

Prefrontal interneuron genes underlie neurobiological processes shared between psychiatric disorders

Giulio Pergola (✉ Giulio.Pergola@uniba.it)

University of Bari Aldo Moro <https://orcid.org/0000-0002-9193-1841>

Leonardo Sportelli

Corina Nagy

Daniel Weinberger

Director and CEO, Lieber Institute for Brain Development, Johns Hopkins School of Medicine:
Departments of Psychiatry, Neurology, Neuroscience and The Institute of Genetic Medicine

<https://orcid.org/0000-0003-2409-2969>

Gustavo Turecki

McGill University <https://orcid.org/0000-0003-4075-2736>

Article

Keywords: Single-cell RNA sequencing, Major Depressive Disorder, Schizophrenia, Tensor decomposition, Dorsolateral prefrontal cortex, Gene co-expression analysis

Posted Date: August 19th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1941910/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Previous reports on brain co-expression networks, mostly applied to tissue homogenates, have utilized cluster-based strategies for assigning genes to a unique co-expression cluster. This feature is incompatible with the diversity of cell populations and cellular components involved. Single-cell RNA-sequencing provides a finer-grained resolution than bulk tissue when interrogating brain cell types, dynamic states, and functional processes. We uncovered co-expression patterns across different brain cell types by applying tensor decomposition to single-nucleus transcriptomes from the prefrontal cortex of male patients with depression who died by suicide. We identified a gene set differentially co-expressed in inhibitory neurons between patients and controls and enriched for genes associated with major depression and schizophrenia.

Introduction

Gene co-expression is thought to be one mechanism governing the role of many putative risk associated genes, i.e., genes linked to variants associated with a disorder at genome-wide significant level, into biological pathways of risk for psychiatric disorders¹. Previous reports on brain co-expression networks have been primarily applied to tissue homogenates with diverse cell types. The high mRNA yield granted by tissue homogenate investigations comes at the price of biological specificity. In tissue homogenates, graph theory and clustering techniques have been used to assign genes expressed in multiple cell types to unique co-expression clusters². However, a hallmark of cell differentiation is that the same genes are expressed in different cell types together with different partners. Hence, different co-expression states in bulk tissue RNA-sequencing may reflect undefined diverse cell types whose relative abundance may vary between individuals³⁻⁴. Accordingly, the biological dissection of postmortem brain tissue into cell population enriched samples reveals genetic associations otherwise undetectable in bulk tissue RNA-sequencing⁵.

Recent single-nuclei RNA-sequencing technologies provide a window to disentangle the role of each cell type in complex psychiatric disorders⁶. For example, Nagy and colleagues⁷ examined single-nucleus transcriptomes from the prefrontal cortex of patients with major depressive disorder (MDD) and found a gene expression dysregulation in almost 60% of the cell types identified – with a potentially important interplay between oligodendrocyte precursor cells and deep layer excitatory neurons. They identified 96 cell-type-specific differentially expressed genes (DEGs) and performed so-called pseudo-bulk Weighted Genes Co-expression Network Analysis (WGCNA)⁸ on these data. WGCNA revealed a set of genes significantly enriched for the identified DEGs and involved in neurotransmitter secretion, suggesting a disruption of intercellular communication between neural cells. The pseudo-bulk approach aggregates the gene counts of each cell type at the sample level, essentially taking the sum of counts for each sample within each cell type. This approach helps to maximize signal from low-yield RNA-sequencing but takes limited advantage of the granularity afforded by the single-cell experiment. The use of WGCNA additionally imposes the constraint of assigning genes to non-overlapping modules.

A different approach to gene co-expression is Sparse Decomposition of Arrays (SDA)⁹. SDA relies on singular value decomposition developed in a Bayesian framework to decompose a 3D array (or tensor) of multiple-tissue or multiple-cell-types gene expression measurements into a set of latent components. The result is a list of gene sets with shared variance in their expression patterns, which can be tracked to separate cell types. Unlike WGCNA, SDA handles higher-dimensional data, representing the biological reality of different cell types expressing common genes which interact with each other in different ensembles depending on cell type. SDA allows gene loadings of each component to have a high sparsity level, i.e., it penalizes models with extensive sets of genes of negligible contribution, which may result from singular-value decomposition techniques.

