stress: A First Step towards Late-Onset Diseases? A Narrative Review. *Biomedicines* **8**, 584 (2020).

69. Volakakis, N. *et al.* NR4A orphan nuclear receptors as mediators of CREB-dependent neuroprotection. *Proc. Natl. Acad. Sci.* **107**, 12317–12322 (2010).

70. Pallarés, M. E. & Antonelli, M. C. Hormonal Modulation of Catecholaminergic Neurotransmission in a Prenatal Stress Model. in *Perinatal Programming of Neurodevelopment* (ed. Antonelli, M. C.) vol. 10 45–59 (Springer New York, 2015).

71. He, Y. *et al.* Protective effect of Nr4a2 (Nurr1) against LPS-induced depressive-like behaviors via regulating activity of microglia and CamkII neurons in anterior cingulate cortex. *Pharmacol. Res.* **191**, 106717 (2023).

72. eQTLGen *et al.* Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50**, 668–681 (2018).

73. Van Der Poel, M. *et al.* Transcriptional profiling of human microglia reveals grey–white matter heterogeneity and multiple sclerosis-associated changes. *Nat. Commun.* **10**, 1139 (2019).

74. Maynard, K. R. *et al.* Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nat. Neurosci.* **24**, 425–436 (2021).

75. Scheepstra, K. W. F. *et al.* Microglia Transcriptional Profiling in Major Depressive Disorder Shows Inhibition of Cortical Gray Matter Microglia. *Biol. Psychiatry* S0006322323012398 (2023) doi:10.1016/j.biopsych.2023.04.020.

76. Herring, C. A. *et al.* Human prefrontal cortex gene regulatory dynamics from gestation to adulthood at single-cell resolution. *Cell* **185**, 4428-4447.e28 (2022).

77. Carpentier, R., Sacchetti, P., Ségard, P., Staels, B. & Lefebvre, P. The glucocorticoid receptor is a co-regulator of the orphan nuclear receptor Nurr1. *J. Neurochem.* **0**, 071106220615011-??? (2007).

78. Xing, G., Zhang, L., Russell, S. & Post, R. Reduction of dopamine-related transcription factors Nurr1 and NGFI-B in the prefrontal cortex in schizophrenia and bipolar disorders. *Schizophr. Res.* **84**, 36–56 (2006).

79. Eells, J. B., Lipska, B. K., Yeung, S. K., Misler, J. A. & Nikodem, V. M. Nurr1-null heterozygous mice have reduced mesolimbic and mesocortical dopamine levels and increased stress-induced locomotor activity. *Behav. Brain Res.* **136**, 267–275 (2002).

80. Eells, J. B., Misler, J. A. & Nikodem, V. M. Early postnatal isolation reduces dopamine levels, elevates dopamine turnover and specifically disrupts prepulse inhibition in Nurr1-null heterozygous mice. *Neuroscience* **140**, 1117–1126 (2006).

81. Fang, C., Wang, H. & Naumann, R. K. Developmental Patterning and Neurogenetic Gradients of Nurr1 Positive Neurons in the Rat Claustrum and Lateral Cortex. *Front. Neuroanat.* **15**, 786329 (2021).

82. Català-Solsona, J. *et al.* Activity-Dependent Nr4a2 Induction Modulates Synaptic Expression of AMPA Receptors and Plasticity via a $Ca^{2+/}$ CRTC1/CREB Pathway. *J. Neurosci.* **43**, 3028–3041 (2023).

83. Bleakman, D., Alt, A. & Witkin, J. AMPA Receptors in the Therapeutic Management of Depression. *CNS Neurol. Disord. - Drug Targets* **6**, 117–126 (2007).

84. Zhang, Y. *et al.* Microglia-specific transcriptional repression of interferon-regulated genes after prolonged stress in mice. *Neurobiol. Stress* **21**, 100495 (2022).

85. Wang, Q., Roy, B., Turecki, G., Shelton, R. C. & Dwivedi, Y. Role of Complex Epigenetic Switching in Tumor Necrosis Factor-α Upregulation in the Prefrontal Cortex of Suicide Subjects. *Am. J. Psychiatry* **175**, 262–274 (2018).

86. Henry, C. J. *et al.* Minocycline attenuates lipopolysaccharide (LPS)-induced neuroinflammation, sickness behavior, and anhedonia. *J. Neuroinflammation* **5**, 15 (2008).

87. O'Connor, J. C. *et al.* Lipopolysaccharide-induced depressive-like behavior is mediated by indoleamine 2,3-dioxygenase activation in mice. *Mol. Psychiatry* **14**, 511–522 (2009).

88. Böttcher, C. *et al.* Single-cell mass cytometry of microglia in major depressive disorder reveals a non-inflammatory phenotype with increased homeostatic marker expression. *Transl. Psychiatry* **10**, 310 (2020).

89. Morrison, F. G. *et al.* Reduced interleukin 1A gene expression in the dorsolateral prefrontal cortex of individuals with PTSD and depression. *Neurosci. Lett.* **692**, 204–209 (2019).

90. Kreisel, T. *et al.* Dynamic microglial alterations underlie stress-induced depressive-like behavior and suppressed neurogenesis. *Mol. Psychiatry* **19**, 699–709 (2014).

91. Zhang, J., Kaye, A. P., Wang, J. & Girgenti, M. J. Transcriptomics of the depressed and PTSD brain. *Neurobiol. Stress* **15**, 100408 (2021).

92. Hannestad, J. *et al.* The neuroinflammation marker translocator protein is not elevated in individuals with mild-to-moderate depression: A [11C]PBR28 PET study. *Brain. Behav. Immun.* **33**, 131–138 (2013).

93. Bolton, J. L. *et al.* Early stress-induced impaired microglial pruning of excitatory synapses on immature CRH-expressing neurons provokes aberrant adult stress responses. *Cell Rep.* **38**, 110600 (2022).

94. Yirmiya, R., Rimmerman, N. & Reshef, R. Depression as a Microglial Disease. *Trends Neurosci.* **38**, 637–658 (2015).

95. Labonté, B. *et al.* Sex-specific transcriptional signatures in human depression. *Nat. Med.* **23**, 1102–1111 (2017).

96. Lee, A. J. *et al.* Characterization of altered molecular mechanisms in Parkinson's disease through cell type–resolved multiomics analyses. *Sci. Adv.* **9**, eabo2467 (2023).

97. Kumasaka, N., Knights, A. J. & Gaffney, D. J. Fine-mapping cellular QTLs with RASQUAL and ATAC-seq. *Nat. Genet.* **48**, 206–213 (2016).

98. Dumais, A. *et al.* Risk Factors for Suicide Completion in Major Depression: A Case-Control Study of Impulsive and Aggressive Behaviors in Men. *Am. J. Psychiatry* **162**, 2116–2124 (2005).

99. Maitra, M. *et al.* Extraction of nuclei from archived postmortem tissues for single-nucleus sequencing applications. *Nat. Protoc.* **16**, 2788–2801 (2021).

100. Lareau, C. A., Ma, S., Duarte, F. M. & Buenrostro, J. D. Inference and effects of barcode multiplets in droplet-based single-cell assays. *Nat. Commun.* **11**, 866 (2020).

101. Huang, Y., McCarthy, D. J. & Stegle, O. Vireo: Bayesian demultiplexing of pooled single-cell RNA-seq data without genotype reference. *Genome Biol.* **20**, 273 (2019).

102. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).

103. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).

104. Hagenauer, M. H. *et al.* Inference of cell type content from human brain transcriptomic datasets illuminates the effects of age, manner of death, dissection, and psychiatric diagnosis. *PLOS ONE* **13**, e0200003 (2018).

105. Hauberg, M. E. *et al.* Common schizophrenia risk variants are enriched in open chromatin regions of human glutamatergic neurons. *Nat. Commun.* **11**, 5581 (2020).

106. Stuart, T., Srivastava, A., Madad, S., Lareau, C. A. & Satija, R. Single-cell chromatin state analysis with Signac. *Nat. Methods* **18**, 1333–1341 (2021).

107. Jäkel, S. *et al.* Altered human oligodendrocyte heterogeneity in multiple sclerosis. *Nature* **566**, 543–547 (2019).

108. Crow, M., Paul, A., Ballouz, S., Huang, Z. J. & Gillis, J. Characterizing the replicability of cell types defined by single cell RNA-sequencing data using MetaNeighbor. *Nat. Commun.* **9**, 884 (2018).

109. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).

110. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).

111. Crowell, H. L. *et al.* muscat detects subpopulation-specific state transitions from multi-sample multi-condition single-cell transcriptomics data. *Nat. Commun.* **11**, 6077 (2020).

112. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).

113. Silveira, P. P., Pokhvisneva, I., Howard, D. M. & Meaney, M. J. A sex-specific genome-wide association study of depression phenotypes in UK Biobank. *Mol. Psychiatry* (2023) doi:10.1038/s41380-023-01960-0.

114. Howard, D. M. *et al.* Genome-wide association study of depression phenotypes in UK Biobank identifies variants in excitatory synaptic pathways. *Nat. Commun.* **9**, 1470 (2018).

115. Sey, N. Y. A. *et al.* A computational tool (H-MAGMA) for improved prediction of brain-disorder risk genes by incorporating brain chromatin interaction profiles. *Nat. Neurosci.* **23**, 583–593 (2020).

116. Gandal, M. J. *et al.* Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science* **359**, 693–697 (2018).

117. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS J. Integr. Biol.* **16**, 284–287 (2012).

118. 23andMe Research Team *et al.* Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. *Nat. Genet.* **50**, 920–927 (2018).

119. Yengo, L. *et al.* Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Hum. Mol. Genet.* **27**, 3641–3649 (2018).

120. Trubetskoy, V. *et al.* Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* **604**, 502–508 (2022).

121. eQTLGen Consortium *et al.* Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat. Genet.* **51**, 793–803 (2019).

122. Demontis, D. *et al.* Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat. Genet.* **51**, 63–75 (2019).

123. Watanabe, K. *et al.* Genome-wide meta-analysis of insomnia prioritizes genes associated with metabolic and psychiatric pathways. *Nat. Genet.* **54**, 1125–1132 (2022).

124. Mullins, N. *et al.* Dissecting the Shared Genetic Architecture of Suicide Attempt, Psychiatric Disorders, and Known Risk Factors. *Biol. Psychiatry* **91**, 313–327 (2022).

125. The 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

126. Gupta, S., Stamatoyannopoulos, J. A., Bailey, T. L. & Noble, W. Quantifying similarity between motifs. *Genome Biol.* **8**, R24 (2007).

Figure Legends

Figure 1. Chromatin Architecture at the single-cell level. A. Schematic overview of the snATAC-seq study design in 84 subjects from nuclei extraction and multiplexing to downstream data analyses B. UMAP of 201,456 Chromatin accessibility defined high-quality cells colored by 38 clusters from 7 broad cell-types. C. Iterative clustering within broad excitatory and inhibitory cell-types revealed 12 excitatory (ExN1-12; top), 5 inhibitory clusters (InN3/PV/SST/LAMP5/VIP; bottom). D. UMAP plot colored by gene-activity computed using ArchR GeneScore matrix for cortical layer (top) and interneuron lineage marker genes (bottom). Top: CUX2: Upper layer (L1-2), RORB, TOX: Middle layer (L3-4), SYNPR, NTNG2, FEZF2: Deep layer (L5-6). Bottom: LHX2, SST, PVALB: medial ganglionic eminence (MGE), ADARB2, VIP, LAMP5: caudal (CGE) lineage interneurons. E. Pseudo-bulk chromatin accessibility at cell-type marker genes (normalized by reads in TSS; Ast: GFAP; ExN1: SLC17A7; InN: GAD2; Mic: CX3CR1; Oli: MAG; OPC: CSPG4). F. Total no. of nuclei in each cluster followed by proportion of nuclei contributed by MDD and control subjects and contributed by male and female subjects. G. Heatmap showing differentially accessible cCREs identified in each cluster compared to all others (marker cCREs; Wilcoxon test, FDR < 0.05, Log2FC > 0.5). H. Heatmap showing mean differences in ChromVar based bias-corrected motif accessibility deviation z-scores in each cluster compared to all others (Top 3 marker TFs are plotted; Wilcoxon test, FDR < 0.05, MeanDiff>1.5; Grey boxes represent NA values). I. Tn5 bias subtracted motif footprints for cell-type marker TFs (JUN: Neuronal marker, SPI1: Microglial marker).

Figure 2. Chromatin profiling in individuals with major depression. A. Differentially accessible regions (DARs) in MDD compared to neurotypical control by cluster (Limma-voom; FDR < 0.05 & abs(LogFC) > log2(1.1)). Points above the line represent increased accessibility, below decreased. Colors represent distance to the nearest transcription start sites (TSS). B. Total number of differentially accessible regions (DARs) that are differentially more (open; n=1142) or less accessible (closed; n=3609) in MDD versus controls subjects across all clusters. Adjacent bar plots show the total number of PsychEncode ChIP-seq histone modification peaks from cortical cells (H3K27ac; marks active enhancers and H3K27ac+H3K4me3; marks promoters) that overlap differentially more (bottom) or less (top) accessible MDD DARs. C. PsyGeNet

identified the total number of gene-disease associations linked to DARs across all clusters (DAR-linked genes (r>0.45); SI: substance induced; 100% association: consistent evidence for the disease; 100% no association: consistently not associated with the disease). D. Gene overlap enrichment (one-sided Fisher's exact test; FDR<0.05) between DAR-linked genes and MDD-associated GWAS risk-genes from Howard et. al (2019) and Als et.al (2023), PsyGeNET genes associated with depressive disorder, and MDD DEGs identified in snRNA-seq meta-analysis across both sexes. E. DAR-linked genes plotted according to log2FC and -log10(adjusted p-values) of each DAR in ExN1 (left) and F. Mic2 (right). Inset: Bar plots depict significant enrichment (p<0.05) of expression (eQTLs) and splicing quantitative trait loci (sQTLs) in ExN1 (left) and Mic2 DARs (right) compared to expected genomic coverage. G-H. Gene ontology (GO) terms enriched in G. ExN1 and H. Mic2 DAR-linked genes converted into networks (emapplot) plotted by significance (FDR<0.05). Nodes in the network show significant GO terms (size of node is scaled by the number of genes associated with each GO term). Edges connect overlapping gene sets (minimum 10% overlap required for an edge criterion between GO terms and thickness of edges are scaled by the overlap percentage).

Figure 3: The role of transcription factors in MDD. A. Circos plot shows TF motif families enriched (Homer; p<0.05) in differentially more (open) or less accessible (closed) ExN1 DARs and differentially less accessible (closed) Mic2 DARs (width of links represent –log10(p-value) scaled within each cluster). Each segment is colored by individual TF families representing TF motifs significantly enriched in DARs. Only TFs present in both ArchR peak-motif matrix and snRNA-seq integrated gene-expression matrix are plotted. B: Workflow for TF prioritization. Left to right: (i) Top 4 most significant TF families (FDR<0.05) in ExN1 (top) and Mic2 (bottom); (ii) DNA motifs of prioritized TFs in each TF family; (iii) Density plots show snRNA-seq integrated gene expression for topmost significant TFs belonging to each family (plotted from the lowest to highest median gene expression) across cells in each cluster (iv) ChromVar based bias-corrected TF motif accessibility deviation z-scores in MDD versus controls (Wilcoxon test, **FDR<0.01, *FDR<0.05; white stars indicate prioritized TFs with the most significant differences in motif accessibility z-scores in MDD versus controls; (v) Number of DARs enriched for the prioritized TF motifs and the number of target-genes associated with

prioritized TFs (% represents TF target-genes which are also MDD DEGs (in MDD bulk or single-nucleus RNA-seq datasets) or MDD-associated genes (in MDD GWAS or PsyGeNET depressive disorder genes). C. TF target-gene regulatory networks for prioritized TFs associated with ExN1 DARs (FDR<0.05; left) and Mic2 DARs (FDR<0.01; right). Nodes in the network are colored by target-gene status of MDD associated genes or MDD DEGs or both. Edges are colored by type of DARs (promoters, introns or distal DARs). An edge is added to the network when a significant DAR contains the prioritized TF motif and connects to genes linked to any DAR through peak-to-gene linkages (r>0.45) and nearest genes to promoter DARs. D. snRNA-seq gene expression module scores for the prioritized TF target-genes in ExN1 (left) and Mic2 (right) compared between MDD versus control groups (Wilcoxon test) in the corresponding snRNA-seq clusters. E-F. Tn5 bias subtracted motif footprints of prioritized TFs in MDD versus control subjects (left) and snRNA-seq gene-expression module scores for target-genes of prioritized TFs in MDD versus controls in corresponding snRNA-seq clusters (right; Wilcoxon test) for E. ExN1 and F. Mic2.

Figure 4. Genetic overlap with cell- and MDD- specific chromatin structure. A-B: Heatmaps scaled by –log10(p-value) (LDSC z-score test) for heritability enrichment of GWAS SNPs belonging to each trait in each cluster using A. cluster-specific marker cCREs B. cluster-specific candidate DARs C. Overrepresentation of cluster-specific candidate DARs (permutation test; n=10,000; p<0.05) within significant heritable LD blocks identified for MDD (HESS; p<0.05). Boxplots represent expected proportions of chromosome-matched permuted cCREs within any LD block. Circles show the observed proportion of candidate DARs within significantly heritable MDD-associated LD blocks. D. Word cloud represents TF motif families significantly enriched (q<0.05; Homer) in both ExN1-specific marker cCREs and candidate DARs located within MDD-associated LD blocks. The size of TF families is scaled by the total number of distinct TF members (belonging to these families) identified to be significantly enriched (q<0.05). Density plots shows snRNA-seq integrated gene-expression of enriched TFs across ExN1 cells.

