



Single-nucleus transcriptomics of the prefrontal cortex in major depressive disorder implicates oligodendrocyte precursor cells and excitatory neurons

Corina Nagy^{1,9}, Malosree Maitra^{1,9}, Arnaud Tanti¹, Matthew Suderman², Jean-Francois Théroux¹, Maria Antonietta Davoli¹, Kelly Perlman¹, Volodymyr Yerko¹, Yu Chang Wang^{3,4}, Shreejoy J. Tripathy^{5,6,8}, Paul Pavlidis⁵, Naguib Mechawar^{1,7}, Jiannis Ragoussis^{1,3,4} and Gustavo Turecki^{1,3,7}

Major depressive disorder (MDD) has an enormous impact on global disease burden, affecting millions of people worldwide and ranking as a leading cause of disability for almost three decades. Past molecular studies of MDD employed bulk homogenates of postmortem brain tissue, which obscures gene expression changes within individual cell types. Here we used single-nucleus transcriptomics to examine ~80,000 nuclei from the dorsolateral prefrontal cortex of male individuals with MDD ($n=17$) and of healthy controls ($n=17$). We identified 26 cellular clusters, and over 60% of these showed differential gene expression between groups. We found that the greatest dysregulation occurred in deep layer excitatory neurons and immature oligodendrocyte precursor cells (OPCs), and these contributed almost half (47%) of all changes in gene expression. These results highlight the importance of dissecting cell-type-specific contributions to the disease and offer opportunities to identify new avenues of research and novel targets for treatment.

MDD is a complex and heterogeneous disorder that affects an estimated 300 million people worldwide¹. The genetic factors underlying the risk for MDD have been investigated using genome-wide association studies, among other approaches². Although some genetic associations have been detected, it remains a challenge to extract causal disease mechanisms from these findings³. It has been posited that MDD results from the dysregulation of monoaminergic transmission, thus largely implicating the serotonergic and noradrenergic systems, and this view has dominated the field for several decades. More recently, other factors have been associated with MDD, including glutamatergic and GABAergic transmission^{4,5}, glial cell function (including astrocytic and oligodendrocytic contributions)^{6–8}, blood-brain barrier integrity⁶ and inflammation⁹. Given the wide variety of cell types in the brain and their complex interactions, investigative approaches using cell-type specificity are especially needed to gain insight into psychiatric phenotypes, including MDD.

The interpretation of differential gene expression in bulk brain tissue homogenates is complicated by the heterogeneous cellular composition of the sample. Single-cell sequencing approaches have revealed that gene expression patterns in the brain are cell-type specific, not only differentiating major classes of cells such as neuronal and glial cells but also differentiating subtypes of glial cells and neurons^{10,11}. Therefore, it is difficult to verify whether subtle molecular differences observed from tissue homogenates are explained by the

disease state or by differences in the cell-type composition between samples¹². Recently developed techniques for high-throughput single-cell RNA sequencing (scRNA-seq) and single-nucleus RNA sequencing (snRNA-seq) provide a solution for addressing this inherent drawback of bulk-tissue experiments^{11,13}.

High-throughput droplet-based snRNA-seq allows the profiling of thousands of nuclear transcriptomes by utilizing nucleus-specific barcodes and unique molecular identifiers (UMIs) to tag individual RNA molecules. snRNA-seq yields comparable, albeit distinct, information¹⁴ from scRNA-seq while facilitating the analysis of frozen tissues, which are not amenable to the isolation of intact cells. While there has been considerable interest in using scRNA-seq and snRNA-seq datasets to gain insight into the processes underlying complex brain disorders^{15–17}, very few direct comparisons of single-nucleus human brain gene expression patterns have been performed in a psychiatric phenotype using high-throughput technologies.

Here, we sequenced ~80,000 nuclear transcriptomes from the prefrontal cortex of MDD cases and psychiatrically healthy controls and identified cell-type-specific differentially expressed genes (DEGs). These results point to gene expression changes in predominantly two cell types: OPCs and deep layer excitatory neurons. The relationships between and functions of the DEGs from these two cell clusters suggest impairments to fibroblast growth factor (FGF) signaling, steroid hormone receptor (SHR) cycling, immune function

¹McGill Group for Suicide Studies, Douglas Mental Health University Institute, Montreal, Quebec, Canada. ²MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK. ³Department of Human Genetics, McGill University, Montreal, Quebec, Canada. ⁴McGill University and Genome Quebec Innovation Centre, Montreal, Quebec, Canada. ⁵Michael Smith Laboratories and Department of Psychiatry, University of British Columbia, Vancouver, British Columbia, Canada. ⁶Krembil Centre for Neuroinformatics, Centre for Addiction and Mental Health, Toronto, Ontario, Canada. ⁷Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada. ⁸Department of Psychiatry, McGill University, Montreal, Quebec, Canada. ⁹These authors contributed equally: Corina Nagy, Malosree Maitra e-mail: gustavo.turecki@mcgill.ca

Table 1 | Sample information

	Controls (n=17)	Cases (n=17)	P value
Age (years) ^a	38.71±4.32	41.06±4.66	0.714
Gender	17 males	17 males	NA
PMI (h) ^a	34.01±4.94	41.69±4.76	0.190 ^b
pH	6.49±0.06	6.60±0.07	0.212
Storage time (years) ^a	14.71±1.44	12.47±1.46	0.543 ^b
Cause of death	Accident (6), natural (11)	Suicide (17)	
Substance dependence	None	None	
Comorbid diagnoses	None	None	
Toxicology	Ethanol (2), cannabinoids (1)	Ethanol (6), benzodiazepines (1), antidepressants (2), cannabinoids (1), cocaine (1)	
Antidepressant treatment	None	3	

^aData are presented as the mean±s.e.m. ^bMann-Whitney test. NA, not applicable.

and altered cytoskeletal regulation (related to changes in synaptic plasticity). This approach to snRNA-seq can effectively interrogate subtle phenotypes with improved resolution in archived brain tissue and provide novel directions for follow-up studies.