Here, our objective was to leverage the richer representation offered by distinct cell types in the dataset reported by Nagy and colleagues⁷ using SDA to identify co-expression components across multiple cell types. We associated components with diagnosis and genomic risk, here operationalized as the ensemble of genes associated with the specific disorder by the latest Genome Wide Association Studies (GWASs) for various psychiatric disorders. This way we identified gene sets implicated in genetic risk and differential co-expression across prefrontal cell types. We further assessed the biological significance of these gene sets via over-representation analysis.

Materials And Methods

1. Single-nuclei RNA-seq data

Nagy and colleagues⁷ sequenced ~ 80,000 nuclear transcriptomes from 33 postmortem prefrontal cortex (Brodmann area 9) of MDD cases who died by suicide (N = 17) and neurotypical non-affected controls (NC; N = 16) and identified 26 cell clusters. They computed the average cell expression for each sample across every cluster.

2. Tensor decomposition applied to single-nuclei RNA-seq data

To maximize the number of genes expressed across all cell types, we removed 4 clusters (Endo, OPCs_2, Oligos_1, Oligos_2) and used RNA-sequencing data of 22 cell clusters, including 6,496 genes.

Log-transformed RPKM values with an offset of 1 were used for the analysis after performing the following pre-processing and normalization steps⁹: (i) removal of genes with < .1 expression median and > 30% zeros in all 22 cell types resulting in 6,477 genes, (ii) quantile normalization of expression data within each cell type, and (iii) rank-based transformation of each gene onto a standard normal.

We used a tensor of 6,477 mRNA expression measurements at total gene level across 22 cell types in 33 male individuals as input for the SDA tool. We obtained latent components defined by three bidimensional matrices: i) the individual score matrix, which represented the magnitude of the effect of each component across individuals and was used to compute the association with diagnosis; ii) the cell-type score matrix, which indicated the activity of the component for each cell-type and was used to identify the contribution of each cell-type to the components; iii) the gene loading matrix, which indicated

the contribution of each gene to components and served to identify genes specific to cell types or shared between them.

We computed SDA ten times to obtain a robust component membership of the genes analyzed⁹. We stored the individual, cell type, and gene loading scores, along with posterior inclusion probabilities (PIPs). Hierarchical clustering grouped components into clusters using a dissimilarity measure ($1 - |R_i|$; R_i = correlation between individual scores for all pairs of components across the ten runs). The clustering stopped when no correlations were above 0.6.

Finally, we combined components within each cluster, taking the mean of the individual scores, cell type scores, gene loadings, and the median of PIPs. To assign genes to a specific component, we set $PIP > .05$ as a parameter as previously done by authors of SDA⁹.

3. Biological and functional enrichment analysis of SDA components

To assess the association of the identified components with diagnosis and technical confounders, we performed a multiple linear regression using the component-specific individual score as dependent variable and diagnosis as independent variable including the number of genes and unique molecular identifiers (UMI) as covariates ($\alpha = .05$).

The association with genomic risk was assessed for components not associated with confounders through enrichment analysis for MDD¹⁰, SCZ¹¹, ASD¹², bipolar disorder (BD)¹³, suicide attempt (SA)¹⁴, obsessive-compulsive disorder (OCD)¹⁵, post-traumatic stress disorder (PTSD)¹⁶, and immune disorders (Crohn's disease: CD; Ulcerative colitis: UC)¹⁷. To this end, we used published lists of prioritized risk genes, i.e., putative causal genes identified by setting a fixed distance around each index GWAS-significant SNP and subsequently integrating genomic functions^{18,19} or chromatin interactions²⁰.

We further assessed over-representation for differentially expressed genes (DEGs) and genes proximal to differentially methylated CpG islands (DMGs) in the dorsolateral prefrontal cortex of patients with SCZ^{21,22} and patients with MDD^{7,23,24} and for loss of function intolerant genes (LoF)²⁵.