Figure 5: MDD sSNPs with significant allelic-effects on chromatin accessibility landscape in ExN1 cluster. A. Left to right: MDD sSNPs with significant allelic-effects in ExN1 cluster; ExN1 sSNP-associated nearest genes, linked genes most significantly (FDR<0.05) correlated

with cCREs associated with sSNPs (peak-to-gene linkages; in grey are genes that did not pass FDR); DLPFC eQTLs and sQTLs genes associated with sSNPs; sSNPs categorized based on accessibility disruption confidence sets; TF binding sites potentially disrupted by sSNPs (listed in the order of significance for TF motif match). B. Example plot for ExN1 sSNP (rs2276138). Top to bottom: snATAC-seq-derived pseudo-bulk tracks for each cell-type and ExN1 cluster, Manhattan plot showing MDD sSNP (chr11: T>C; rs2276138) overlapping ExN marker cCRE (peak-to-gene linkage: SPTBN2 gene) and ExN1 marker cCRE (peak-to-gene linkage: SYT12 gene), ExN1 candidate DARs and marker cCREs within the plotted region, DLPFC neuronal and non-neuronal (NeuN+/-) H3K27ac and H3K4me3 ChIP-seq peaks merged across each subject 25, DLPFC NeuN+/- promoter anchored HiC fragments and HiC loops49, snATAC-seq peak-to-peak co-accessible loops (r>0.45) within 10kbps of sSNP associated cCRE, SYT12 gene locus (chr11:67006771-67050863) with the most significant peak-to-gene linkage (r=0.42) with cCRE associated with sSNP, gkmExplain importance scores for each base in 50-bp region surrounding sSNP for A1 and A2 alleles from ExN1 gkm-SVM model, the predicted TF binding site (RFX3) affected by sSNP is shown at the bottom. C. Example plot for ExN1 sSNP (rs11601579). Top to bottom: snATAC-seq-derived pseudo-bulk tracks for each cell-type and ExN1 cluster (normalized by reads in TSS), Manhattan plot showing MDD lead SNP, SNPs in LD (r>0.8), and MDD sSNP (chr11: C>A; rs11601579) overlapping ExN and ExN1 marker cCRE (peak-to-gene linkage: SHANK2-AS1 gene), ExN1 candidate DARs and marker cCREs within the plotted region, DLPFC neuronal and non-neuronal (NeuN+/-) H3K27ac and H3K4me3 ChIP-seq peaks merged across each subject25, DLPFC NeuN+/- promoter anchored HiC fragments and HiC loops49, snATAC-seq peak-to-peak co-accessible loops (r>0.45) within 10kbps of sSNP associated cCRE, SHANK2-AS1 gene locus (chr11-70625973-70635733) with the most significant peak-to-gene linkage (r=0.6) with cCRE associated with sSNP, gkmExplain importance scores for each base in 50-bp region surrounding sSNP for A1 and A2 alleles from ExN1 gkm-SVM model, the predicted TF binding site (GLI2) affected by the sSNP is shown at the bottom.

Figure 6: MDD sSNPs with significant allelic-effects on chromatin accessibility landscape in Mic2 cluster. A. Left to right: MDD sSNPs with significant allelic-effects in Mic2 cluster;

nearest genes to sSNPs, linked genes most significantly (FDR<0.05) correlated with cCREs associated with sSNPs (peak-to-gene linkages, in grey are genes that did not pass FDR); DLPFC eQTLs and sQTLs genes associated with sSNPs; sSNPs categorized based on accessibility disruption confidence sets; TF binding sites potentially disrupted by sSNPs (listed in the order of significance for TF motif match) B. Example plot for Mic2 sSNP (rs10210757). Top to bottom: snATAC-seq-derived pseudo-bulk tracks for each cell-type and Mic2 cluster (normalized by reads in TSS), Manhattan plot showing MDD lead SNP, SNPs in LD (r>0.8), and MDD sSNP (chr2: C>T; rs10210757) overlapping Mic marker cCRE (also a candidate DAR in Mic; FDR=0.1), Mic candidate DARs and marker cCREs within the plotted region, DLPFC neuronal and non-neuronal (NeuN+/-) H3K27ac and H3K4me3 ChIP-seq peaks merged across each subject25, DLPFC NeuN+/- promoter anchored HiC fragments and HiC loops49, snATAC-seq peak-to-peak co-accessible loops (r>0.45) within 10kbps of sSNP associated cCRE, ARL4C gene locus (chr2:234493041-234497081) has HiC linkages with sSNP region, gkmExplain importance scores for each base in 50-bp region surrounding sSNP for A1 and A2 alleles from Mic2 gkm-SVM model, the predicted TF binding site (FLI1) affected by the sSNP is shown at the bottom. C. Example plot for Mic2 sSNP (rs5995992). Top to bottom: snATAC-seq-derived pseudo-bulk tracks for each cell-type and Mic2 cluster (normalized by reads in TSS), Manhattan plot showing MDD lead SNP, SNPs in LD (r>0.8), and MDD sSNP (chr22: C>T; rs5995992) overlapping Mic2 candidate DAR (FDR=0.06) (peak-to-gene linkage: TOB2 gene), Mic2 candidate DARs and marker peaks in the plotted region, DLPFC neuronal and non-neuronal (NeuN+/-) H3K27ac and H3K4me3 ChIP-seq peaks merged across each subject25, DLPFC NeuN+/- promoter anchored HiC fragments and HiC loops49, snATAC-seq peak-to-peak co-accessibility (r>0.45) for cCRE within 10kbps of sSNP associated cCRE, TOB2 gene locus (chr22:41433494-41446801) with the most significant peak-to-gene linkage (r=0.8) with sSNP associated cCRE, No TF binding site was identified to be significantly disrupted by this sSNP.

Extended Figure 1. Annotation of snATAC-seq cell-types and clusters. A. UMAP plot based on chromatin accessibility in cCREs identified at cluster resolution and colored by cell-type labels predicted from MDD snRNA-seq from the same samples. B. Histogram shows MDD snRNA-seq cell-type label-transfer prediction scores across snATAC-seq cells (median

prediction score: 0.95; in red). C. K-nearest neighbors (KNN; k=10) based on snATAC-seq peak-accessibility z-scores (left) and snRNA-seq integrated gene-expression z-scores (right) for significant peak-to-gene linkages (r>0.45) grouped by snATAC-seq cell-types. D. UMAP plot based on chromatin accessibility in cCREs identified at cluster resolution and colored by snATAC-seq clusters. E. Distribution of MDD snRNA-seq cell-type label-transfer prediction scores across snATAC-seq cells in each cluster. F. MetaNeighbor best hit plot depicting correspondence (AUROC) between cell-types in MDD snATAC-seq and snRNA-seq datasets. G. Dotplot depicting module scores (size of dots represents percentage of cells) in each snATAC-seq cluster computed based on gene-activity (ArchR GeneScore matrix) for cell-type marker genes catalogued in BrainInABlender104. H. Heatmap shows the overlap enrichment of the most significant fluorescence-activated nuclei (FAN)-sorted bulk ATAC-seq cell-type marker peaks in snATAC-seq cluster-specific marker peaks (Hypergeometric test; FDR<0.05 & Log2FC > 0.5; n=1; MGAS: Microglia and Astrocytes; ExN: Excitatory neurons; InN: Inhibitory neurons; Oli: Oligodendrocyte-lineage cells). I. Overlap of PsychEncode ChIP-seq histone modification peaks (H3K27ac, marks active enhancers; H3K27ac+H3K4me3: marks promoters) from cortical cells with cluster-specific marker cCREs (top) and cCREs in significant peak-to-gene linkages (r>0.45) (bottom).

Extended Figure 2: Comparison of snATAC-seq with published single-cell datasets. A. UMAP plot based on chromatin accessibility in cluster resolution cCREs colored by cell-type labels predicted from the reference PFC snATAC-seq24. Histogram shows reference label-transfer prediction scores across cells in our data (median prediction score: 0.99; in red). B. Distribution of reference cell-type label-transfer prediction scores across snATAC-seq cells in each cluster. C-D. MetaNeighbor AUROC plots show correspondence of MDD snATAC-seq gene promoter-accessibility with C. MDD snRNA-seq clusters20 and D. White matter postmorterm human brain tissue snRNA-seq clusters107. E. Violin plot shows NR4A2 gene expression in excitatory neuronal clusters identified in MDD snRNA-seq20. F. Gene expression module score differences in MDD versus controls in corresponding snRNA-seq clusters. NR4A2+ excitatory neuronal cluster (ExN16_L56) for ExN1 DAR-linked genes (left) linked to significantly less accessible (down; FDR<0.05) or more accessible DARs in MDD (up;

FDR<0.05) and gene expression module score differences in snRNA-seq Mic cluster for genes liked to differentially less accessible Mic2 DARs (FDR<5%; right).

Extended Figure 3: MDD risk variants identified in each cluster. A. Flowchart for identification of significant MDD SNPs (sSNPs) with allelic-effects on cluster-specific chromatin accessibility predicted to be significant across all three cluster-specific gkmSVM models (ISM, deltaSVM, gkmexplain) B. Stacked bar plots show overview of cdSNPs (top) and sSNPs (bottom) in each cluster (left to right: counts in each cluster, % SNPs that are significant eQTLs and sQTLs in the DLPFC (FDR<0.05), % SNPs mapping to cluster-specific marker cCREs and candidate DARs. C. Pie charts depicting (first row) percentage of cdSNPs in MDD-associated SNPs, percentage of sSNPs in cdSNPs, and sSNPs categorized based on accessibility disruption confidence sets; (second row) percentage of significant DLPFC eQTLs in MDD-associated SNPs, cdSNPs, sSNPs. Arrow represents significant enrichment (p<0.001; hypergeometric test) of eQTLs in cdSNPs compared to all MDD-associated SNPs; (third row) depicting percentage of sQTLs in MDD-associated sSNPs, cdSNPs, sSNPs. Arrows represents the significant enrichment (p<5e-08; hypergeometric test) of sQTLs in cdSNPs compared to all MDD-associated SNPs. D. DLPFC eQTLs and sQTLs were significantly enriched (binomial test, p<0.05) in marker cCREs and candidate DARs compared to the expected genomic coverage. E. Heatmap showing sSNP overlap across cell-types. Each cell in the plot is scaled with respect to the total number of sSNPs available for the cell type in the column.

Extended Figure 4: Gene ontology (GO) pathways associated with MDD risk variants. GO pathways enriched (FDR<0.05) in genes associated with MDD sSNPs A. nearest genes, B linked genes (peak-to-gene linkages, r>0.45). C. GO pathway enrichment (FDR<0.05) for ExN1 sSNP

associated genes (nearest, linked, and eQTLs). D. GO pathway enrichment (FDR<0.05) for Mic2 sSNP associated genes (nearest, linked, and eQTLs).

Extended Figure 5. Examples of MDD sSNPs with allelic-effects exclusively on chromatin accessibility in ExN1 cluster. A. Example plot for ExN1 sSNP (rs9299525). Top to bottom: snATAC-seq-derived pseudo-bulk tracks for each cell-type and ExN1 cluster (normalized by reads in TSS), Manhattan plot showing MDD sSNP (chr10: G>A; rs9299525) overlapping ExN marker cCRE (peak-to-gene linkage: KAT6B gene), ExN1 candidate DARs and marker cCREs within the plotted region, DLPFC neuronal and non-neuronal (NeuN+/-) H3K27ac and H3K4me3 ChIP-seq peaks merged across each subject25, DLPFC neuronal and non-neuronal (NeuN+/-) H3K27ac and H3K4me3 ChIP-seq peaks merged across each subject25, DLPFC NeuN+/- promoter anchored HiC fragments and HiC loops49, snATAC-seq peak-to-peak co-accessible loops (r>0.45) within 10kbps of sSNP associated cCRE, KAT6B gene locus (chr10:74824927-75032624) with the most significant peak-to-gene linkage (r=0.4) with cCRE associated with sSNP, gkmExplain importance scores for each base in 50-bp region surrounding sSNP for A1 and A2 alleles from ExN1 gkm-SVM model corresponding, the predicted TF binding site (ZFP3) affected by the sSNP is shown at the bottom. B. Example plot for ExN1 sSNP (rs6063348). Top to bottom: snATAC-seq-derived pseudo-bulk tracks for each cell-type and ExN1 cluster (normalized by reads in TSS), Manhattan plot showing MDD sSNP (chr20: C>G; rs6063348) overlapping ExN1 candidate DAR (FDR=0.15) (peak-to-gene linkage: KCNB1 gene), ExN1 candidate DARs and marker cCREs within the plotted region, DLPFC neuronal and non-neuronal (NeuN+/-) H3K27ac and H3K4me3 ChIP-seq peaks merged across each subject25, DLPFC NeuN+/- promoter anchored HiC fragments and HiC loops49, snATAC-seq peak-to-peak co-accessible loops (r>0.45) within 10kbps of sSNP associated cCRE, KCNB1 gene locus (chr20-49293394-49484297) with the most significant peak-to-gene linkage (r=0.4) with cCRE associated with sSNP, No TF binding site was identified to be significantly disrupted by this sSNP.

Extended Figure 6. Functional characterization of ExN1 and Mic2 clusters. A. MetaNeighbor AUROC plot shows the strongest correspondence of snATAC-seq ExN1 cluster

with ExN layer 6 IT Car3 neurons in Allen brain institute (ABI) snRNA-seq from the motor cortex (M1)59. B. Violin plot depicts promoter-accessibility for NR4A2 gene in ExN clusters. C. Tn5 bias subtracted NR4A2 motif footprints aggregated in each of the ExN clusters. D. StringDB analysis depicts protein-protein interactions (PPI) between ExN1 marker NR4A2 TF (circled in blue) and MDD-associated TFs significantly enriched ($q<0.05$; Homer) in ExN1 DARs (circled in red) (interaction evidence>0.7 (high); direct interactions: FOS, ASCL1, FOSB and indirect: EGR1). E-F. Module scores computed using E. promoter-accessibility and F. snRNA-seq integrated gene-expression for GM and WM specific microglia marker genes. Boxplot shows significant differences in module scores between snATAC-seq microglial clusters (Mic1 and Mic2; Wilcoxon, $p<0.05$). G. 10x Visium spatial gene-expression in male and female DLPFC sections colored by the label-transfer prediction scores from snATAC-seq microglial clusters, Mic1 (top) and Mic2 (bottom), using promoter-accessibility (left) and snRNA-seq integrated gene-expression (right) H-I. Module scores computed using H. promoter accessibility and I. snRNA-seq integrated gene-expression for MDD-associated significantly down-regulated and up-regulated genes in cortical GM depression disease-associated microglia (depDAM)75. Boxplot shows significant differences in module scores between snATAC-seq microglial clusters (Mic1 and Mic2; Wilcoxon, $p<0.05$).

Extended Figure 7: Proposed model for MDD associated regulatory mechanisms in deep-layer excitatory neurons (ExN1) and grey matter microglia (Mic2) through which genetic and epigenetic factors may mediate disease pathology.

Supplementary Figure 1: Demultiplexing of pooled male and female subjects. A. UMAP plot based on chromatin accessibility in 500bp genomic bins and colored by sex. B. The table shows R2 values reflecting cluster, individual and sex-specific variation explained by UMAP1 and UMAP2 embeddings. C. PCA of top 1% most variable cCREs identified using pseudo-bulked counts per subject and colored by the sex identified for each demultiplexed subject. D.

Best match between genotypes of demultiplexed subjects piled-up using 1000 Genomes based common variants and genotypes obtained from the blood of these individuals. E. Pseudo-bulk chromatin accessibility (normalized by reads in TSS) at *XIST* gene in each of the demultiplexed subject.

Supplementary Figure 2: **Quality-control of snATAC-seq**. A. snATAC-seq fragment-size distribution in each subject. B. Normalized Tn5 insertions profiles aggregated across TSS in each subject. C. Density scatter plot depicting relationship between TSS Enrichment and log(10) unique fragments across 201456 cells. D. Ridge plots showing distribution of total number of fragments (nFrags), TSSEnrichment, and fraction reads in peaks (FRIP) in each of the demultiplexed subjects.

Supplementary Fig 3: **Quality-control of snATAC-seq**. Comparison of following mean variable values across cells in MDD versus control subjects (Wilcoxon test) A. no. of fragments B. TSSEnrichment. C. fraction reads in peaks (FRIP). D. Age. E. post-mortem interval (PMI). F. pH of the brain tissue.

Supplementary Fig 4: Quality-control of snATAC-seq. A. UMAP plots before (left) and after (right) batch-correction colored by batches. B. Bar plots showing proportions of cells contributed by each batch in each cluster. C. Bar plots shows proportions of cells contributed by each subject in each cluster. D. UMAP plot colored by PMI (left) and Bar plots (right) showing proportions of cells contributed by PMI groups in each cluster. E. UMAP plot colored by age of subjects (left) and Bar plots (right) showing proportions of cells contributed by each age group in each cluster. F. Violin plots depicting distribution of FRIP across individual cells in each cluster.