Results

To assess the involvement of individual cell types in the pathophysiology of MDD, we examined nuclei from the dorsolateral prefrontal cortex (dlPFC), a region implicated in the pathology of MDD¹⁸. We used a droplet-based single-nucleus method optimized for use with postmortem brain tissue to assess a large number of nuclei. We measured 78,886 nuclei from 34 brain samples: half from patients who died during an episode of MDD and the other half from matched psychiatrically healthy individuals (Table 1 and Supplementary Tables 1–3). The experimental design is depicted in Fig. 1. On average, we sequenced to a depth of almost 200 million reads per sample (Supplementary Table 1). Given that glial cells consistently have fewer transcripts than neuronal cells^{10,11}, we used custom filtering criteria based on the distribution of UMIs per nucleus detected to recover a substantial number of glial cells (Methods, Supplementary Fig. 1a–e and Supplementary Table 4). In an initial subset of 20 participants, applying our custom filtering criteria increased the total number of cells by 1.8-fold, but it also increased the number of non-neuronal cells by almost sixfold (data not shown). More than 90% of the nuclei passing these filtering criteria had less than 5% reads from mitochondrially encoded genes (Supplementary Fig. 1f). The average gene count across nuclei ranged from 2,144 in neurons to 1,144 genes in glia (Supplementary Table 5). UMI counts were approximately twice the gene count for all cell types, as expected for this level of sequencing depth (Supplementary Table 5). Between sample groups, there were no significant differences between cases and controls in the median gene count per nucleus (*t*-test, *P*=0.12), the median UMI count per nucleus (*t*-test, *P*=0.14) or the number of cells detected per individual (*t*-test, *P*=0.07) (Supplementary Table 1).

Identification of 26 distinct cell types in the dlPFC. To identify different cell types present in the brain samples, we applied unsupervised graph-based clustering¹⁹ using the first 50 principal

components (PCs) derived from the 2,135 most variable genes across individual nuclei (Methods and Supplementary Fig. 2a–b). After stringent quality control (Methods), we identified 26 distinct clusters (Fig. 2a). Each cluster was annotated using a combination of known cell-type markers for excitatory and inhibitory neurons and non-neuronal cells, including astrocytes, oligodendrocytes, OPCs, endothelial cells and microglia (see Methods for a full list of markers; Supplementary Table 6 and Supplementary Fig. 3a–p). Gene expression patterns specific to cell-type clusters were visualized using various methods, including a dot plot (Fig. 2b), average and median gene expression heatmaps (Supplementary Fig. 4a–b) and violin plots (Fig. 2c–e), to form a consensus for annotation.

Refined cell subtypes reflect cortical cellular architecture. The clusters generated from our data were consistent with those previously reported for a snRNA-seq analysis of the human prefrontal cortex¹¹ (Supplementary Fig. 5). Gene expression patterns previously linked to specific cortical layers (Methods) coincided with our clustering of excitatory cells. In Fig. 2c, the genes are arranged from left to right in order of their expression across the cortical layers (from layer I/II to layer VI). There was a gradient of expression of these genes across the excitatory clusters (designated ‘Ex’). For example, clusters Ex1, Ex4 and Ex7–9 had high expression of *TLE4* (specific for layer VI). Ex1, Ex8 and Ex9 showed concurrent expression of layer V/VI markers such as *TOX*. Ex6 and Ex7 additionally showed expression of the layer-IV-specific gene *RORB*. *HTR2C*, which is specific to a subset of layer V neurons, was prominent in Ex1 alone. *PCP4*, which is also specific to layer V, was present in Ex1–3, Ex7 and Ex9. Superficial layer (I–III) markers such as *CUX2* and *RASGRF2* were mainly seen in the large cluster Ex10. Likewise, inhibitory cell types (designated ‘Inhib’) demonstrated subtype-specific gene expression patterns. For example, Inhib_7 was classified as inhibitory parvalbumin (PVALB) because it expressed *GAD1* and *PVALB* and lacked *VIP* and *SST* (Fig. 2d). Multiple astrocytic clusters (designated ‘Astros’) were also identified, and while the typical subclassification of astrocytes is based on their morphology within gray or white matter²⁰, we used only gray matter for these samples. As such, based on the higher percentage of *GFAP* expression in Astros_3 (38%) compared to Astros_2 (21%), we suspect that Astros_3 is more likely to represent reactive astrocytes²¹ (Supplementary Table 6).

Reconstruction of oligodendrocyte developmental trajectory. We identified five distinct cell type clusters that fell into the oligodendrocyte lineage (OL), including two that we classified as OPCs (Fig. 2e). OPCs express a characteristic set of markers, such as *PDGFRA* and *PCDH15*, which decline as these cells mature into oligodendrocytes (designated as ‘Oligos’), whereas other lineage markers, such as *OLIG2* or *SOX10*, are present in both mature and immature cells. Given these developmental-stage-specific markers, it was possible to plot a pseudotime trajectory²² using gene expression for OPC1, OPC2, Oligos1, Oligos2 and Oligos3. Our results indicated that OPC2 were the youngest cells within the dataset followed by OPC1, then Oligos2 and Oligos3, with Oligos1 being the most mature (Fig. 2e, top). The expression of thousands of genes varied according to pseudotime (*q*<0.01). Approximately half of the genes associated with pseudotime overlapped in the cases and the controls (Supplementary Fig. 6a). However, among the genes exclusively associated with pseudotime in the cases, there was a 2.7-fold enrichment of apoptosis signaling according to the PANTHER²³ pathway analysis (false discovery rate (FDR) *P*=9.01×10⁻³), while no enrichment was observed in the controls. Given that certain stages of oligodendrocyte differentiation are associated with heightened susceptibility to apoptosis²⁴, this may indicate differences in OL development between cases and controls. To assess the individual profiles of important developmental gene markers, we plotted