We explored the overlap between SDA and WGCNA outputs obtained from the same input data. We computed the correlation between SDA individual scores and WGCNA module eigengenes (the first principal component of gene expression in the modules, abbreviated as ME) and assessed the overrepresentation of SDA components for the five WGCNA modules that Nagy et al. found significantly associated with MDD⁷. In addition, we explored the correlation of SDA individual scores with each sample's respective composition in terms of the 22 single-cell type clusters they were composed of, expressed as a percentage, to evaluate potential effects of cell composition.

We further asked whether SDA components mapped onto specific brain cell types. To this end, we characterized components in terms of cell-type specificity with two methods: cell type loading analysis and marker gene overrepresentation following two reference atlases²⁶. Cell type loading represents the

covariance between the overall gene expression derived from one cell type and the component identified by SDA. Intuitively, this measure suggests that certain cell types contributed more to determine the inter-individual variation within one component. For the cell type loading analysis we scaled the cell-type score matrix so that the largest score was equal |1| and used a cut-off of |0.5|, as done by SDA developers⁹, to obtain a binary representation of scaled cell-type scores and to highlight the cell-type specificity of the components. For marker gene overrepresentation, we used gene sets reported by Skene and colleagues²⁶ including cell-type specificity indices available at http://www.hjerling-leffler-lab.org/data/scz_singlecell/. They computed specificity indices for each gene and range between 1 (high specificity for a given cell type) and 0 (low specificity). We used specificity indices derived from single-nuclei RNA-seq of human brains (Allen Institute for Brain Science, AIBS, unpublished; DroNc-seq in prefrontal cortex and hippocampus²⁷). We used Mean-rank Gene Set Test in the limma R package²⁸ to evaluate the enrichment of our components for the specificity indices of each cell type. We then computed a metanalytical pvalue for each component for each cell type considered in the two atlases.

Finally, the gene ontology of diagnosis-associated components was explored through the clusterProfiler R package²⁹.

Results

Of 23 identified components, eight were associated with technical confounders and discarded (Table 1; $p < .05$). Of the 15 clean components, four overrepresented MDD risk genes¹⁰ (permutation hypothesis test: $p[\text{empirical}] < .05$). Only one of them (C3; 1770 genes) also differed between MDD and NC ($t[29] = 2.9$; $p[\text{FDR}] = .024$). The same component was also enriched for schizophrenia (SCZ)¹¹ risk genes (permutation hypothesis test: $p[\text{empirical}] < .05$) as well as for genes differentially expressed and proximal to differentially methylated CpG islands in SCZ^{21,22}) and for loss of function intolerant genes (Fig. 1A). Interestingly, some components overrepresented genetic risk for MDD and other psychiatric disorders, including bipolar disorder and obsessive-compulsive disorder (C15, C22) although they were not at the same time differentially expressed between cases and controls like C3 (Fig. 1A). Moreover, some components (C2,C4,C6,C20) showed cell-type score specificity for the same excitatory neuronal types previously highlighted by Nagy et al.⁷ (Ex7, Fig. 1B), whereas others (C10,C11) were ascribed to oligodendrocyte and oligodendrocyte precursor cells (Fig. 1B). These results therefore converged with the previous analyses, whereas C3 represented a novel insight.

Table 1
Association with diagnosis and technical confounders. *Highlighted rows show components not associated with any technical confounder.*