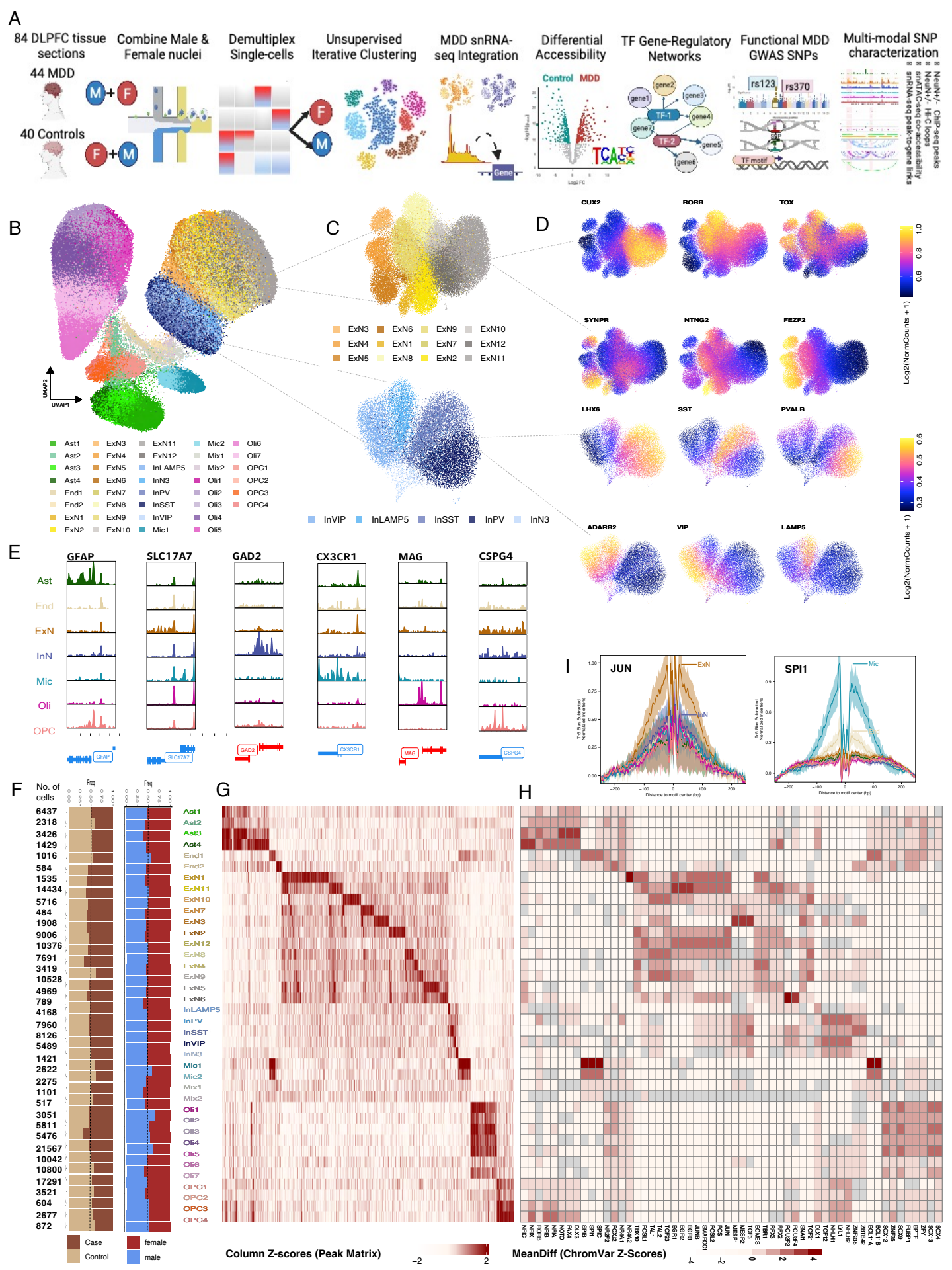
Supplementary Figure 5: **Differential accessibility in MDD**. A. Differential accessibility in MDD versus control subjects in each cell-type (Limma-voom; FDR < 0.05 & abs(LogFC) > log2(1.1)) split by the direction of effect in MDD versus controls and colored by the distance to the nearest TSS. B. B. Pie chart depicts the total number of differentially accessible regions across cell-type DARs that are differentially more (open; n=121) or less accessible (closed; n=176) in MDD versus controls subjects. C. PsyGeNet analysis for psychiatric disorders shows the total number of gene-disease associations for genes linked to cell-type DARs (r>0.45). D.

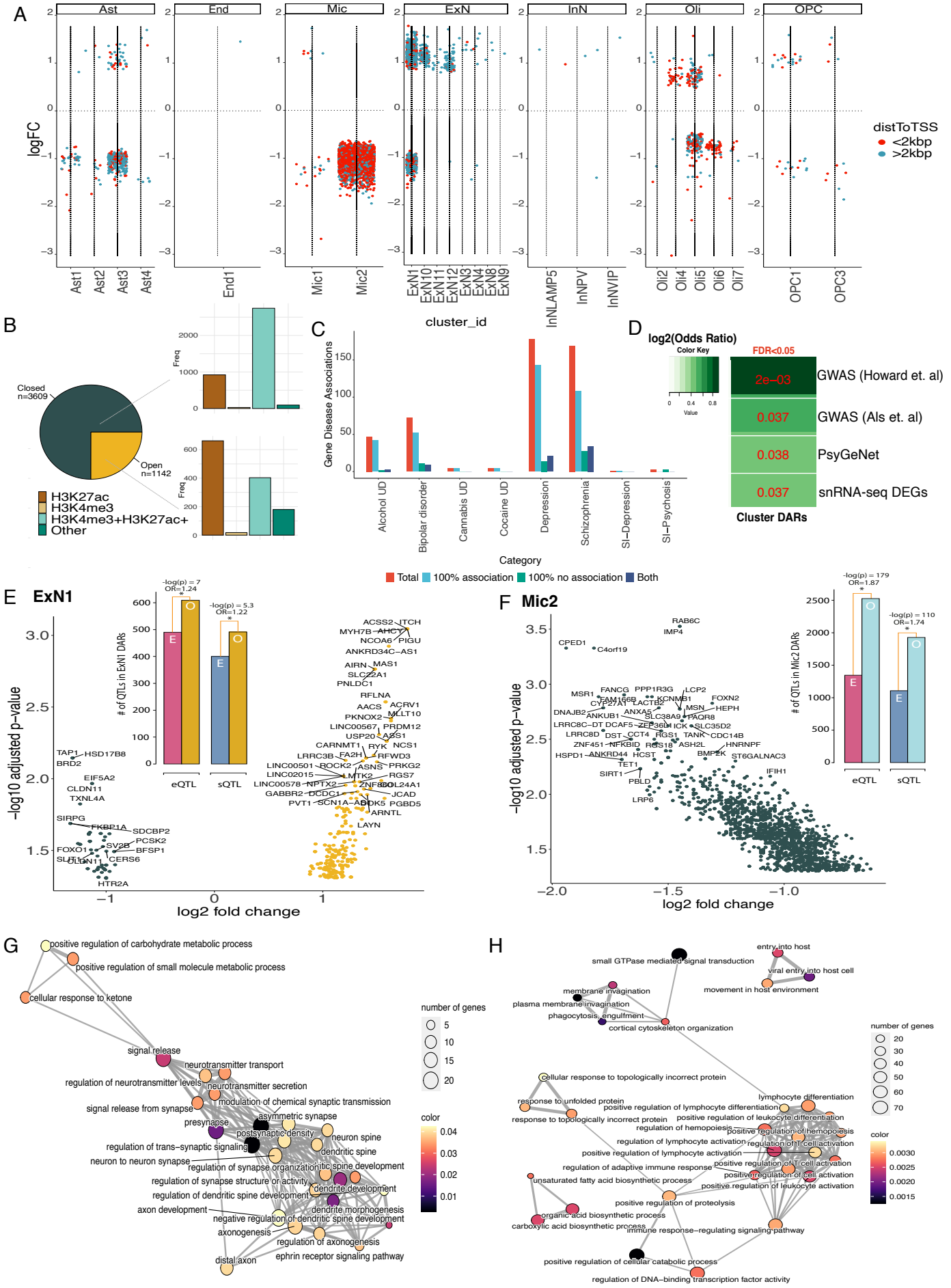
Heatmap (colored by –log10(p-values)) shows GO pathways (Top 10; FDR<0.2) enriched in genes linked (r>0.4) to differentially less (down) or more accessible (up) MDD DARs (FDR<5%) in broad microglia cell-type (Mic) and microglial clusters (Mic1, Mic2). E. Flowchart overview of TF prioritization workflow and construction of target gene-regulatory networks in MDD.

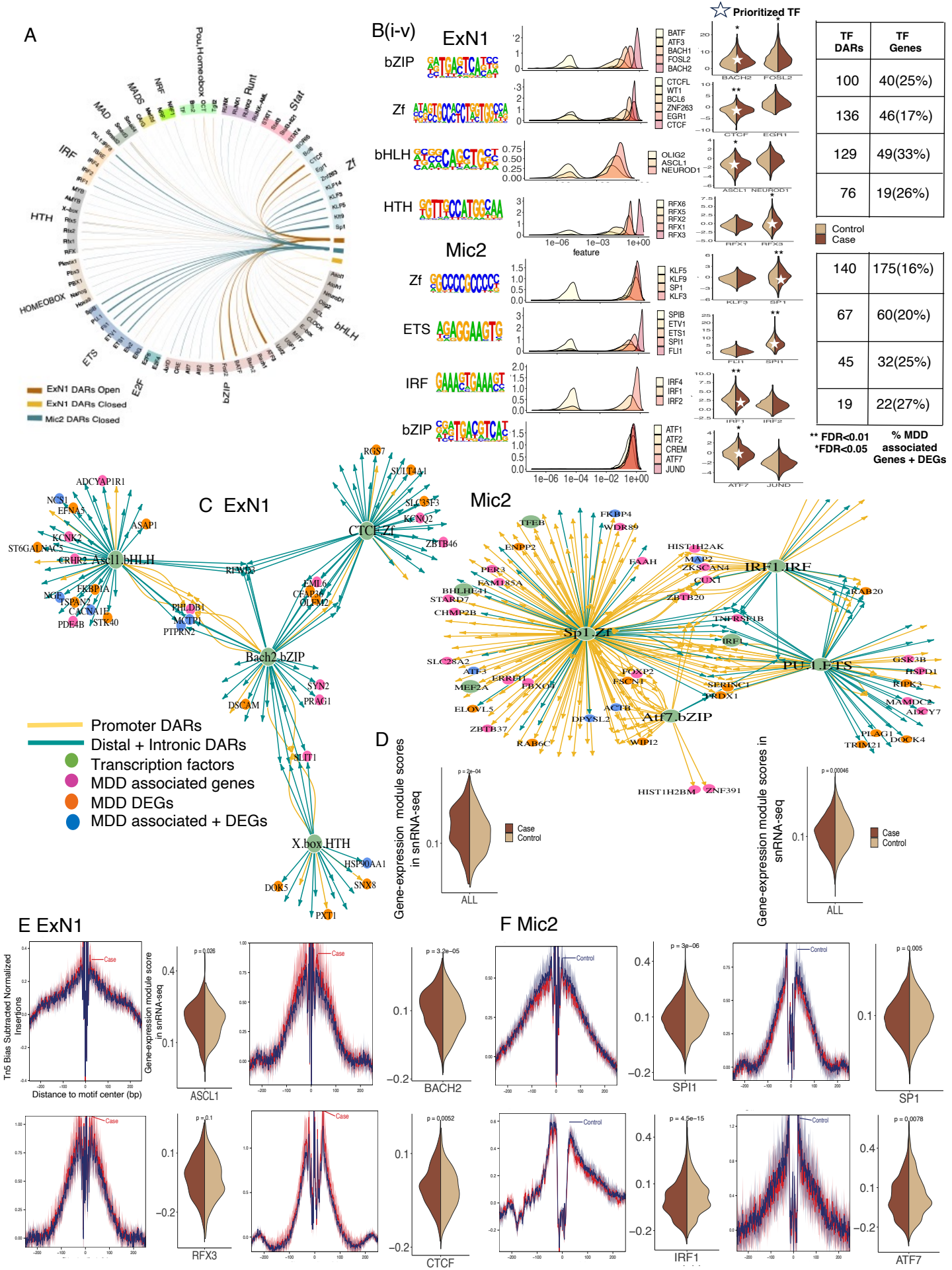
Supplementary Figure 6: **Schematic describing gkm-SVM based variant-scoring workflow employed for each snATAC cluster.** A. Cluster-specific gkm-SVM classifiers were trained to predict cCRE status of input 1001-bp genome sequences via 5-fold cross validation. B. MDD SNPs located within 1001-bp marker cCREs and DARs are pooled to obtain candidate SNP set per cluster (cdSNPs). C. gkm-SVM-based cdSNP variant effect scores were computed by running ISM, deltaSVM and gkmexplain on 201-bp allelic sequences of cdSNPs. D. Statistical significance of cdSNPs were assessed with respect to score-specific null t-distributions fitted to null variant effect scores computed on shuffled (di-nucleotide preserved) 201-bp allelic sequences. For each score type and cluster, cdSNPs whose variant effect scores unanimously lie outside of 95%CI of respective null distributions were coined as statistically significant (sSNP). E. sSNPs were further partitioned into three groups according to the prominence and magnitude of their allelic impact to their immediate (seqlet) as well as 51-bp neighborhood. Additionally, seqlets of sSNPs having high or medium importance were interrogated for potential TFBS matches in CIS-BP 2.0 (TOMTOM, q-value < 0.1). Please see supplementary methods for implementation specific details regarding the workflow, and supplementary results for reliability analysis of workflow components.

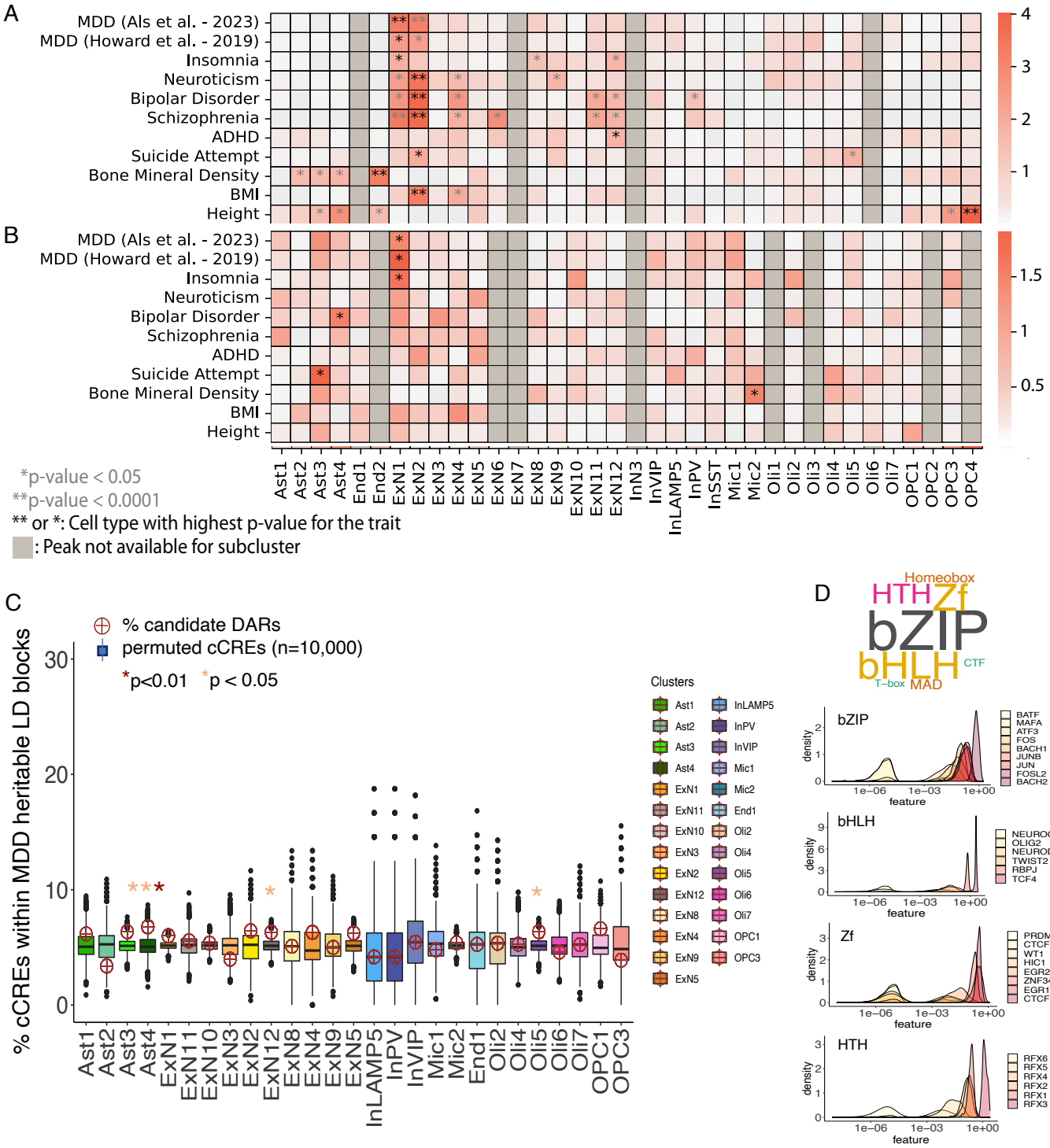
Supplementary Figure 7: **Cluster-specific SVM models for identification of MDD sSNPs.** A. Performance comparison of cluster-specific gkm-SVM classifiers on cross validation test splits (top) and marker cCREs (bottom) with respect to AUC-ROC (left) and AUC-PR (right). B. Pairwise correlation of cdSNP variant effect scores across snATAC clusters. C. Statistical significance of cdSNPs according to ISM, deltaSVM and gkmexplain variant effect scores. D. Distribution of potentially significant cdSNPs into confidence sets. E. Composition of MDD