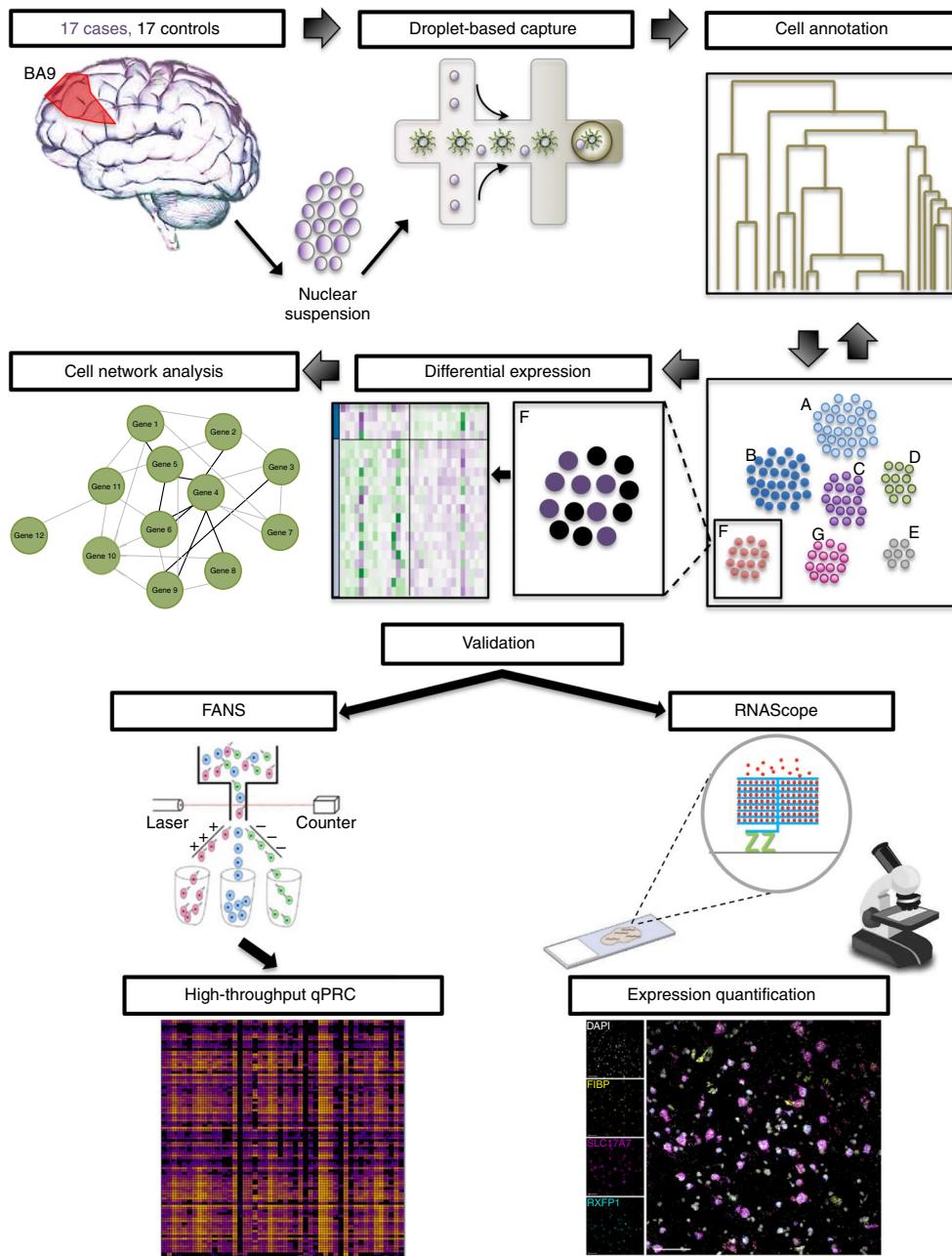


Fig. 1 | Experimental flow. Schematic representation of the experimental procedures. Nuclei were extracted from Brodmann area 9 (BA9) in the dlPFC of 17 cases and 17 controls. Single nuclei were then captured in droplets for RNA-seq. Unsupervised clustering and cell-type annotation were followed by differential expression analysis between the cases and the controls within each cluster. Bioinformatic analyses were performed to link the changes to the phenotype. Two validation approaches, FANS high-throughput qPCR and fluorescence in situ hybridization, were applied for validating the differential expression results.

their expression across pseudotime (Supplementary Fig. 6b–i); this revealed their expected pattern of expression.

To compare our OL cells with previously described OL cell types, we performed bioinformatic deconvolution (Fig. 2e, bottom). Our OPC2 gene expression profile was entirely represented by the ‘OPCs’ gene expression profile described in the Jäkel et al. study²⁵. The OPC1 profile also primarily corresponded to the OPCs; however, consistent with this cluster being further along the pseudotime trajectory, it showed a small correspondence to the committed OPCs. Our oligodendrocyte clusters showed varying degrees of correspondence to the published data, with a decreasing overlap to the published ‘OPCs’ expression profile with increasing maturity

of the cell type (ranging from 70% to 11% correspondence). Interestingly, among our oligodendrocytes, Oligos3 showed the highest correspondence to the ‘immune oligodendroglia’ as defined by Jäkel et al. study²⁵. The immune gene expression feature of Oligos3 is highlighted in our hierarchical clustering dendrogram (Fig. 2b), in which Oligos3 is located closer to the ‘Micro/Macro’ cluster compared with the other OL clusters.

Cell-type-specific patterns of altered gene expression in MDD. We set out to assess gene expression differences between the cases and the controls within each cluster. However, one limitation of droplet-based single-nucleus technology is the possibility of