	Diagnosis		Number of genes		Number of UMI	
	t-value	p-value	t-value	p-value	t-value	p-value
C1	0.311	0.758	3.31	0.00253	-1.11	0.276
C2	-0.389	0.7	0.899	0.376	0.0379	0.97
C3	2.95	0.00616	0.065	0.949	0.542	0.592
C4	0.0919	0.927	1.22	0.232	-0.936	0.357
C5	-0.476	0.637	-2.96	0.00607	-2.14	0.0406
C6	0.836	0.41	-0.479	0.636	-1.29	0.206
C7	-0.642	0.526	-2.77	0.00956	0.853	0.401
C8	-0.527	0.603	0.423	0.675	-1.18	0.249
C9	1.57	0.126	-0.823	0.417	0.57	0.573
C10	-0.0518	0.959	-0.48	0.635	-1.27	0.215
C11	0.48	0.635	0.57	0.573	-0.988	0.331
C12	0.956	0.347	0.835	0.41	-0.446	0.659
C13	-1.73	0.0943	-2.5	0.0185	0.878	0.387
C14	-0.093	0.927	-5.15	1.69e-05	2.14	0.0412
C15	-0.885	0.383	0.318	0.752	1.36	0.185
C16	0.839	0.408	-1.28	0.209	0.506	0.617
C17	-1.89	0.0681	-0.304	0.764	0.121	0.905
C18	0.361	0.721	-0.782	0.441	0.136	0.893
C19	1.07	0.296	2.3	0.029	-0.937	0.357
C20	0.235	0.816	-1.06	0.299	0.811	0.424
C21	0.261	0.796	-1.83	0.0781	2.42	0.0221
C22	-0.178	0.86	0.791	0.435	-0.152	0.88
C23	-2.3	0.0291	-3.15	0.00375	1.24	0.224

It is noteworthy that the two components ascribed to oligodendrocyte lineage (C10, C11) showed no relation to diagnosis nor MDD risk genes (Fig. 1A) while including genes shared in neuronal cell types, like most of the components identified (Fig. 1B, 1C). This could be related to the lower number of expressed genes in the oligodendrocyte clusters analyzed than the neuronal ones, meaning SDA weighted the contribution of neuronal clusters more than others or it might suggest that oligodendrocytes represent an epiphenomena or unclear origin. Moreover, one component was ascribed to Microglia cluster (C17, Fig. 1B), but the reference atlases did not support this specificity (Fig. 1C). Notably, C17 did not show significant association with any pathology examined, suggesting neuroinflammation might not play a crucial role in the shared etiopathogenesis of psychiatric disorders identified here. Cell type loadings ascribed C3 to parvalbumin-positive inhibitory neurons and astrocytes (Fig. 1B); the reference atlases highlighted neuronal specificity, hence providing convergent support to the specificity of this gene set for parvalbumin-positive inhibitory neurons (Fig. 1C). Gene ontologies for this component included neuron projection morphogenesis (135 genes, fold enrichment = 1.3, $p[\text{FDR}] = .045$), postsynaptic (146 genes, fold enrichment = 1.4, $p[\text{FDR}] = .003$) and somatodendritic compartments (174 genes, fold enrichment = 1.4, $p[\text{FDR}] = 5.7 \times 10^{-4}$) (Fig. 2A). Finally, the intersection between genes comprised in C3, SCZ, and MDD top risk loci identified two genes: *NEGR1*, a neuronal growth regulator whose variants have been previously associated with MDD¹⁰, and *RERE*, a SCZ risk gene¹¹ encoding a member of the atrophin family of arginine-glutamic acid (RE) dipeptide repeat-containing proteins that have been associated with cortical thickness reductions in patients with MDD³⁰ and whose expression was higher in patients with SCZ¹¹ (Fig. 2B).

The overlap analysis between SDA and WGCNA revealed that some components, including C3 (permutation hypothesis test: $p[\text{empirical}] < .05$), were significantly enriched for WGCNA modules (Fig. 1) but overall, there was a lack of correspondence with the ME correlation (Fig. 1), thus suggesting that genes within the components we found might have different co-expression patterns or biological pathway involvement than the ones identified by WGCNA. C3 results are not driven by cell type composition examined for each sample (no association survived correction for multiple comparisons).

Discussion

These findings complement the previous reports about gene expression changes in immature oligodendrocyte precursors and deep layer excitatory neurons⁷ by revealing a gene set of shared genetic risk between disorders that has not been identified with pseudo-bulk approaches. The gene set we identified is differentially co-expressed in male patients with MDD relative to NC and is associated with shared risk for MDD and SCZ, another disorder characterized by increased suicide rate relative to the general population. Previous studies provided evidence of widespread pleiotropy among psychiatric disorders with a moderate genetic correlation between MDD and SCZ and with pleiotropic genes involved in gene regulation, synaptic function, and neuronal and dendritic development³¹. Moreover, both MDD and SCZ risk genes are preferentially expressed in neurons^{11,31}. SDA uncovered interneuron gene co-expression that might play a role in local prefrontal neuronal circuits implicated in multiple psychiatric

disorders. This gene expression analysis thus points to a molecular pathway within inhibitory neurons that may be relevant for treatment across diagnoses.