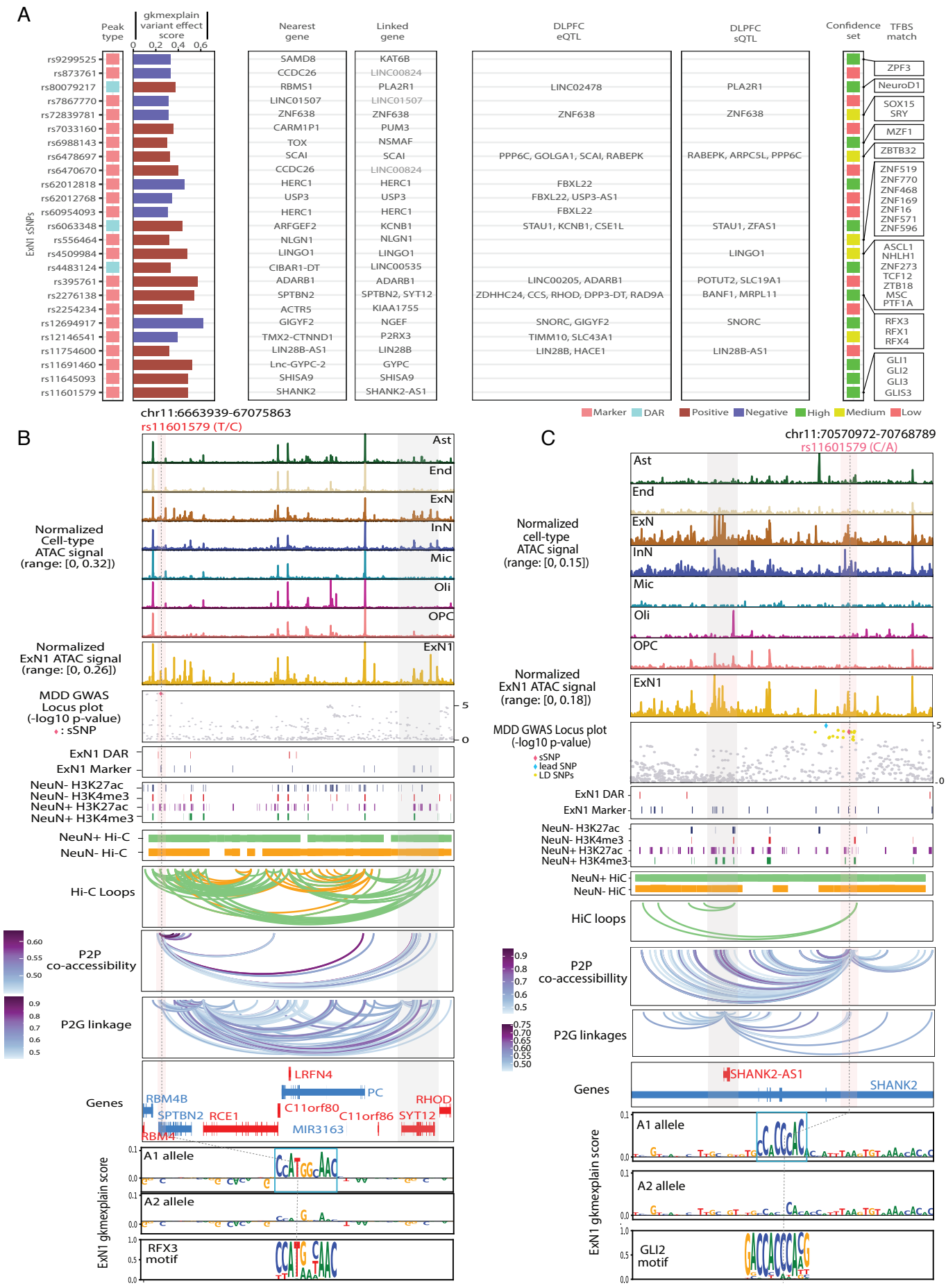
SNP (left), cdSNP (middle) and sSNP (right). (Panel F) Genomic location based categorization of MDD SNPs (top), cdSNPs (middle) and sSNPs (bottom).

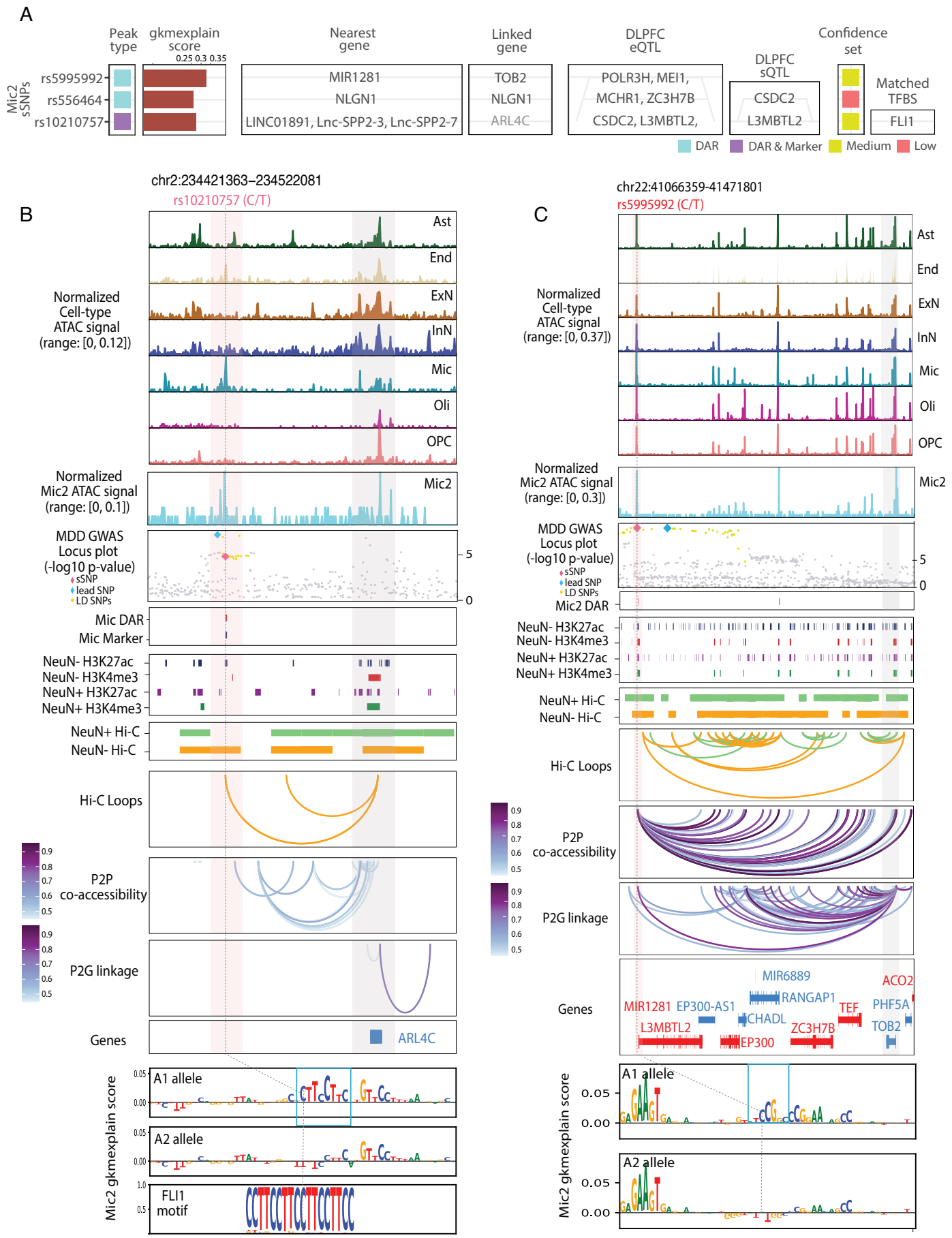

A
84 DLPFC tissue sections
44 MDD
40 Controls
Combine Male & Female nuclei
Demultiplex Single-cells
Unsupervised Iterative Clustering
MDD snRNA-seq Integration
Differential Accessibility
Control
MDD
TF Gene-Regulatory Networks
Functional MDD GWAS SNPs
rs123
rs370
Multi-modal SNP characterization
B
C
D
E
F
G
H
I
UMAP1
UMAP2
Ast1
Ast2
Ast3
Ast4
End1
End2
ExN1
ExN2
ExN3
ExN4
ExN5
ExN6
ExN7
ExN8
ExN9
ExN10
ExN11
ExN12
InN3
InVIP
InLAMP5
InPV
InSST
Mic1
Mic2
Mix1
Mix2
Oli1
Oli2
Oli3
Oli4
Oli5
Oli6
Oli7
OPC1
OPC2
OPC3
OPC4
CUX2
RORB
TOX
SYNPR
NTNG2
FEZF2
LHX6
SST
PVALB
ADARB2
VIP
LAMP5
Log2(NormCounts + 1)
GFAP
SLC17A7
GAD2
CX3CR1
MAG
CSPG4
Ast
End
ExN
InN
Mic
Oli
OPC
JUN
SPI1
Distance to motif center (bp)
No. of cells
Case
Control
female
male
Column Z-scores (Peak Matrix)
MeanDiff (ChromVar Z-Scores)

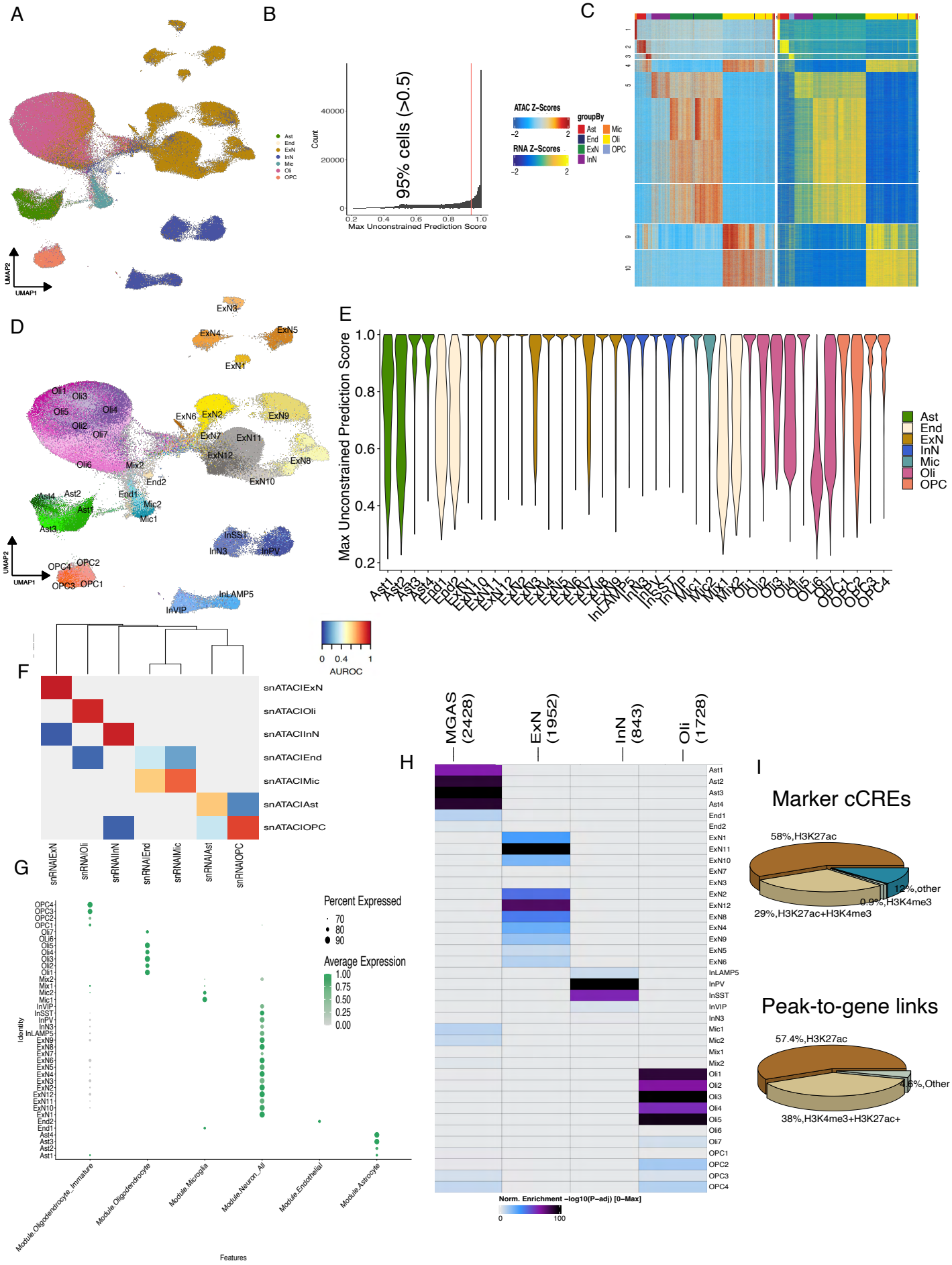

A
Ast
End
Mic
ExN
InN
Oli
OPC
logFC
distToTSS
<2kbp
>2kbp
Ast1
Ast2
Ast3
Ast4
End1
Mic1
Mic2
ExN1
ExN10
ExN11
ExN12
ExN3
ExN4
ExN8
ExN9
InNLAMP5
InNPV
InNVIP
Oli2
Oli4
Oli5
Oli6
Oli7
OPC1
OPC3
B
Closed
n=3609
Open
n=1142
Freq
H3K27ac
H3K4me3
H3K4me3+H3K27ac+
Other
C
cluster_id
Gene Disease Associations
Alcohol UD
Bipolar disorder
Cannabis UD
Cocaine UD
Depression
Schizophrenia
SI−Depression
SI−Psychosis
Category
Total
100% association
100% no association
Both
D
log2(Odds Ratio)
Color Key
Value
FDR<0.05
2e−03
0.037
0.038
0.037
GWAS (Howard et. al)
GWAS (Als et. al)
PsyGeNet
snRNA-seq DEGs
Cluster DARs
E
ExN1
-log(p) = 7
OR=1.24
-log(p) = 5.3
OR=1.22
# of QTLs in ExN1 DARs
eQTL
sQTL
−log10 adjusted p−value
log2 fold change
F
Mic2
-log(p) = 179
OR=1.87
-log(p) = 110
OR=1.74
# of QTLs in Mic2 DARs
G
positive regulation of carbohydrate metabolic process
positive regulation of small molecule metabolic process
cellular response to ketone
signal release
neurotransmitter transport
regulation of neurotransmitter levels
neurotransmitter secretion
signal release from synapse
modulation of chemical synaptic transmission
presynaptic
asymmetric synapse
postsynaptic density
neuron spine
regulation of trans−synaptic signaling
dendritic spine
neuron to neuron synapse
regulation of synapse organization
dendritic spine development
regulation of synapse structure or activity
dendrite development
regulation of dendritic spine development
dendrite morphogenesis
axon development
negative regulation of dendritic spine development
axonogenesis
regulation of axonogenesis
distal axon
ephrin receptor signaling pathway
number of genes
color
H
small GTPase mediated signal transduction
entry into host
viral entry into host cell
movement in host environment
membrane invagination
plasma membrane invagination
phagocytosis, engulfment
cortical cytoskeleton organization
cellular response to topologically incorrect protein
response to unfolded protein
response to topologically incorrect protein
lymphocyte differentiation
positive regulation of lymphocyte differentiation
positive regulation of leukocyte differentiation
regulation of hemopoiesis
positive regulation of hemopoiesis
regulation of lymphocyte activation
regulation of T cell activation
positive regulation of lymphocyte activation
positive regulation of T cell activation
regulation of adaptive immune response
positive regulation of cell activation
unsaturated fatty acid biosynthetic process
positive regulation of leukocyte activation
positive regulation of proteolysis
organic acid biosynthetic process
carboxylic acid biosynthetic process
immune response−regulating signaling pathway
positive regulation of cellular catabolic process
regulation of DNA−binding transcription factor activity
number of genes
color

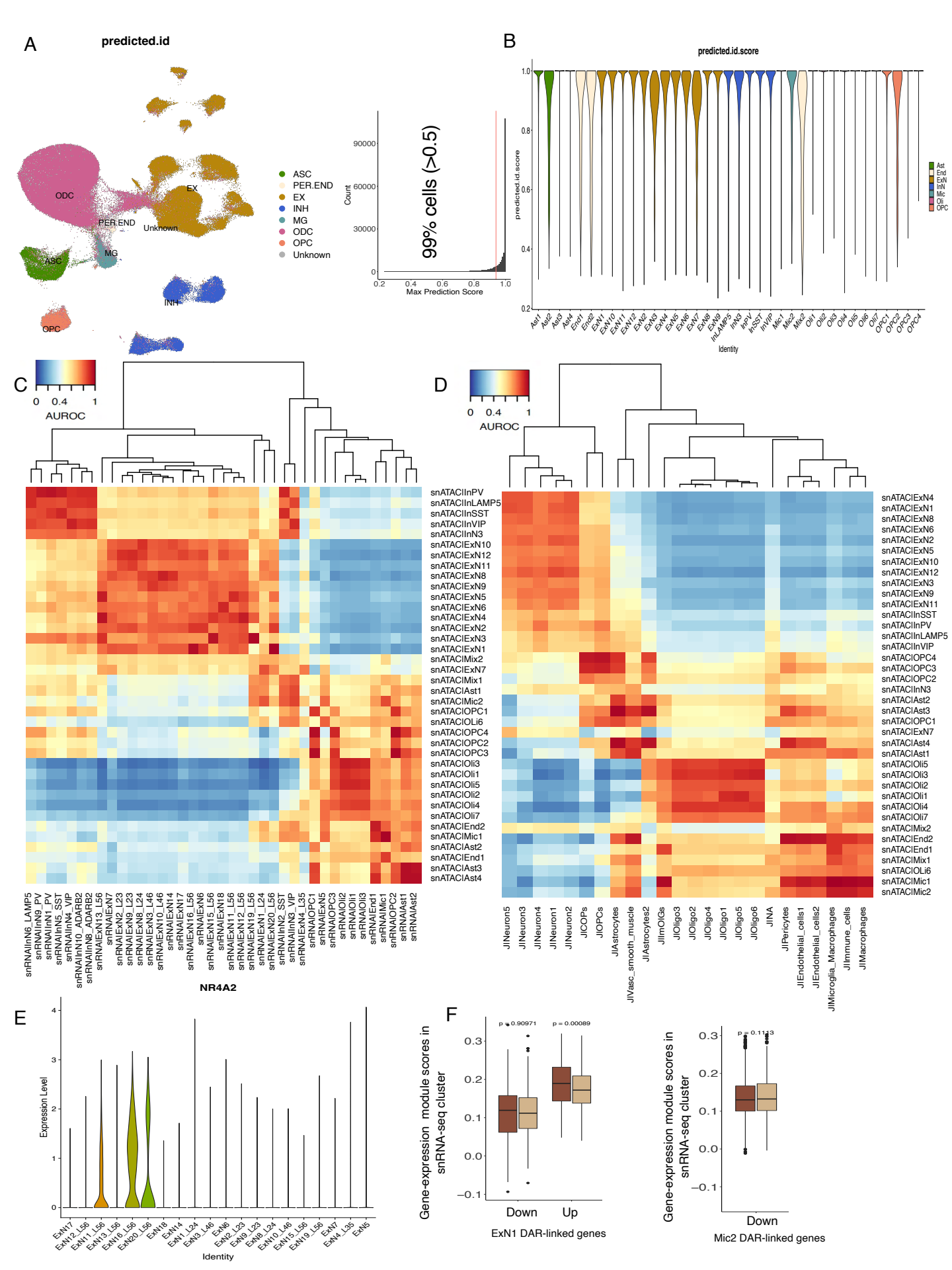

A
B(i-v)
ExN1
Mic2
Prioritized TF
bZIP
Zf
bHLH
HTH
ETS
IRF
feature
TF DARs
TF Genes
100 40(25%)
136 46(17%)
129 49(33%)
76 19(26%)
140 175(16%)
67 60(20%)
45 32(25%)
19 22(27%)
Control
Case
** FDR<0.01
*FDR<0.05
% MDD associated Genes + DEGs
ExN1 DARs Open
ExN1 DARs Closed
Mic2 DARs Closed
C ExN1
Mic2
Promoter DARs
Distal + Intronic DARs
Transcription factors
MDD associated genes
MDD DEGs
MDD associated + DEGs
D
Gene-expression module scores in snRNA-seq
ALL
E ExN1
F Mic2
Tn5 Bias Subtracted Normalized Insertions
Distance to motif center (bp)
ASCL1
BACH2
SPI1
SP1
RFX3
CTCF
IRF1
ATF7