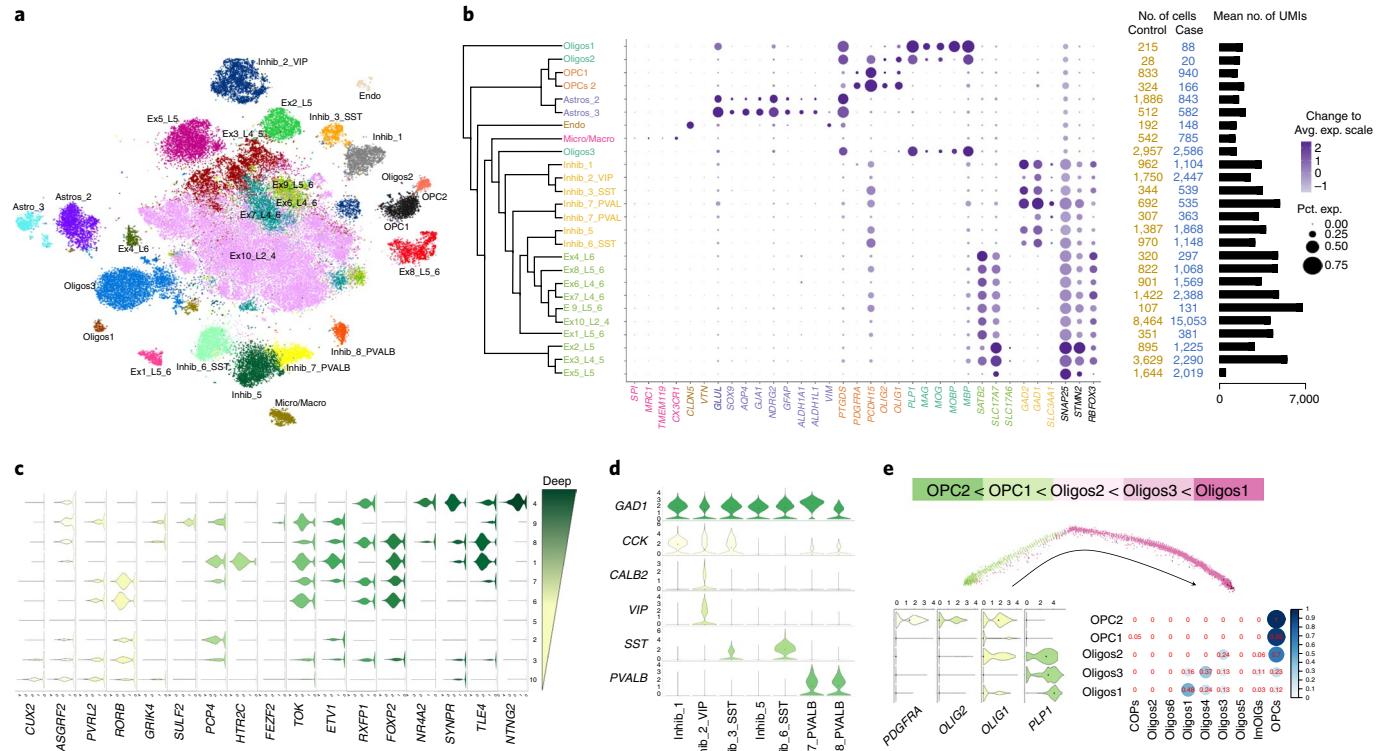


Fig. 2 | Identification of cell types. **a**, A t-distributed stochastic neighbor embedding plot depicting the ~73,000 cells in 26 clusters identified after strict quality control of initial clusters. Endo, endothelial; L, layer; Micro/Macro, microglia/macrophage. **b**, Cell-type annotation was performed on the basis of the expression of well-established marker genes. Left: dendrogram representing the relationship between identified cell-type clusters based on gene expression. Middle: dot plot depicting the expression of known marker genes in the 26 clusters of interest. Marker genes are color coded according to the cell type in which they should be detected. The size of the dots represents the proportion of cells expressing the gene (Pct. exp.), whereas the color intensity represents the average expression level (Avg. exp. scale). Right: the columns list the number of cells per group and the bar plot depicts the mean number of UMIs per cell in each cluster. **c**, Cortical-layer-specific markers varied in expression within the excitatory neuronal clusters. The violin plots depict the expression per cluster of layer-specific marker genes, from the more superficial layers (I/II) on the bottom to the deeper layers (V/VI) on the top. **d**, Known classes of inhibitory neurons were identifiable based on the expression pattern of peptide genes (VIP, SST and CCK) and calcium-binding protein genes (PVALB). **e**, Top: OL cells from five clusters were analyzed to produce a pseudotime trajectory to gauge their developmental stages. Bottom left: violin plots of cells belonging to the OL expressed the expected markers. Bottom right: the location of these clusters along the trajectory was consistent with deconvolution²⁵. The numbers represent the percentage contribution of each of the previously published²⁵ cluster signatures to the corresponding clusters in our dataset. COPs, committed OPCs; ImOIGs, immune oligodendroglia. For the violin plots in **c–e**, values extend from minimum to maximum, the median value is indicated by a dot, and the *n* value per cluster corresponds to the total ‘no. of cells’ for cases and controls combined listed in **b**. Nuclei were derived from 34 brains.

capturing doublet or multiplet nuclei, which we estimated was minimal in our case, as only 5.2% of captured nuclei were doublets or multiplets based on a species-mixing experiment (Supplementary Fig. 1g). This, however, represented a potential confounding factor when assessing differential gene expression between groups. We therefore eliminated doublets and multiplets from the dataset by calculating the correlation of each cell to the median expression value of its assigned cluster (Methods; Supplementary Fig. 7), and cells with low correlation were removed (Supplementary Table 7a,b). We also excluded any genes expressed in less than 10% of the cells in that cluster. We then performed a differential gene expression analysis using only these purified clusters and filtered genes (median 5,212 per cluster; Supplementary Tables 8–31).

A total of 96 genes (FDR < 0.10) were differentially expressed in 16 out of the 25 clusters analyzed (Fig. 3a), and 45 of those remained significant at FDR < 0.05 (12 out of 25 clusters). FDR correction considering all clusters together yielded 41 significant genes (FDR < 0.10) in 16 clusters (Supplementary Table 32). This provides further support that our statistical analyses are able to detect differences in gene expression between the groups. To retain a larger

set of genes to better capture functional enrichments within individual cell types, we considered all genes that passed FDR < 0.10, corrected per cluster. The majority, 83% (80 genes), were downregulated in line with findings from previous transcriptomics studies of MDD^{3,4}. Differential expression analysis treated each cell as a sample (Supplementary Fig. 8a–f), but per-participant contributions were visualized using heatmaps of average gene expression to assess biases in participant contributions. Patterns of gene expression averaged by participant reflected the expected differences between cases and controls (Supplementary Fig. 9a–p). Thirty-nine out of the 96 DEGs were found in excitatory cell clusters and 34 of these were downregulated (Fig. 3a, inset). Some neuronal clusters contained both upregulated and downregulated genes, but it was more common for affected neuronal clusters to contain only downregulated genes (8 out of 12, 67%). All but one inhibitory cluster showed altered gene expression, and non-neuronal clusters tended to have both upregulated and downregulated genes (Fig. 3b).