A limitation of this analysis is that postmortem data were derived from male individuals only all of whom died by suicide, hence restricting the scope of our findings. It is also noteworthy that we do not have access to the methods of suicide adopted, thus preventing us from studying any role these can play in cell-specific gene co-expression. Because of the study design, the molecular processes leading to death may have differed between the patient and the control groups. Finally, we analyzed 6,477 genes expressed in nuclei from 22 cell types without information about cytoplasmatic gene expression.

Declarations

Acknowledgments and Disclosures

This work was supported by the Lieber Institute for Brain Development.

The single-nuclei RNA-seq data analyzed here has been previously published by Nagy et al. in 2020⁷.

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

References

1. Gandal MJ, Haney JR, Parikshak NN, Leppa V, Ramaswami G, Hartl C *et al*. Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science* 2018; **359**(6376): 693–697.
2. Pergola G, Di Carlo P, Jaffe AE, Papalino M, Chen Q, Hyde TM *et al*. Prefrontal Coexpression of Schizophrenia Risk Genes Is Associated With Treatment Response in Patients. *Biol Psychiatry* 2019; **86**(1): 45–55.
3. Hartl CL, Ramaswami G, Pembroke WG, Muller S, Pintacuda G, Saha A *et al*. Coexpression network architecture reveals the brain-wide and multiregional basis of disease susceptibility. *Nature neuroscience* 2021; **24**(9): 1313–1323.
4. Panagiotakos G, Pasca SP. A matter of space and time: Emerging roles of disease-associated proteins in neural development. *Neuron* 2022; **110**(2): 195–208.
5. Jaffe AE, Hoeppner DJ, Saito T, Blanpain L, Ukaigwe J, Burke EE *et al*. Profiling gene expression in the human dentate gyrus granule cell layer reveals insights into schizophrenia and its genetic risk. *Nature neuroscience* 2020; **23**(4): 510–519.
6. Mancarci BO, Toker L, Tripathy SJ, Li B, Rocco B, Sibille E *et al*. Cross-Laboratory Analysis of Brain Cell Type Transcriptomes with Applications to Interpretation of Bulk Tissue Data. *eNeuro* 2017; **4**(6).
7. Nagy C, Maitra M, Tanti A, Suderman M, Theroux JF, Davoli MA *et al*. Single-nucleus transcriptomics of the prefrontal cortex in major depressive disorder implicates oligodendrocyte precursor cells and excitatory neurons. *Nat Neurosci* 2020; **23**(6): 771–781.

8. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Statistical applications in genetics and molecular biology* 2005; **4**: Article17.
9. Hore V, Vinuela A, Buil A, Knight J, McCarthy MI, Small K *et al*. Tensor decomposition for multiple-tissue gene expression experiments. *Nat Genet* 2016; **48**(9): 1094–1100.
10. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A *et al*. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet* 2018; **50**(5): 668–681.
11. Schizophrenia Working Group of the Psychiatric Genomics Consortium, Ripke S, Walters JT, O'Donovan MC. Mapping genomic loci prioritises genes and implicates synaptic biology in schizophrenia. *medRxiv* 2020.
12. Grove J, Ripke S, Als TD, Mattheisen M, Walters RK, Won H *et al*. Identification of common genetic risk variants for autism spectrum disorder. *Nat Genet* 2019; **51**(3): 431–444.
13. Stahl EA, Breen G, Forstner AJ, McQuillin A, Ripke S, Trubetskoy V *et al*. Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat Genet* 2019; **51**(5): 793–803.
14. Erlangsen A, Appadurai V, Wang Y, Turecki G, Mors O, Werge T *et al*. Genetics of suicide attempts in individuals with and without mental disorders: a population-based genome-wide association study. *Mol Psychiatry* 2020; **25**(10): 2410–2421.
15. International Obsessive Compulsive Disorder Foundation Genetics C, Studies OCD CGA. Revealing the complex genetic architecture of obsessive-compulsive disorder using meta-analysis. *Mol Psychiatry* 2018; **23**(5): 1181–1188.
16. Nievergelt CM, Maihofer AX, Klengel T, Atkinson EG, Chen CY, Choi KW *et al*. International meta-analysis of PTSD genome-wide association studies identifies sex- and ancestry-specific genetic risk loci. *Nat Commun* 2019; **10**(1): 4558.
17. Liu JZ, van Sommeren S, Huang H, Ng SC, Alberts R, Takahashi A *et al*. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nature Genetics* 2015; **47**(9): 979–986.
18. Thyme SB, Pieper LM, Li EH, Pandey S, Wang Y, Morris NS *et al*. Phenotypic Landscape of Schizophrenia-Associated Genes Defines Candidates and Their Shared Functions. *Cell* 2019; **177**(2): 478–491.e420.
19. Huo Y, Li S, Liu J, Li X, Luo X-J. Functional genomics reveal gene regulatory mechanisms underlying schizophrenia risk. *Nature Communications* 2019; **10**(1): 670.
20. Won H, de la Torre-Ubieta L, Stein JL, Parikshak NN, Huang J, Opland CK *et al*. Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature* 2016; **538**(7626): 523–527.
21. Fromer M, Roussos P, Sieberts SK, Johnson JS, Kavanagh DH, Perumal TM *et al*. Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nature neuroscience* 2016; **19**(11): 1442–1453.

22. Jaffe AE, Gao Y, Deep-Soboslay A, Tao R, Hyde TM, Weinberger DR *et al.* Mapping DNA methylation across development, genotype and schizophrenia in the human frontal cortex. *Nature neuroscience* 2016; **19**(1): 40–47.
23. Huls A, Robins C, Conneely KN, De Jager PL, Bennett DA, Epstein MP *et al.* Association between DNA methylation levels in brain tissue and late-life depression in community-based participants. *Translational psychiatry* 2020; **10**(1): 262.
24. Pantazatos SP, Huang YY, Rosoklija GB, Dwork AJ, Arango V, Mann JJ. Whole-transcriptome brain expression and exon-usage profiling in major depression and suicide: evidence for altered glial, endothelial and ATPase activity. *Mol Psychiatry* 2017; **22**(5): 760–773.
25. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016; **536**(7616): 285–291.
26. Skene NG, Bryois J, Bakken TE, Breen G, Crowley JJ, Gaspar HA *et al.* Genetic identification of brain cell types underlying schizophrenia. *Nat Genet* 2018; **50**(6): 825–833.
27. Habib N, Avraham-Davidi I, Basu A, Burks T, Shekhar K, Hofree M *et al.* Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat Methods* 2017; **14**(10): 955–958.
28. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015; **43**(7): e47.
29. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 2012; **16**(5): 284–287.
30. Katsuki A, Kakeda S, Watanabe K, Igata R, Otsuka Y, Kishi T *et al.* A single-nucleotide polymorphism influences brain morphology in drug-naive patients with major depressive disorder. *Neuropsychiatr Dis Treat* 2019; **15**: 2425–2432.
31. Sey NYA, Hu B, Mah W, Fauni H, McAfee JC, Rajarajan P *et al.* A computational tool (H-MAGMA) for improved prediction of brain-disorder risk genes by incorporating brain chromatin interaction profiles. *Nature neuroscience* 2020; **23**(4): 583–593.