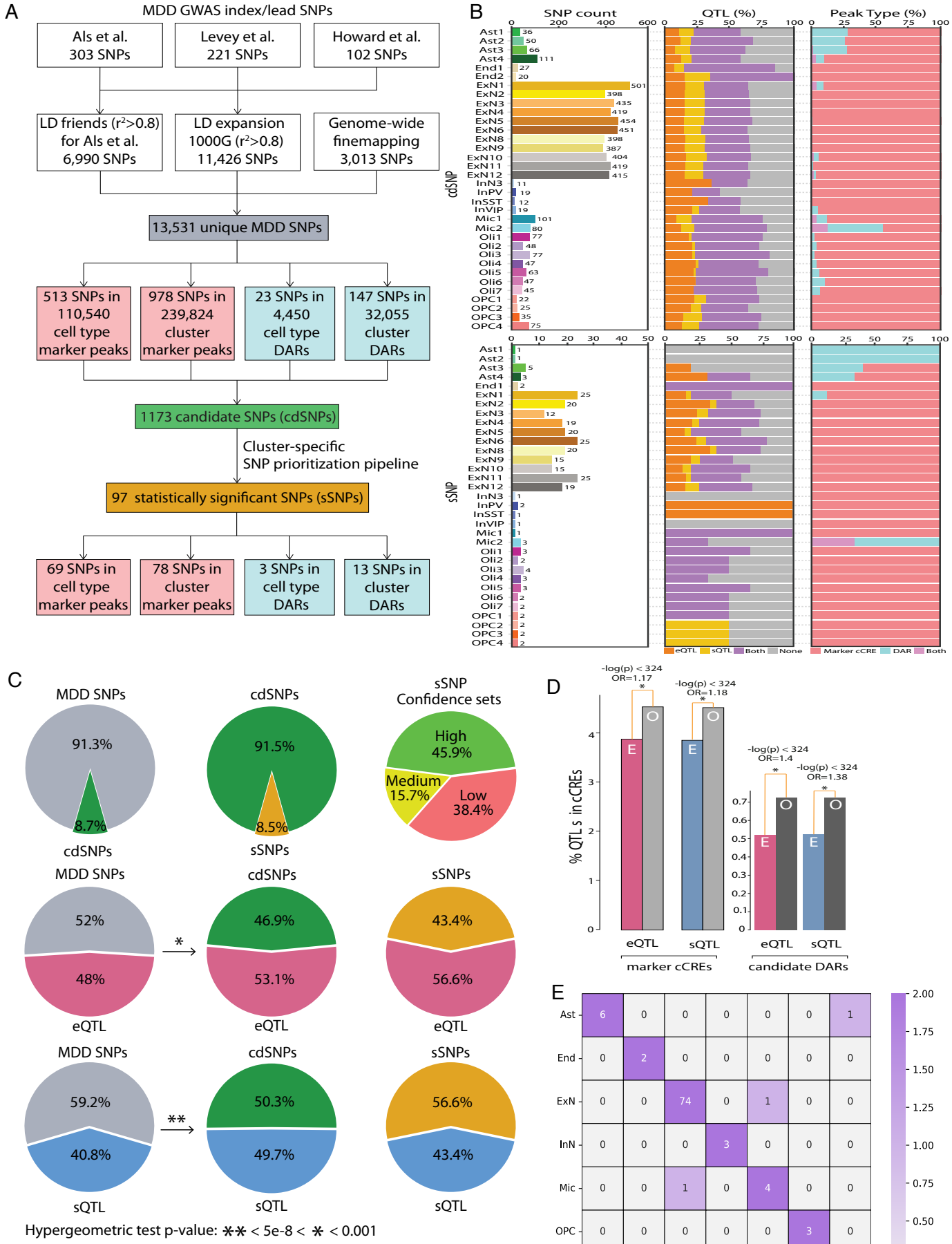

A
MDD (Als et al. - 2023)
MDD (Howard et al. - 2019)
Insomnia
Neuroticism
Bipolar Disorder
Schizophrenia
ADHD
Suicide Attempt
Bone Mineral Density
BMI
Height
B
Ast1 Ast2 Ast3 Ast4 End1 End2 ExN1 ExN2 ExN3 ExN4 ExN5 ExN6 ExN7 ExN8 ExN9 ExN10 ExN11 ExN12 InN3 InVIP InLAMP5 InPV InSST Mic1 Mic2 Oli1 Oli2 Oli3 Oli4 Oli5 Oli6 Oli7 OPC1 OPC2 OPC3 OPC4
*p-value < 0.05
**p-value < 0.0001
** or *: Cell type with highest p-value for the trait
: Peak not available for subcluster
C
% candidate DARs
permuted cCREs (n=10,000)
*p<0.01 *p < 0.05
% cCREs within MDD heritable LD blocks
Clusters
Ast1 Ast2 Ast3 Ast4 ExN1 ExN11 ExN10 ExN3 ExN2 ExN12 ExN8 ExN4 ExN9 ExN5 InLAMP5 InPV InVIP Mic1 Mic2 End1 Oli2 Oli4 Oli5 Oli6 Oli7 OPC1 OPC3
D
Homeobox
HTH Zf
bZIP
bHLH CTF
T-box MAD
bZIP
BATF MAFA ATF3 FOS BACH1 JUNB JUN FOSL2 BACH2
bHLH
NEUROG2 OLIG2 NEUROD1 TWIST2 RBPJ TCF4
Zf
PRDM9 CTCFL WT1 HIC1 EGR2 ZNF341 EGR1 CTCF
HTH
RFX6 RFX5 RFX4 RFX2 RFX1 RFX3
density
feature


A
Peak type
gkmexplain variant effect score
0 0.2 0.4 0.6
Nearest gene
Linked gene
DLPFC eQTL
DLPFC sQTL
Confidence set
TFBS match
ExN1 sSNPs
rs9299525 SAMD8 KAT6B
rs873761 CCDC26 LINC00824
rs80079217 RBMS1 PLA2R1 LINC02478 PLA2R1
rs7867770 LINC01507 LINC01507
rs72839781 ZNF638 ZNF638 ZNF638 ZNF638
rs7033160 CARM1P1 PUM3
rs6988143 TOX NSMAF
rs6478697 SCAI SCAI PPP6C, GOLGA1, SCAI, RABEPK RABEPK, ARPC5L, PPP6C
rs6470670 CCDC26 LINC00824
rs62012818 HERC1 HERC1 FBXL22
rs62012768 USP3 USP3 FBXL22, USP3-AS1
rs60954093 HERC1 HERC1 FBXL22
rs6063348 ARFGEF2 KCNB1 STAU1, KCNB1, CSE1L STAU1, ZFAS1
rs556464 NLGN1 NLGN1
rs4509984 LINGO1 LINGO1 LINGO1
rs4483124 CIBAR1-DT LINC00535
rs395761 ADARB1 ADARB1 LINC00205, ADARB1 POTUT2, SLC19A1
rs2276138 SPTBN2 SPTBN2, SYT12 ZDHHC24, CCS, RHOD, DPP3-DT, RAD9A BANF1, MRPL11
rs2254234 ACTR5 KIAA1755
rs12694917 GIGYF2 NGEF SNORC, GIGYF2 SNORC
rs12146541 TMX2-CTNND1 P2RX3 TIMM10, SLC43A1
rs11754600 LIN28B-AS1 LIN28B LIN28B, HACE1 LIN28B-AS1
rs11691460 Lnc-GYPC-2 GYPC
rs11645093 SHISA9 SHISA9
rs11601579 SHANK2 SHANK2-AS1
ZPF3
NeuroD1
SOX15 SRY
MZF1
ZBTB32
ZNF519 ZNF770 ZNF468 ZNF169 ZNF16 ZNF571 ZNF596
ASCL1 NHLH1 ZNF273 TCF12 ZTB18 MSC PTF1A
RFX3 RFX1 RFX4
GLI1 GLI2 GLI3 GLIS3
Marker DAR Positive Negative High Medium Low
B
chr11:6663939-67075863
rs11601579 (T/C)
Normalized Cell-type ATAC signal (range: [0, 0.32])
Ast
End
ExN
InN
Mic
Oli
OPC
Normalized ExN1 ATAC signal (range: [0, 0.26])
ExN1
MDD GWAS Locus plot (-log10 p-value)
: sSNP
5
0
ExN1 DAR
ExN1 Marker
NeuN- H3K27ac
NeuN- H3K4me3
NeuN+ H3K27ac
NeuN+ H3K4me3
NeuN+ Hi-C
NeuN- Hi-C
Hi-C Loops
P2P co-accessibility
0.60 0.55 0.50
P2G linkage
0.9 0.8 0.7 0.6 0.5
Genes
LRFN4
PC
RHOD
RBM4B
SPTBN2
RCE1
C11orf80
C11orf86
SYT12
MIR3163
RBM4
ExN1 gkmexplain score
A1 allele
A2 allele
RFX3 motif
C
chr11:70570972-70768789
rs11601579 (C/A)
Normalized cell-type ATAC signal (range: [0, 0.15])
Ast
End
ExN
InN
Mic
Oli
OPC
Normalized ExN1 ATAC signal (range: [0, 0.18])
ExN1
MDD GWAS Locus plot (-log10 p-value)
sSNP
lead SNP
LD SNPs
5
0
ExN1 DAR
ExN1 Marker
NeuN- H3K27ac
NeuN- H3K4me3
NeuN+ H3K27ac
NeuN+ H3K4me3
NeuN+ HiC
NeuN- HiC
HiC loops
P2P co-accessibility
0.9 0.8 0.7 0.6 0.5
P2G linkages
0.75 0.70 0.65 0.60 0.55 0.50
Genes
SHANK2-AS1
SHANK2
ExN1 gkmexplain score
A1 allele
A2 allele
GLI2 motif

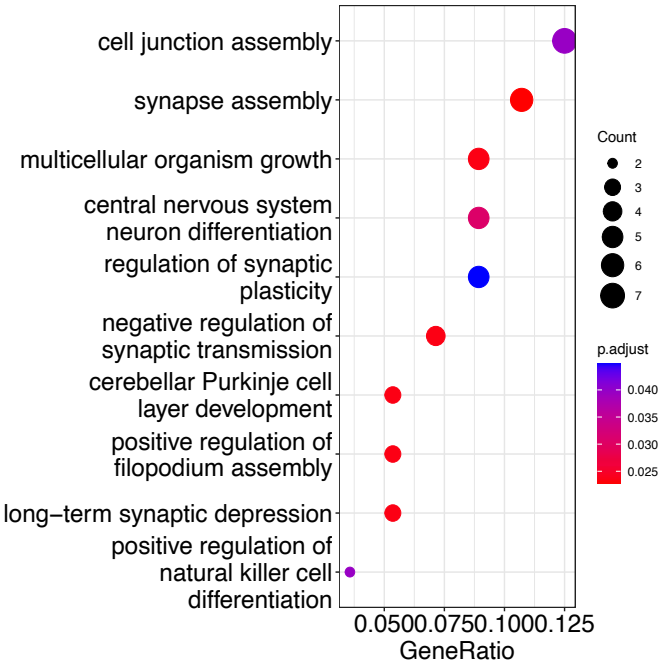

A
Peak type
gkmexplain score
0.25 0.3 0.35
Nearest gene
Linked gene
DLPFC eQTL
DLPFC sQTL
Confidence set
Mic2 sSNPs
rs5995992
rs556464
rs10210757
MIR1281
NLGN1
LINC01891, Lnc-SPP2-3, Lnc-SPP2-7
TOB2
NLGN1
ARL4C
POLR3H, MEI1,
MCHR1, ZC3H7B
CSDC2, L3MBTL2,
CSDC2
L3MBTL2
Matched TFBS
FLI1
DAR
DAR & Marker
Medium
Low
B
chr2:234421363–234522081
rs10210757 (C/T)
Normalized Cell-type ATAC signal (range: [0, 0.12])
Ast
End
ExN
InN
Mic
Oli
OPC
Normalized Mic2 ATAC signal (range: [0, 0.1])
Mic2
MDD GWAS Locus plot (-log10 p-value)
sSNP
lead SNP
LD SNPs
Mic DAR
Mic Marker
NeuN- H3K27ac
NeuN- H3K4me3
NeuN+ H3K27ac
NeuN+ H3K4me3
NeuN+ Hi-C
NeuN- Hi-C
Hi-C Loops
P2P co-accessibility
P2G linkage
Genes
ARL4C
A1 allele
A2 allele
FLI1 motif
Mic2 gkmexplain score
C
chr22:41066359-41471801
rs5995992 (C/T)
Normalized Cell-type ATAC signal (range: [0, 0.37])
Normalized Mic2 ATAC signal (range: [0, 0.3])
Mic2 DAR
MIR6889
ACO2
RANGAP1
MIR1281
EP300-AS1
TEF
PHF5A
CHADL
L3MBTL2
ZC3H7B
TOB2
EP300


A
UMAP2
UMAP1
Ast
End
ExN
InN
Mic
Oli
OPC
B
95% cells (>0.5)
Count
Max Unconstrained Prediction Score
C
ATAC Z-Scores
RNA Z-Scores
groupBy
D
E
Max Unconstrained Prediction Score
F
AUROC
snATACIExN
snATACIOli
snATACIInN
snATACIEnd
snATACIMic
snATACIAst
snATACIOPC
G
Percent Expressed
Average Expression
Identity
Features
Module.Oligodendrocyte_Immature
Module.Oligodendrocyte
Module.Microglia
Module.Neuron_All
Module.Endothelial
Module.Astrocyte
H
MGAS (2428)
ExN (1952)
InN (843)
Oli (1728)
Norm. Enrichment −log10(P−adj) [0−Max]
I
Marker cCREs
58%,H3K27ac
12%,other
0.9%,H3K4me3
29%,H3K27ac+H3K4me3
Peak-to-gene links
57.4%,H3K27ac
4.6%,Other
38%,H3K4me3+H3K27ac+

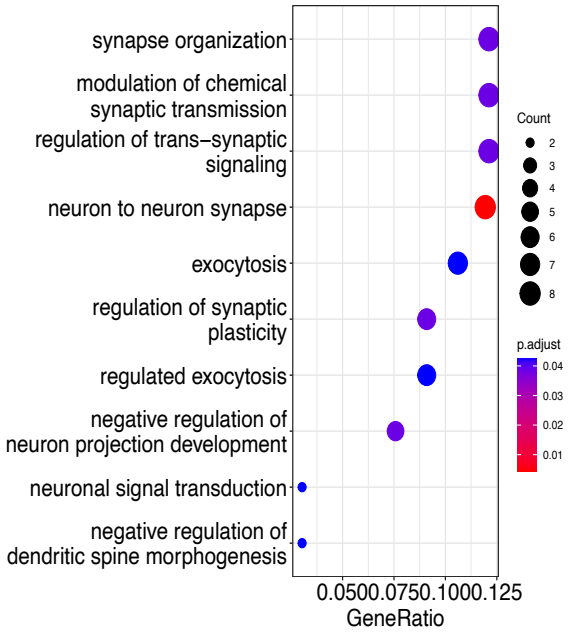

A
predicted.id
ODC
EX
PER.END
Unknown
ASC
MG
INH
OPC
ASC
PER.END
EX
INH
MG
ODC
OPC
Unknown
Count
Max Prediction Score
99% cells (>0.5)
B
predicted.id.score
Identity
C
AUROC
D
AUROC
NR4A2
E
Expression Level
Identity
F
Gene-expression module scores in snRNA-seq cluster
p = 0.90971
p = 0.00089
Down
Up
ExN1 DAR-linked genes
p = 0.1113
Down
Mic2 DAR-linked genes