Of particular interest, two clusters—one composed of immature OPCs (OPC2) and one composed of deep layer excitatory neurons (Ex7)—accounted for almost half (47%) of the dysregulated genes

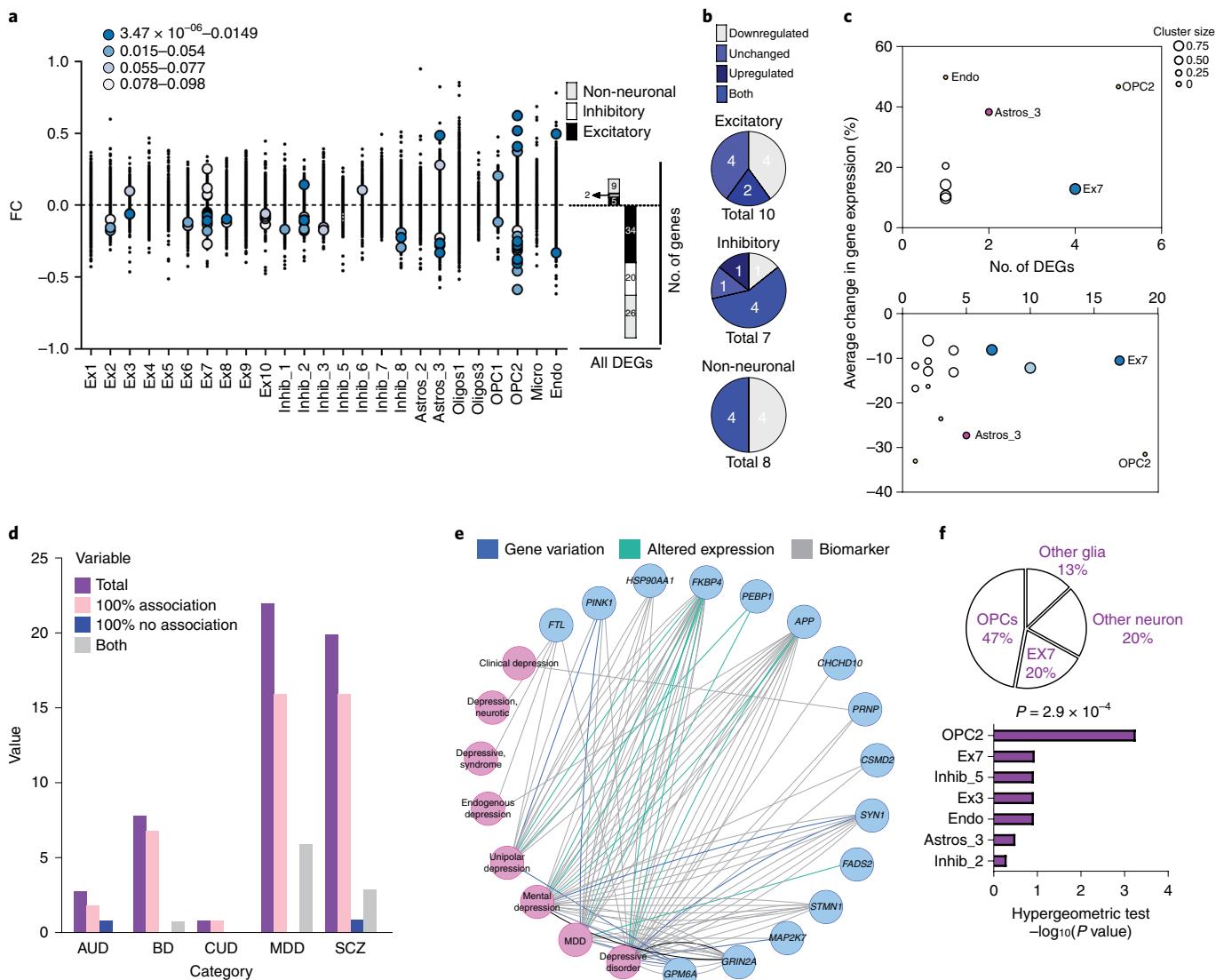


Fig. 3 | DEGs. **a**, For each cluster, the percentage change in expression between the cases and the controls of all detected genes are plotted such that decreased expression is toward the bottom of the midline and increased expression is toward the top. Ninety-six significantly changed genes (16 were upregulated and 80 were downregulated) are marked in color, based on their corrected FDRs as shown in the legend. The number of nuclei from the cases and the controls per cluster (n) are available in Supplementary Tables 8–31. P values were obtained using a mixed linear model (see Methods). Nuclei were derived from 34 individuals. Sixteen of the 26 clusters contained significantly DEGs. The inset shows a stacked bar graph of the contribution of different cell-type clusters to DEGs. **b**, The number of clusters in each broad category showing upregulated and downregulated genes in MDD cases. **c**, Scatter plots representing the number of DEGs and the average percentage change in expression for each cluster. The cluster size is depicted by the size of the circle. Colors indicate cell types. The upper graph depicts upregulated genes, while the lower graph depicts downregulated genes. Clusters OPC2 and Ex7 show the highest levels of both upregulated and downregulated genes. **d–f**, The number of genes with known relationships to psychiatric phenotypes, using the publicly available databases PsyGeNET and DisGeNET, were assessed. **d**, A total of 26 of the 96 dysregulated genes were found in PsyGeNET and showed an enrichment for MDD. ‘Total’ represents all the genes that overlapped the database for a given disorder, ‘100% association’ represents the genes positively associated with the disease, ‘100% no association’ represents the genes negatively associated with the disease, and ‘Both’ represents mixed findings (positive and negative) for a given gene related to the disease. AUD, alcohol use disorder; BD, bipolar disorder; CUD, cocaine use disorder; SCZ, schizophrenia. **e**, A total of 15 genes were associated with depression-related terms in DisGeNET. **f**, Top: the percentage of genes per cluster associated with MDD from DisGeNET, along with cluster-specific enrichment of DisGeNET MDD-associated genes. Bottom: for hypergeometric tests, the number of depression-associated genes in DisGeNET was 1,199 and the number of unique genes in DisGeNET was 17,545 for all tests. The number of DEGs in DisGeNET (k) and the number of depression-associated DEGs (x) are as follows: all clusters: $k=85$, $x=15$; OPC2: $k=24$, $x=7$; Ex7: $k=19$, $x=3$; Endo: $k=2$, $x=1$; Astros_3: $k=6$, $x=1$; Ex3: $k=2$, $x=1$; Inhib_2: $k=11$, $x=1$; Inhib_5: $k=2$, $x=1$.