Figures

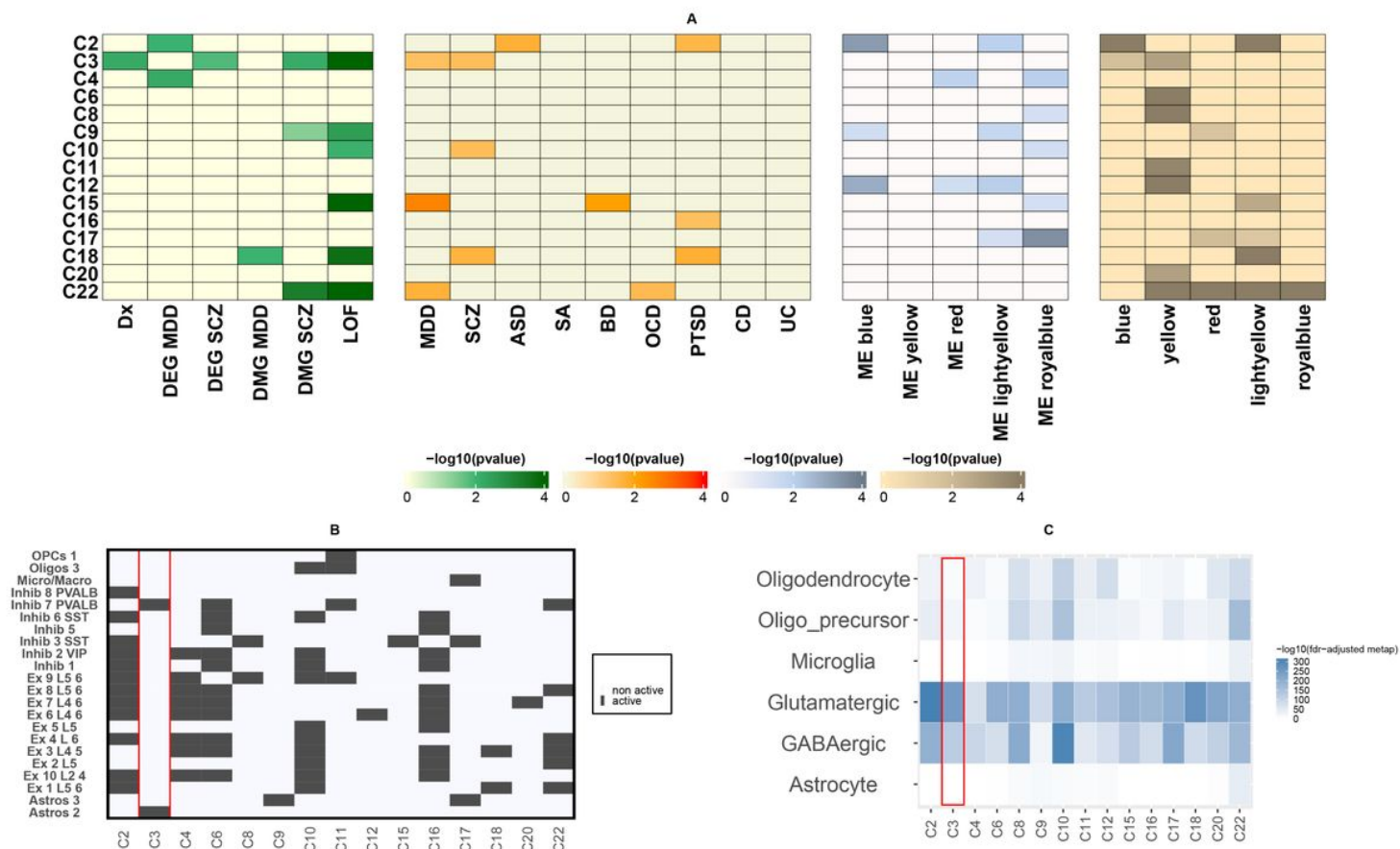


Figure 1

Biological and Functional enrichment analysis. **A** Heatmap of the functional characterization of the 15 components identified with SDA. Starting from left, first panel shows associations with diagnosis (Dx) and empirical enrichment pvalues for DEGs (DEG MDD, DEG SCZ), DMGs (DMG MDD, DMG SCZ) and LOF. Second panel refers to the empirical enrichment pvalues for the pathology-specific prioritized GWAS risk genes. Third and fourth panels refer to the Nagy et al. WGCNA ME correlation and module enrichment, respectively. C3 Component is the only one significantly associated with MDD diagnosis. **B** Binary representation of scaled cell-type scores (obtained by thresholding scores at 0.5) to highlight the cell-type specificity of the components. **C** Enrichment for human brain cell types. Cell types are displayed on the y-axis (snRNA-seq data from the Allen Institute for Brain Science and DroNc-seq). Tiles show the component fdr-corrected metanalytical p-values as gradient of color in the -log10 scale.