A
MDD GWAS index/lead SNPs
Als et al. 303 SNPs
Levey et al. 221 SNPs
Howard et al. 102 SNPs
LD friends (r²>0.8) for Als et al. 6,990 SNPs
LD expansion 1000G (r²>0.8) 11,426 SNPs
Genome-wide finemapping 3,013 SNPs
13,531 unique MDD SNPs
513 SNPs in 110,540 cell type marker peaks
978 SNPs in 239,824 cluster marker peaks
23 SNPs in 4,450 cell type DARs
147 SNPs in 32,055 cluster DARs
1173 candidate SNPs (cdSNPs)
Cluster-specific SNP prioritization pipeline
97 statistically significant SNPs (sSNPs)
69 SNPs in cell type marker peaks
78 SNPs in cluster marker peaks
3 SNPs in cell type DARs
13 SNPs in cluster DARs
B
SNP count
QTL (%)
Peak Type (%)
cdSNP
sSNP
eQTL sQTL Both None
Marker cCRE DAR Both
C
MDD SNPs
cdSNPs
sSNP Confidence sets
91.3%
8.7%
91.5%
8.5%
High 45.9%
Medium 15.7%
Low 38.4%
cdSNPs
sSNPs
52%
48%
46.9%
53.1%
43.4%
56.6%
eQTL
59.2%
40.8%
50.3%
49.7%
56.6%
43.4%
sQTL
Hypergeometric test p-value: ** < 5e-8 < * < 0.001
D
-log(p) < 324 OR=1.17
-log(p) < 324 OR=1.18
-log(p) < 324 OR=1.4
-log(p) < 324 OR=1.38
% QTLs in cCREs
eQTL
sQTL
marker cCREs
candidate DARs
E
Ast End ExN InN Mic OPC Oli

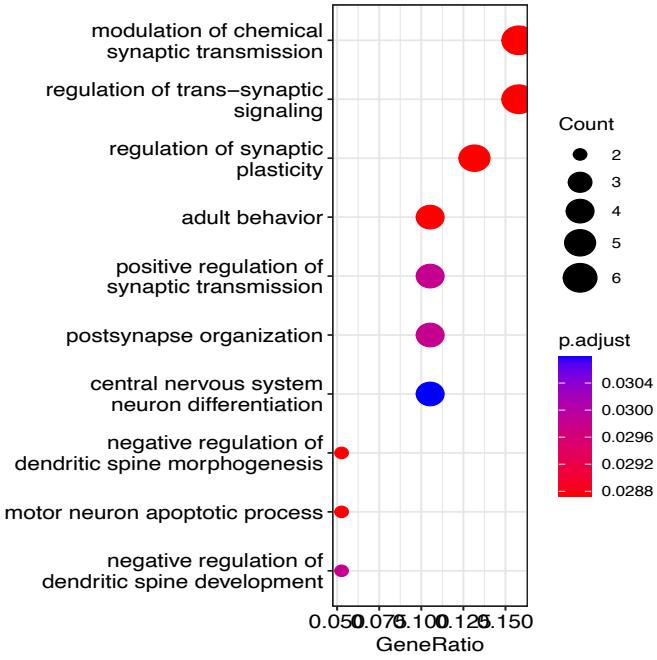
A MDD sSNPs GO (Nearest genes)



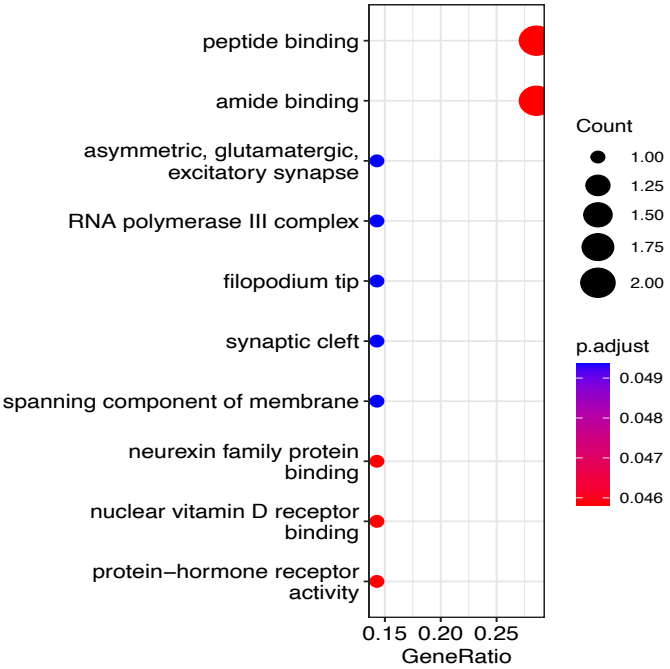
B MDD sSNPs GO (Linked genes)

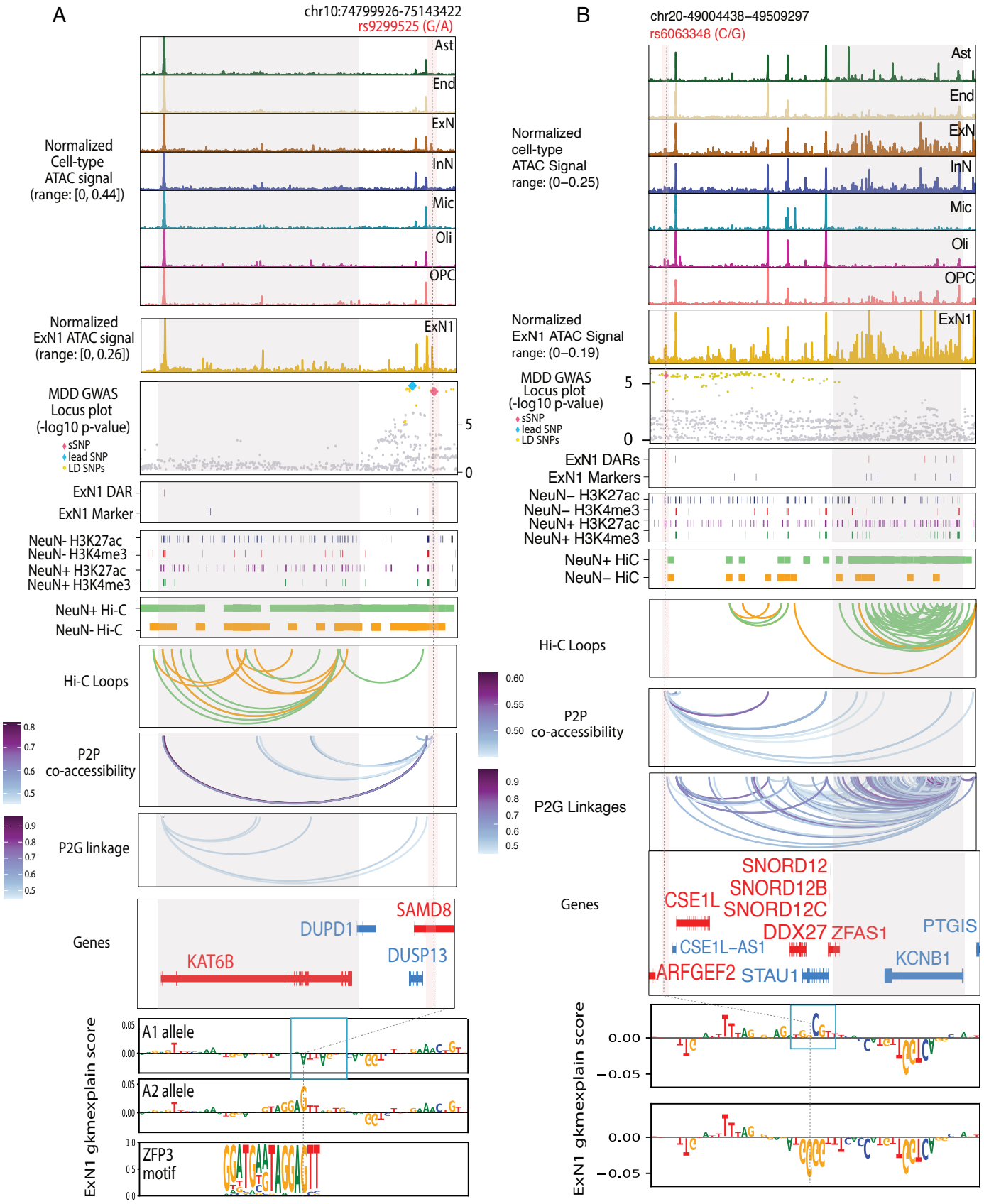


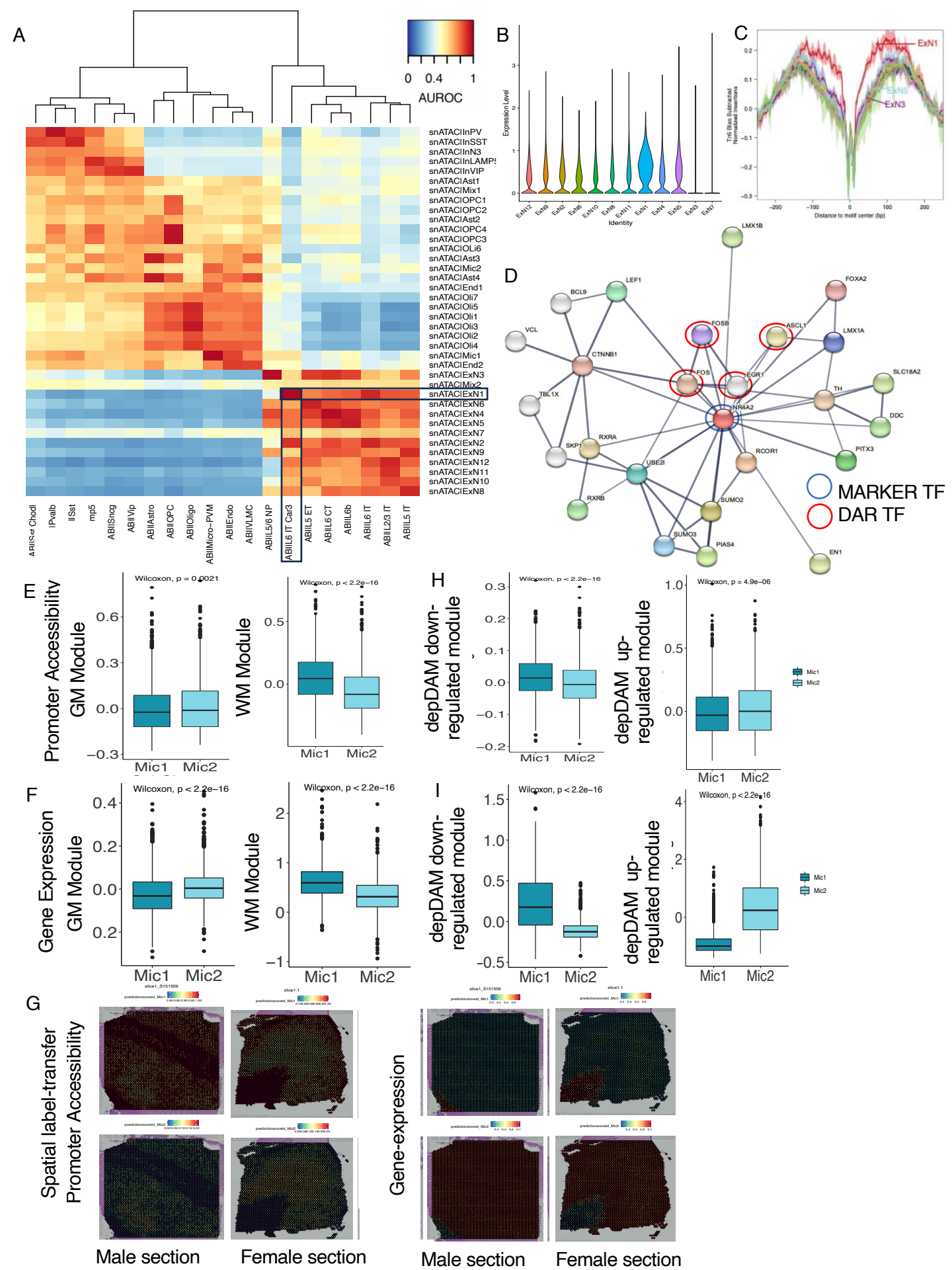
C ExN1 sSNPs GO



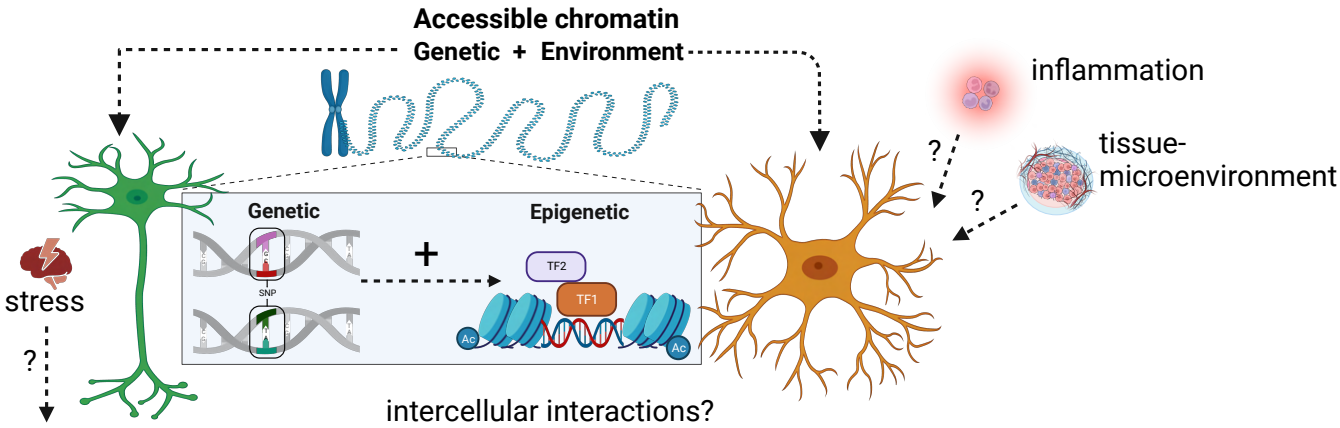
D Mic2 sSNPs GO




A
chr10:74799926-75143422
rs9299525 (G/A)
Normalized Cell-type ATAC signal (range: [0, 0.44])
Ast
End
ExN
InN
Mic
Oli
OPC
Normalized ExN1 ATAC signal (range: [0, 0.26])
ExN1
MDD GWAS Locus plot (-log10 p-value)
sSNP
lead SNP
LD SNPs
ExN1 DAR
ExN1 Marker
NeuN- H3K27ac
NeuN- H3K4me3
NeuN+ H3K27ac
NeuN+ H3K4me3
NeuN+ Hi-C
NeuN- Hi-C
Hi-C Loops
P2P co-accessibility
P2G linkage
Genes
SAMD8
DUPD1
KAT6B
DUSP13
ExN1 gkmexplain score
A1 allele
A2 allele
ZFP3 motif
B
chr20-49004438–49509297
rs6063348 (C/G)
Normalized cell-type ATAC Signal range: (0–0.25)
Normalized ExN1 ATAC Signal range: (0–0.19)
ExN1 DARs
ExN1 Markers
NeuN– H3K27ac
NeuN– H3K4me3
NeuN+ H3K27ac
NeuN+ H3K4me3
NeuN+ HiC
NeuN– HiC
SNORD12
SNORD12B
SNORD12C
CSE1L
DDX27
ZFAS1
PTGIS
CSE1L-AS1
KCNB1
ARFGEF2
STAU1


A
0
0.4
1
AUROC
snATACIInPV
snATACIInSST
snATACIInN3
snATACIInLAMP5
snATACIInVIP
snATACIAst1
snATACIMix1
snATACIOPC1
snATACIOPC2
snATACIAst2
snATACIOPC4
snATACIOPC3
snATACIOli6
snATACIAst3
snATACIMic2
snATACIAst4
snATACIEnd1
snATACIOli7
snATACIOli5
snATACIOli1
snATACIOli3
snATACIOli2
snATACIOli4
snATACIMic1
snATACIEnd2
snATACIExN3
snATACIMix2
snATACIExN1
snATACIExN6
snATACIExN4
snATACIExN5
snATACIExN7
snATACIExN2
snATACIExN9
snATACIExN12
snATACIExN11
snATACIExN10
snATACIExN8
ABIISst ChodI
IPvalb
IISst
mp5
ABIISncg
ABIIVip
ABIIAstro
ABIIOPC
ABIIOligo
ABIIMicro-PVM
ABIIEndo
ABIIVLMC
ABIIL5/6 NP
ABIIL6 IT Car3
ABIIL5 ET
ABIIL6 CT
ABIIL6b
ABIIL6 IT
ABIIL2/3 IT
ABIIL5 IT
B
Expression Level
Identity
ExN12
ExN9
ExN2
ExN6
ExN10
ExN8
ExN11
ExN1
ExN4
ExN5
ExN3
ExN7
C
ExN1
ExN5
ExN3
Tn5 Bias Subtracted Normalized Insertions
Distance to motif center (bp)
D
LMX1B
LEF1
BCL9
FOXA2
VCL
FOSB
ASCL1
LMX1A
CTNNB1
FOS
EGR1
SLC18A2
TBL1X
TH
NR4A2
DDC
RXRA
SKP1
RCOR1
PITX3
UBE2I
RXRB
SUMO2
SUMO3
PIAS4
EN1
MARKER TF
DAR TF
E
Promoter Accessibility
GM Module
Wilcoxon, p = 0.0021
WM Module
Wilcoxon, p < 2.2e−16
Mic1
Mic2
H
depDAM down-regulated module
Wilcoxon, p < 2.2e−16
depDAM up-regulated module
Wilcoxon, p = 4.9e−06
Mic1
Mic2
F
Gene Expression
GM Module
Wilcoxon, p < 2.2e−16
WM Module
Wilcoxon, p < 2.2e−16
Mic1
Mic2
I
depDAM down-regulated module
Wilcoxon, p < 2.2e−16
depDAM up-regulated module
Wilcoxon, p < 2.2e−16
Mic1
Mic2
G
Spatial label-transfer
Promoter Accessibility
Gene-expression
Male section
Female section
Male section
Female section