(Fig. 3c). Finally, two genes were differentially expressed in more than one cluster: *PRKAR1B* showed decreased expression in excitatory clusters Ex7 ($FDR=0.087$, fold change (FC)=0.87) and Ex2 ($FDR=0.047$, FC=0.82), and *TUBB4B* in excitatory clusters Ex7 ($FDR=0.079$, FC=0.87) and Ex6 ($FDR=0.073$, FC=0.86).

Cell-type-specific DEGs recapitulate published MDD findings. Three of our DEGs (*FADS2*, *CKB* and *KAZN*) have previously been identified in genome-wide association studies of MDD^{2,26}. To further compare our DEGs with previously reported findings in MDD, we took advantage of the publicly available databases PsyGeNET²⁷

and DisGeNET²⁸. Using PsyGeNET, we found that 26 of our DEGs have been previously linked to mental illness in the literature. The highest number of associations (22 out of 54 associations) were for depressive disorders, followed by associations for schizophrenia spectrum and other psychotic disorders (20 out of 54; Fig. 3d). Using DisGeNET, we found 15 genes associated with MDD-related terms (hypergeometric test, $P=0.00029$; Fig. 3e). Hypergeometric tests for overlap between DEGs in individual clusters and genes related to depression in DisGeNET revealed a specific enrichment in OPC2 DEGs ($P=5.7 \times 10^{-4}$; Fig. 3e). Interestingly, we found that 67% of these genes were contributed by the OPC2 and Ex7 clusters (Fig. 3e). Complete results from PsyGeNET and DisGeNET are presented in Supplementary Tables 33–35.

Functional implications of cell-type-specific DEGs. We used Gene Ontology and Reactome Pathway enrichment analyses to identify the relationship of our 96 DEGs to biological functions. There were strong enrichments of Gene Ontology terms for ‘neuron projection maintenance’ (84-fold enrichment, FDR = 0.011) and ‘negative regulation of long-term synaptic potentiation’ (75-fold enrichment, FDR = 0.012). Both of these terms were hierarchically related with the more general term ‘regulation of synaptic plasticity’, also enriched in the set of 96 genes (9-fold enrichment, FDR = 0.012). Reactome Pathway enrichments included ‘kinesins’ (21.74-fold enrichment, FDR = 6.24×10^{-4}), ‘HSP90 chaperone cycle for steroid hormone receptors’ (15.79-fold enrichment, FDR = 3.4×10^{-2}) and ‘innate immune system’ (3.01-fold enrichment, FDR = 3.29×10^{-2}). A full list of all the enrichment analyses performed is provided in Supplementary Tables 36–41.

The majority (excluding AC133680.1, MEG3 and FAM66C) of the DEGs were protein-coding. We used STRING DB network analysis²⁹ to plot the interactions between these protein-coding DEGs. This enabled us to identify common pathways and systems, within which these proteins, contributed by different cell types, functionally interact. The overall connectivity between proteins encoded by our DEGs was significantly higher than that expected for a random subset of genes ($P=3.64 \times 10^{-4}$). While distinct genes were dysregulated in different clusters, common pathways and biological processes dysregulated across clusters included cytoskeletal function, immune system function and SHR chaperone cycling (Fig. 4a), all of which have been previously implicated in MDD^{9,30}.

Interestingly, certain genes were present in multiple pathways and processes; for example, HSP90AA1 (OPC2) links SHR chaperone cycling, immune system functioning and cytoskeletal function (Fig. 4b). Likewise, KIF16B from lower layer neurons (Ex7) and KIF26B and KLC2 in two inhibitory cell types (Inhib_2_VIP and Inhib_3_SST, respectively), belong to both the kinesin pathway and cytoskeletal function (Fig. 4c). Of note, KAZN, a gene previously associated with MDD²⁶, interacts with KIF16B (Ex7), both of which represent some of the few upregulated genes in the dataset.

Weighted gene co-expression network analysis. In addition to directly measuring gene expression changes between groups, we performed weighted gene co-expression network analysis (WGCNA). To circumvent the challenges posed by the sparseness of the snRNA-seq data, we performed WGCNA on the average gene expression profile for each participant across all cell types and included the percentage contribution of different cell types as a correlate. Our results indicated that five modules were significantly associated with MDD (Supplementary Table 42).

Four out of the five modules were also strongly associated with Ex7, representing the highest cluster–phenotype overlap. We chose to focus on the largest module (blue), which included 2,699 genes and significantly overlapped with our identified DEGs (Fig. 5a; 44%, $P=6.04 \times 10^{-19}$, hypergeometric test for overlap). To identify the most connected genes within the blue module, we performed a hub

gene analysis, which resulted in 285 hub genes (Fig. 5b and Methods), and plotted the top 50, which included 10 DEGs (Fig. 5c). The top term for a Gene Ontology analysis of the hub gene list was ‘neurotransmitter secretion’ (8.69-fold enrichment, FDR = 7.21×10^{-3}), which suggests that there is a disruption of intercellular communication between neural cells. Furthermore, we found that 26 out of the 41 DEGs that overlapped with the blue module were also hub genes ($P=4.95 \times 10^{-31}$, hypergeometric test for overlap).

Validation of gene expression changes. We performed validation of our DEGs using fluorescence-assisted nuclei sorting (FANS) to separate broad cell types followed by high-throughput quantitative PCR (qPCR). As expected, given that the FANS fractions are much broader than the single-cell clusters, with the 26 clusters combined into 4 sorted populations, levels of validation varied in part as a function of the relative representation of the cluster in the sorted fraction (Supplementary Figs. 10 and 11 and Supplementary Tables 43–46). Figure 5d highlights the validated genes that overlapped with the WGCNA results.