Abbreviations: ASD: autism spectrum disorder; BD: bipolar disorder; CD: Crohn's disease; Dx: diagnosis; DEG: differentially expressed genes; DMG: differentially methylated genes; LOF: loss of function intolerant genes; MDD: major depression disorder; ME: module eigengene; OCD: obsessive-compulsive disorder; PTSD: post-traumatic stress disorder; SA: suicide attempt; SCZ: schizophrenia; UC: ulcerative colitis.

Astros: astrocytes; Ex1-10: excitatory neurons; Inhib1-8: inhibitory neurons; L2-6: cortical layers; Micro/Macro: micro/macrogia; Oligos: oligodendrocytes; OPCs: oligodendrocyte precursor cells; PVALB:

parvalbumin positive neurons; SST: somatostatin-expressing neurons; VIP: vasoactive intestinal peptide-expressing neurons.

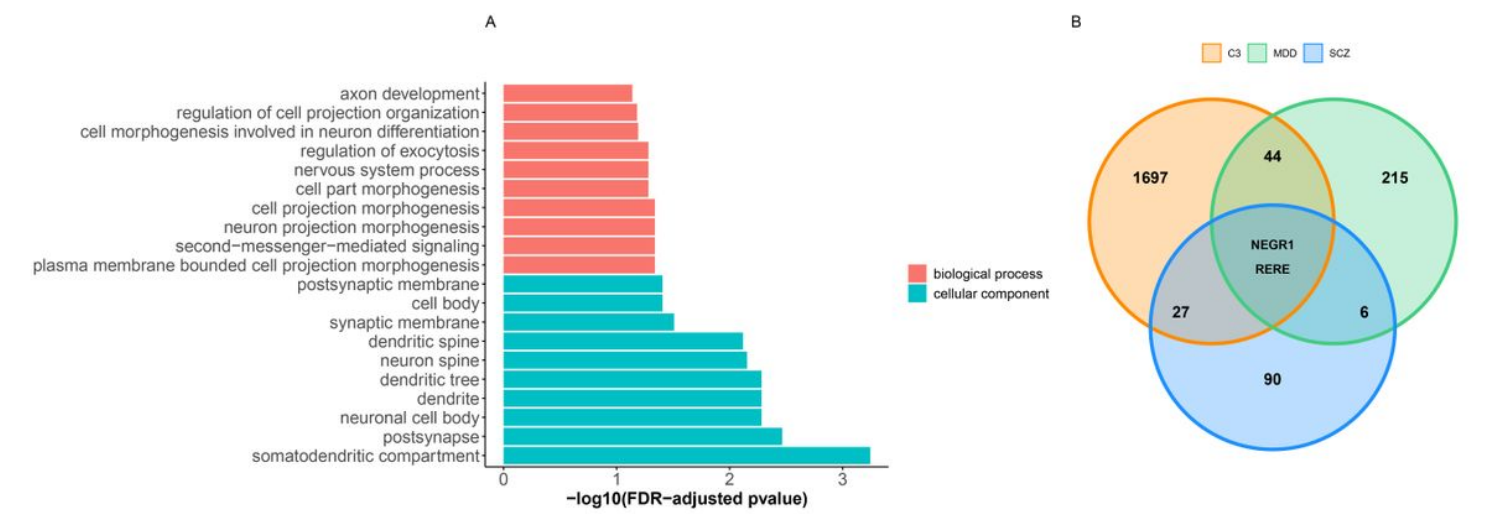


Figure 2

C3 Gene Ontology and intersection with MDD and SCZ risk genes. **A** Gene Ontology bar plot showing biological processes and cellular components for which C3 is significantly enriched. **B** Venn diagram showing the intersection between C3 genes and SCZ and MDD top risk loci genes.