MDD



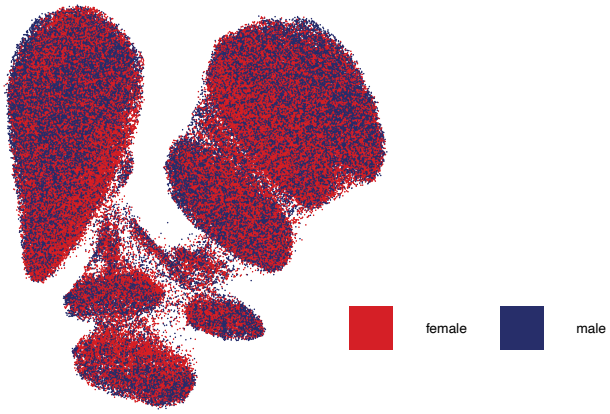
NR4A2+ deep-layer neurons

- MDD risk heritability in marker cCREs and MDD-associated differential chromatin
- Neurodevelopment (bHLH/HTH) & Activity-dependent TFs (bZIP)
- Genes involved in synaptic communication

GM microglia

- Differentially closed chromatin in MDD
- No significant enrichment for MDD heritability
- Microglial lineage- & immune-regulatory TFs (ETS domain/IRFs)
- Genes involved in phagocytosis and immune functions