Intercommunication between lower layer excitatory neurons and OPCs. Next, to better understand how cells are interacting, we applied a predictive tool to explore the relationship of ligands of one cluster to the receptors expressed in another cluster. We focused our analysis on Ex7 and OPC2, the two clusters showing the most DEGs and with the greatest overlap of genes associated with phenotype from the literature and from our WGCNA. We found a total of 90 significantly changed ligand–receptor combinations between Ex7 and OPC2 after random permutations ($P<0.01$). Fifty-eight Ex7 ligand to OPC2 receptor (Fig. 6a, left; and Supplementary Table 47a) and 32 OPC2 ligand to Ex7 receptor interactions were altered between the cases and the controls (Fig. 6a, right; and Supplementary Table 47b). We found significant changes to FGF signaling originating from both cell types. Although these results are exploratory and need to be interpreted with caution, they are consistent with previous literature implicating the FGF system in MDD, and, in particular, changes in FGF signaling in OPCs^{31,32} leading to depressive phenotypes, and provide an intriguing avenue for future experiments.

Based on the DEGs found in Ex7 and OPC2, we modeled the potential interaction that indicated the class of protein and change in expression of the gene (Fig. 6b). To add support to the model, we used RNAscope fluorescence in situ hybridization to selected genes for further study. Given the important change in FGF signaling, we chose to investigate FIBP (which encodes FGF1 intercellular binding protein), KAZN (which encodes a potential junction protein) and HSP90AA1 (which encodes a co-chaperone involved in stress hormone receptor cycling). FIBP was downregulated, as expected, in deep layer excitatory neurons (Fig. 6c; unpaired *t*-test, $t_{217}=2.5$, $P=0.013$, $n=100$ nuclei for controls and 119 nuclei for cases), while KAZN was upregulated in OPCs (Fig. 6d; unpaired *t*-test, $t_{188}=-2.7$, $P=0.008$, $n=95$ nuclei each for controls and cases), and HSP90AA1 was also downregulated in OPCs (Fig. 6e; unpaired *t*-test, $t_{186}=2.2$, $P=0.027$, $n=102$ nuclei for controls, $n=86$ nuclei for cases).

Discussion

Our examination of single-nucleus transcriptomes from the dlPFC of patients with MDD revealed the dysregulation of gene expression in almost 60% of the cell types identified, with a total of 96 DEGs. There were prominent gene expression changes in immature OPCs (cluster OPC2) and in deep layer excitatory neurons (cluster Ex7), and a large percentage of their DEGs overlapped with genes previously implicated in MDD.

Given the complexity of psychiatric disorders such as MDD, disentangling the role of each cell type in the brain is important and requires single-cell resolution. For example, the ability to distinguish

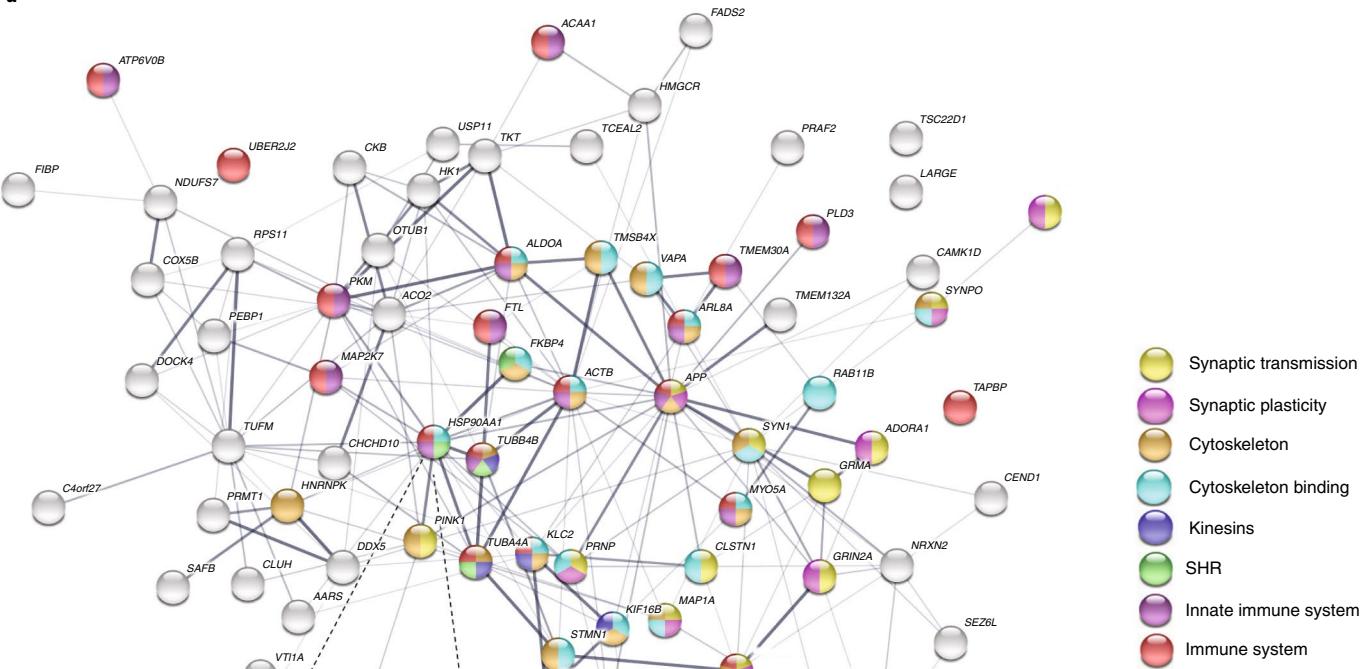
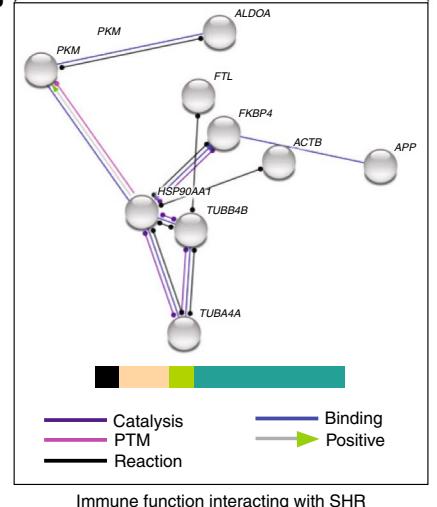
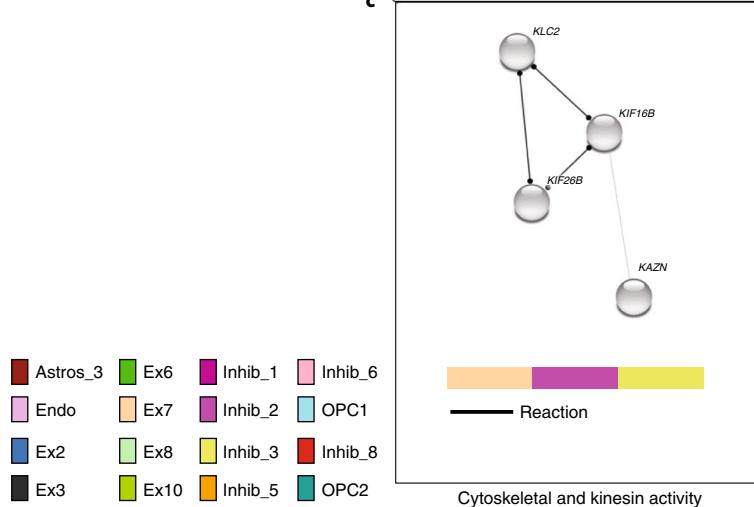
a**b****c**