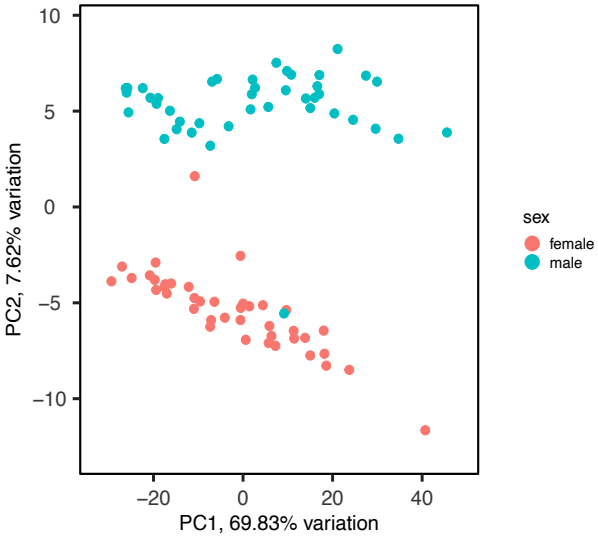
A



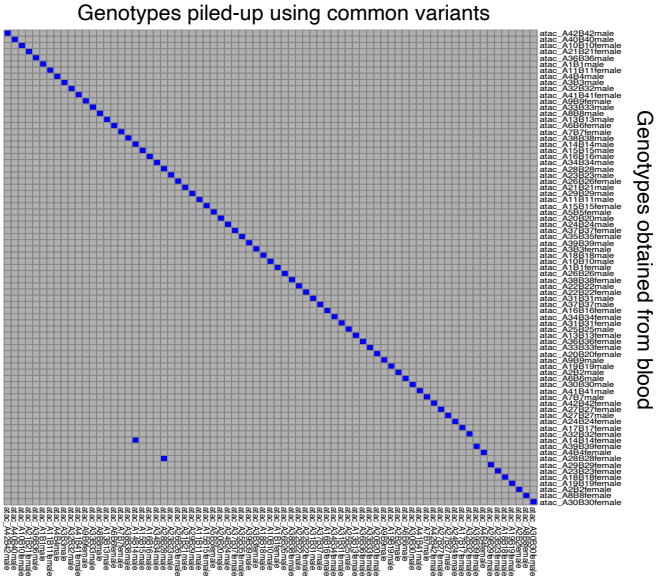
B

Variables	UMAP1	UMAP2
Clusters	0.89	0.68
Subjects	0.15	0.09
Sex	0.0007	0.003

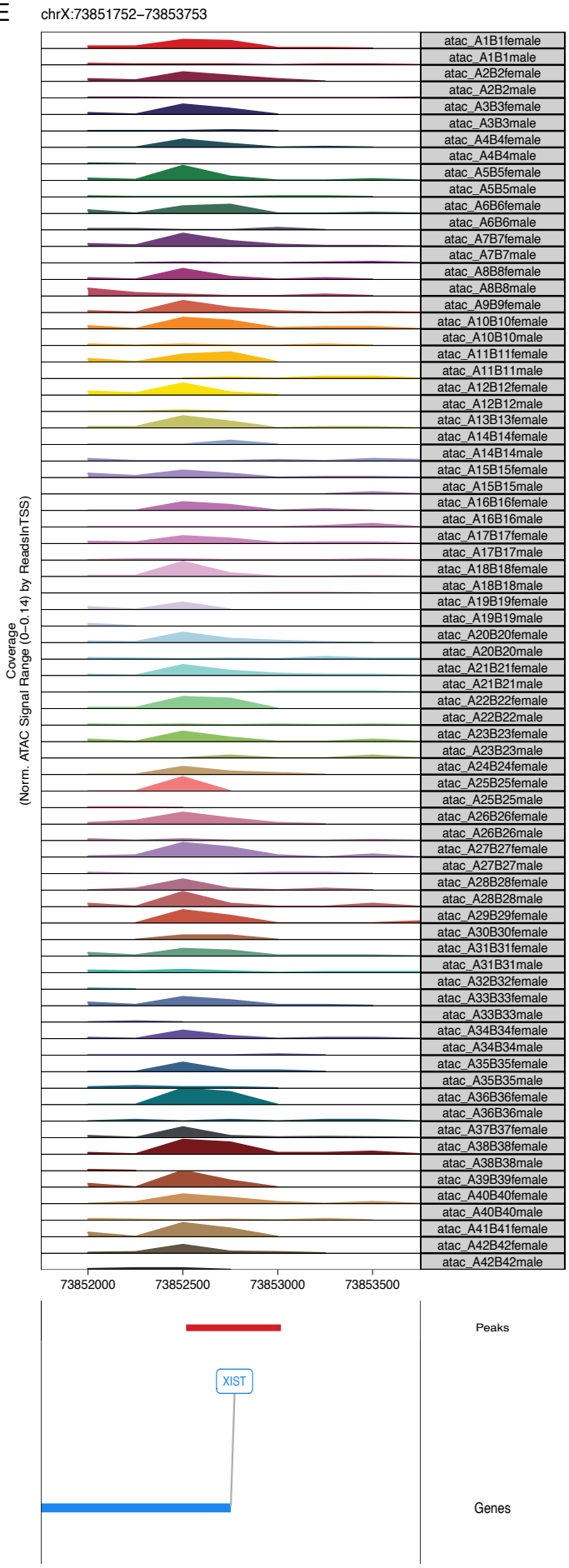
C

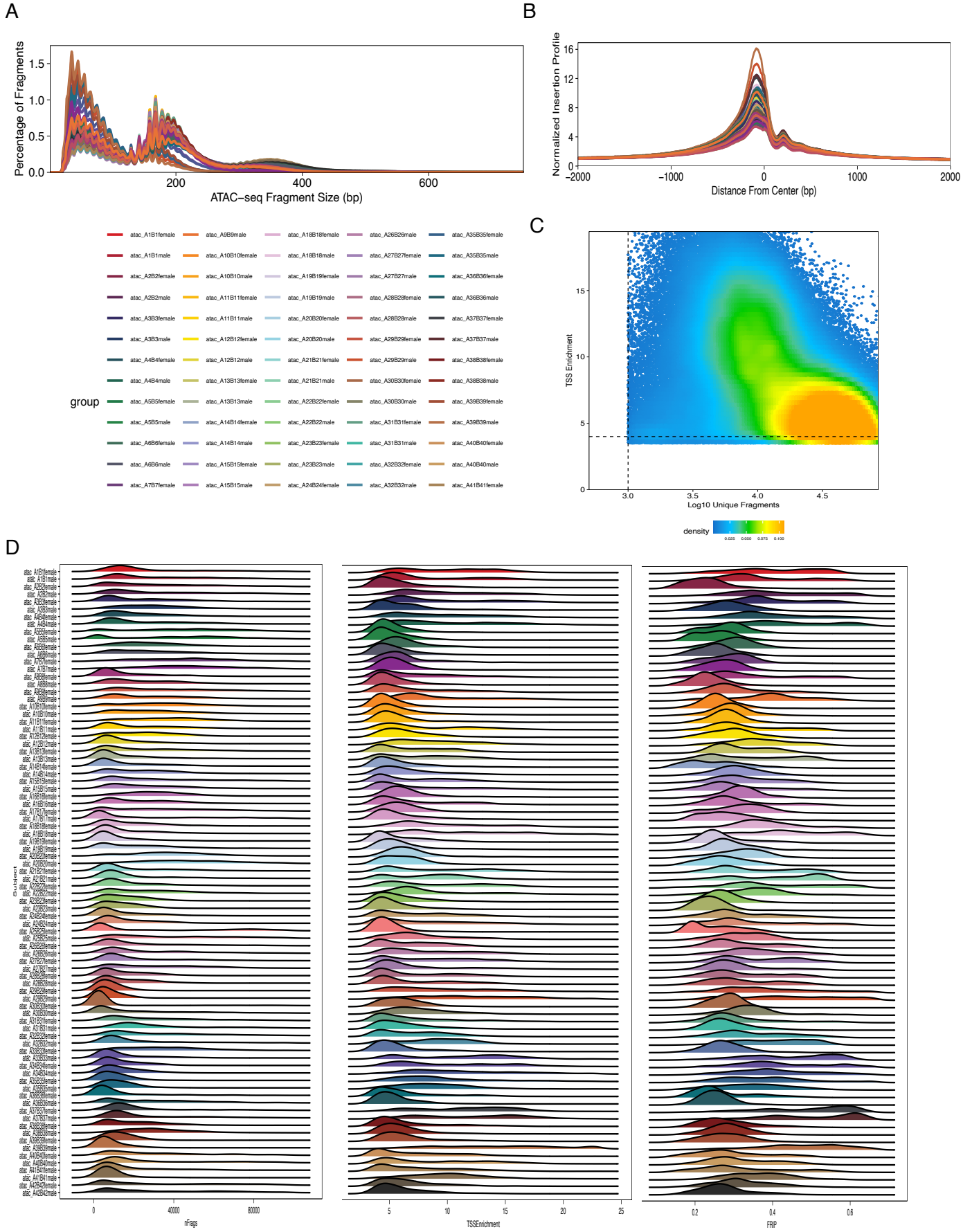


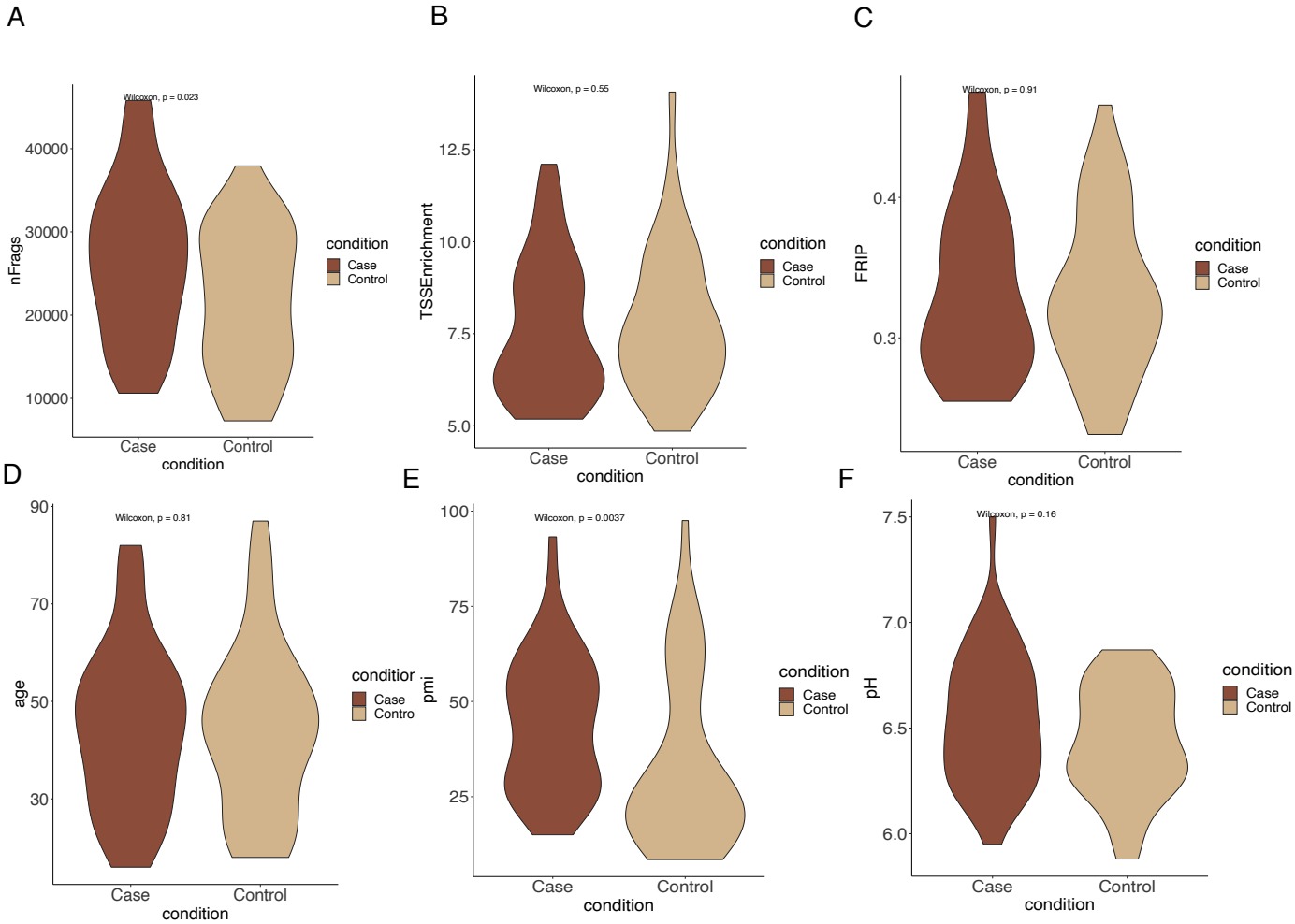
D

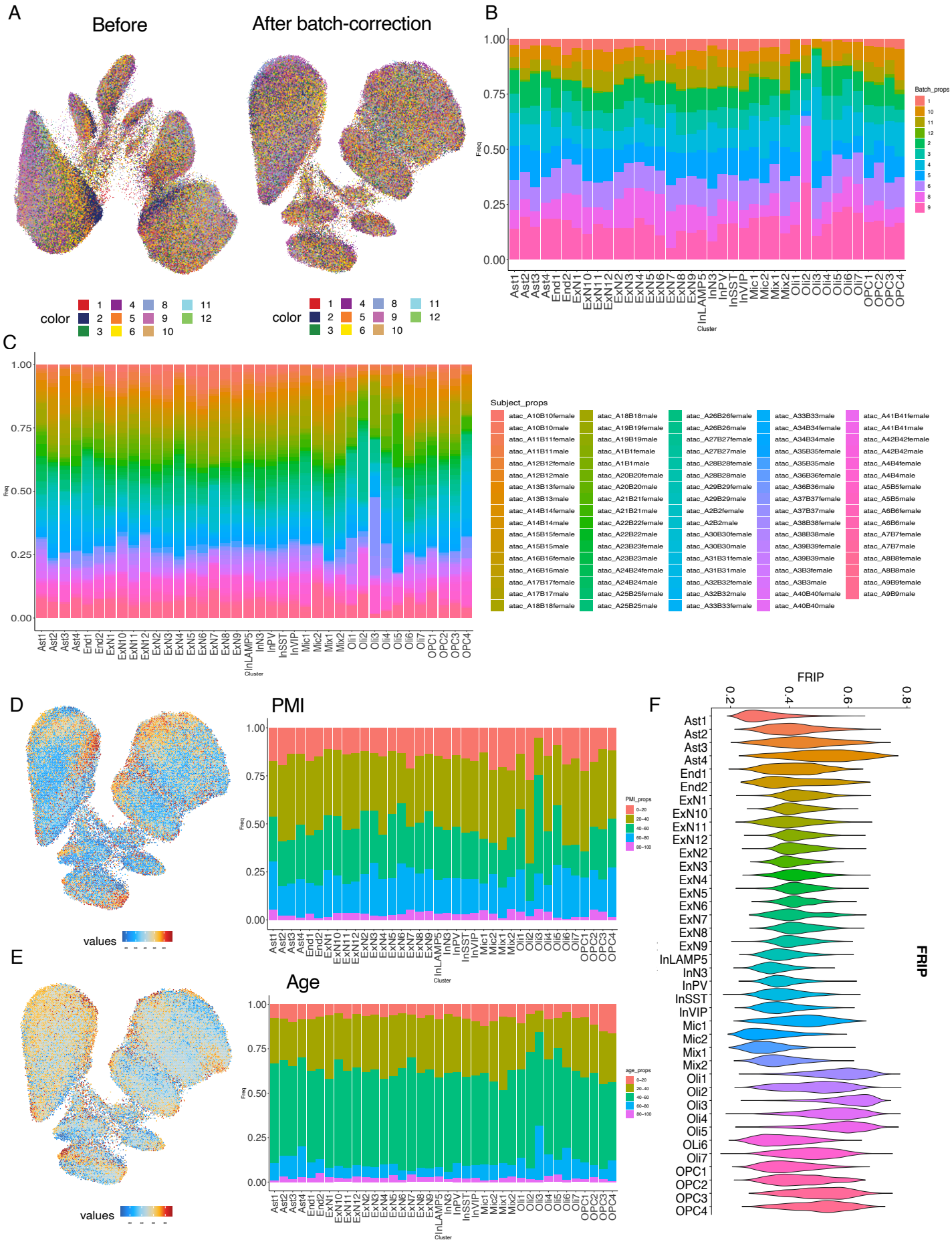


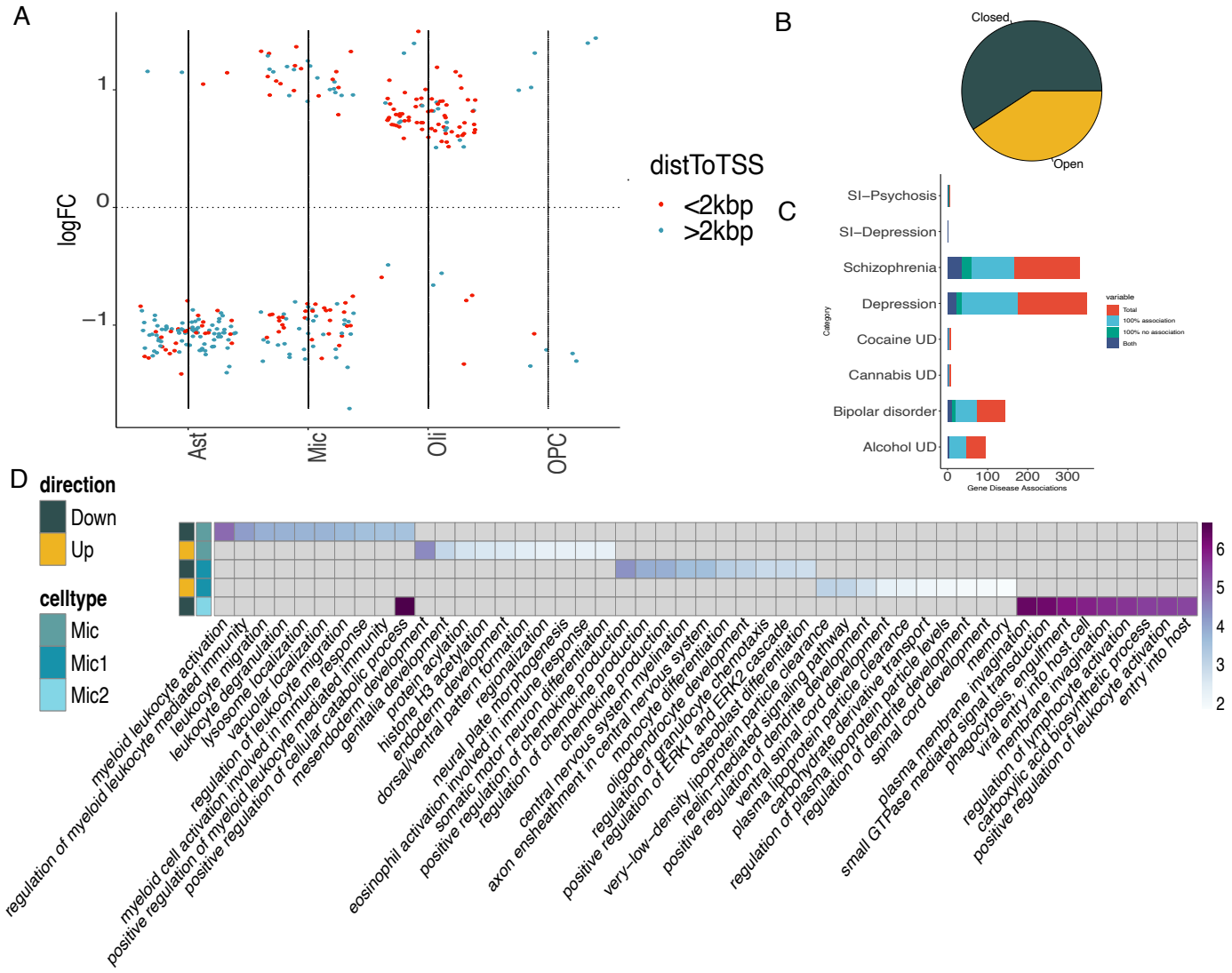
E




A
Percentage of Fragments
ATAC−seq Fragment Size (bp)
B
Normalized Insertion Profile
Distance From Center (bp)
group
C
TSS Enrichment
Log10 Unique Fragments
density
D
Subject
nFrags
TSSEnrichment
FRIP


A
Wilcoxon, p = 0.023
nFrags
40000
30000
20000
10000
Case
Control
condition
condition
Case
Control
B
Wilcoxon, p = 0.55
TSSEnrichment
12.5
10.0
7.5
5.0
Case
Control
condition
condition
Case
Control
C
Wilcoxon, p = 0.91
FRIP
0.4
0.3
Case
Control
condition
condition
Case
Control
D
Wilcoxon, p = 0.81
age
90
70
50
30
Case
Control
condition
condition
Case
Control
E
Wilcoxon, p = 0.0037
pmi
100
75
50
25
Case
Control
condition
condition
Case
Control
F
Wilcoxon, p = 0.16
pH
7.5
7.0
6.5
6.0
Case
Control
condition
condition
Case
Control


A
Before
After batch-correction
color
1 2 3 4 5 6 8 9 10 11 12
B
Freq
Cluster
Batch_props
1 10 11 12 2 3 4 5 6 8 9
C
Subject_props
atac_A10B10female
atac_A10B10male
atac_A11B11female
atac_A11B11male
atac_A12B12female
atac_A12B12male
atac_A13B13female
atac_A13B13male
atac_A14B14female
atac_A14B14male
atac_A15B15female
atac_A15B15male
atac_A16B16female
atac_A16B16male
atac_A17B17female
atac_A17B17male
atac_A18B18female
atac_A18B18male
atac_A19B19female
atac_A19B19male
atac_A1B1female
atac_A1B1male
atac_A20B20female
atac_A20B20male
atac_A21B21female
atac_A21B21male
atac_A22B22female
atac_A22B22male
atac_A23B23female
atac_A23B23male
atac_A24B24female
atac_A24B24male
atac_A25B25female
atac_A25B25male
atac_A26B26female
atac_A26B26male
atac_A27B27female
atac_A27B27male
atac_A28B28female
atac_A28B28male
atac_A29B29female
atac_A29B29male
atac_A2B2female
atac_A2B2male
atac_A30B30female
atac_A30B30male
atac_A31B31female
atac_A31B31male
atac_A32B32female
atac_A32B32male
atac_A33B33female
atac_A33B33male
atac_A34B34female
atac_A34B34male
atac_A35B35female
atac_A35B35male
atac_A36B36female
atac_A36B36male
atac_A37B37female
atac_A37B37male
atac_A38B38female
atac_A38B38male
atac_A39B39female
atac_A39B39male
atac_A3B3female
atac_A3B3male
atac_A40B40female
atac_A40B40male
atac_A41B41female
atac_A41B41male
atac_A42B42female
atac_A42B42male
atac_A4B4female
atac_A4B4male
atac_A5B5female
atac_A5B5male
atac_A6B6female
atac_A6B6male
atac_A7B7female
atac_A7B7male
atac_A8B8female
atac_A8B8male
atac_A9B9female
atac_A9B9male
D
values
PMI
PMI_props
0–20
20–40
40–60
60–80
80–100
E
values
Age
age_props
0–20
20–40
40–60
60–80
80–100
F
FRIP



E

Get transcription factor (TF) families enriched (p<0.05; Homer) in cluster-specific MDD DARs

Filter TFs present in ArchR peak-motif matrix and snRNA-seq integrated gene-expression matrix (Fig3A)

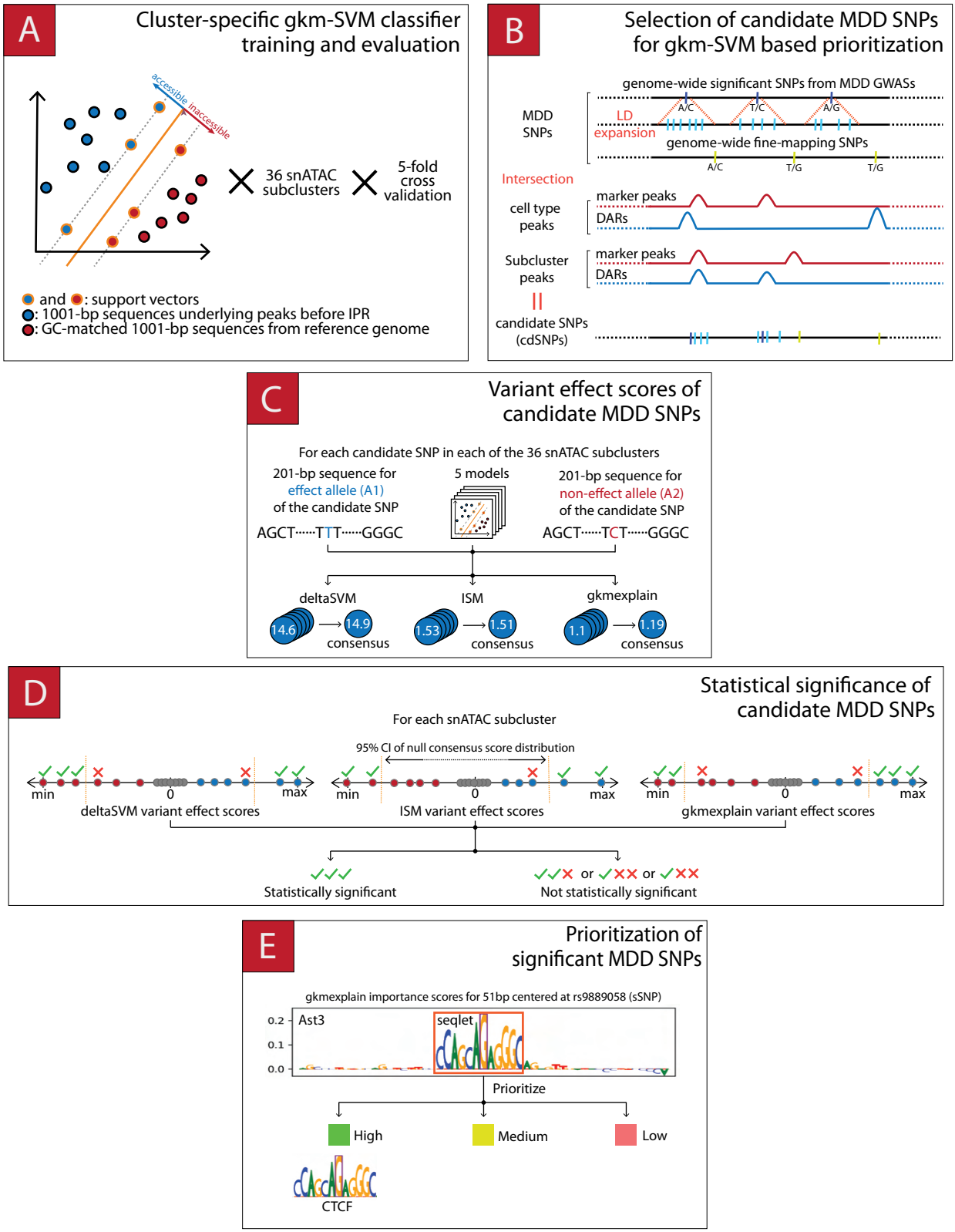
Keep TF families having (>2) distinct TFs enriched in cluster-specific DARs

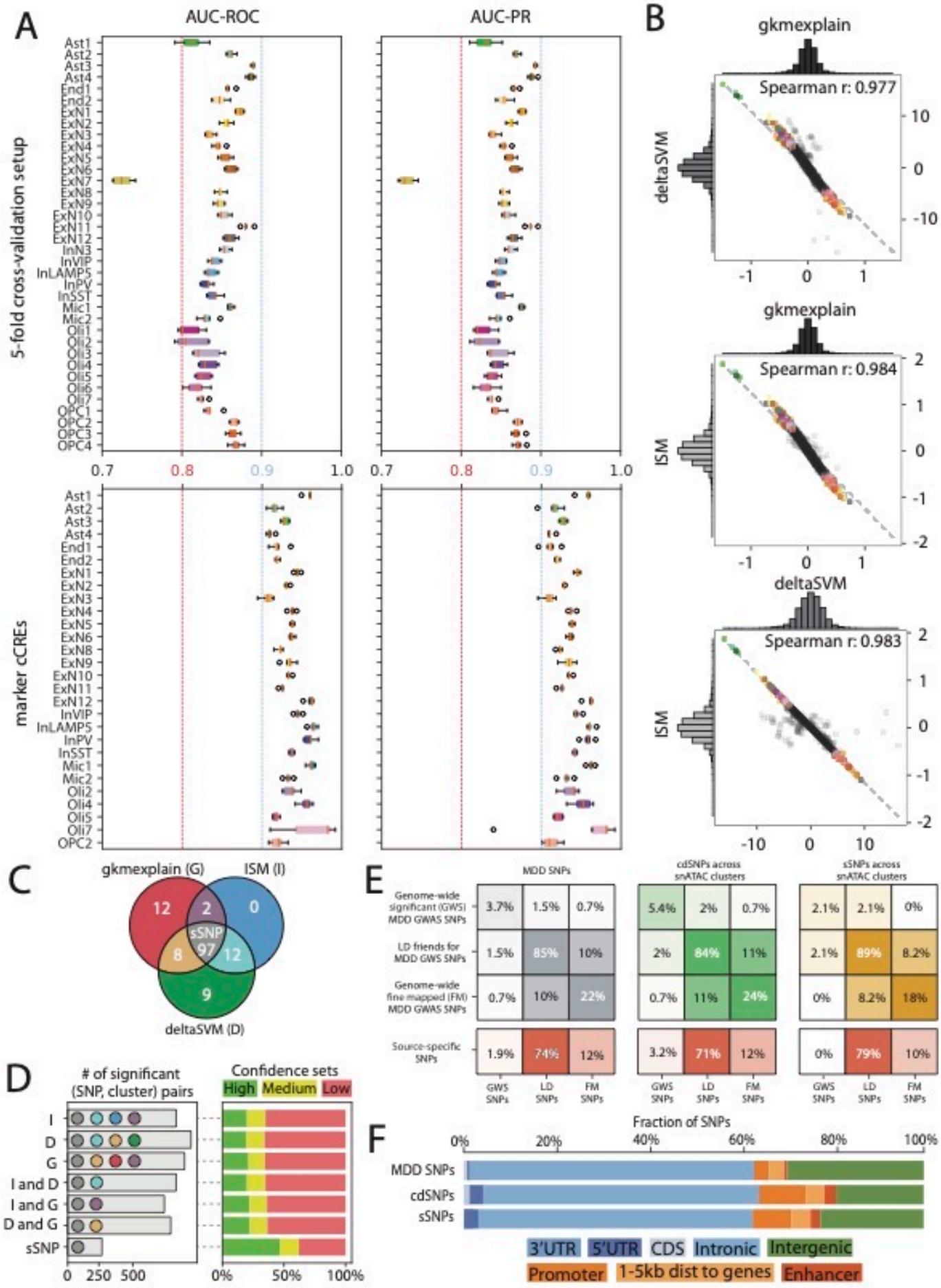
Rank TF motif families and TFs by enrichment in DARs (FDR<0.05)

Filter top two TFs with highest median gene-expression using the integrated snRNA-seq

Prioritize TFs with most significant differences (FDR<0.05; ChromVar) across motif-matching peaks in MDD versus controls (Fig3B)

Construct MDD-related cluster-specific TF target gene-regulatory networks (Fig3C)


A
Cluster-specific gkm-SVM classifier training and evaluation
accessible
inaccessible
36 snATAC subclusters
5-fold cross validation
and : support vectors
: 1001-bp sequences underlying peaks before IPR
: GC-matched 1001-bp sequences from reference genome
B
Selection of candidate MDD SNPs for gkm-SVM based prioritization
genome-wide significant SNPs from MDD GWASs
MDD SNPs
LD expansion
A/C
T/C
A/G
genome-wide fine-mapping SNPs
A/C
T/G
T/G
Intersection
cell type peaks
marker peaks
DARs
Subcluster peaks
marker peaks
DARs
candidate SNPs (cdSNPs)
C
Variant effect scores of candidate MDD SNPs
For each candidate SNP in each of the 36 snATAC subclusters
201-bp sequence for effect allele (A1) of the candidate SNP
AGCT·····TTT·····GGGC
5 models
201-bp sequence for non-effect allele (A2) of the candidate SNP
AGCT·····TCT·····GGGC
deltaSVM
ISM
gkmexplain
14.6
14.9
consensus
1.53
1.51
consensus
1.1
1.19
consensus
D
Statistical significance of candidate MDD SNPs
For each snATAC subcluster
95% CI of null consensus score distribution
min
0
max
deltaSVM variant effect scores
min
0
max
ISM variant effect scores
min
0
max
gkmexplain variant effect scores
Statistically significant
or or
Not statistically significant
E
Prioritization of significant MDD SNPs
gkmexplain importance scores for 51bp centered at rs9889058 (sSNP)
0.2
0.1
0.0
Ast3
seqlet
Prioritize
High
Medium
Low
CTCF


A
AUC-ROC
AUC-PR
5-fold cross-validation setup
marker cCREs
0.7
0.8
0.9
1.0
B
gkmexplain
Spearman r: 0.977
deltaSVM
gkmexplain
Spearman r: 0.984
ISM
deltaSVM
Spearman r: 0.983
ISM
C
gkmexplain (G)
ISM (I)
deltaSVM (D)
12
2
0
sSNP
97
8
12
9
D
# of significant (SNP, cluster) pairs
Confidence sets
High
Medium
Low
I
D
G
I and D
I and G
D and G
sSNP
0
250
500
0%
50%
100%
E
MDD SNPs
cdSNPs across snATAC clusters
sSNPs across snATAC clusters
Genome-wide significant (GWS) MDD GWAS SNPs
LD friends for MDD GWS SNPs
Genome-wide fine mapped (FM) MDD GWAS SNPs
Source-specific SNPs
GWS SNPs
LD SNPs
FM SNPs
F
Fraction of SNPs
0%
20%
40%
60%
80%
100%
MDD SNPs
cdSNPs
sSNPs
3'UTR
5'UTR
CDS
Intronic
Intergenic
Promoter
1-5kb dist to genes
Enhancer