Fig. 4 | Differential expression and biological associations. **a**, STRING DB network results for all DEGs, with nodes corresponding to a set of biological processes and pathways highlighted (legend on the right). **b**, Subset of genes shared between the immune-function-related terms and the SHR cycling pathway. **c**, Subset of genes involved in cytoskeletal function and kinesin activity. The color bars beneath the networks provide a proportional representation of the contributing clusters.

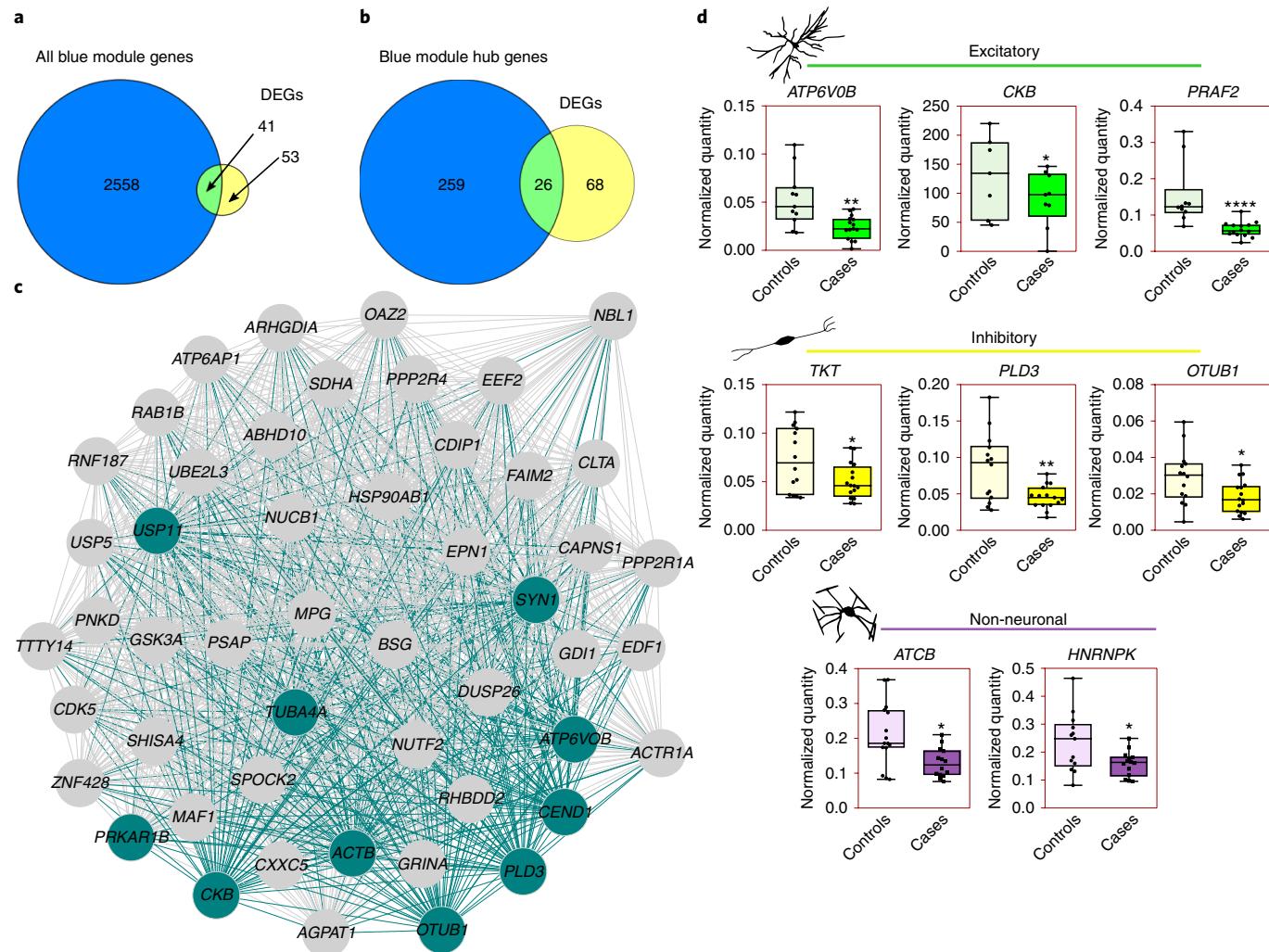


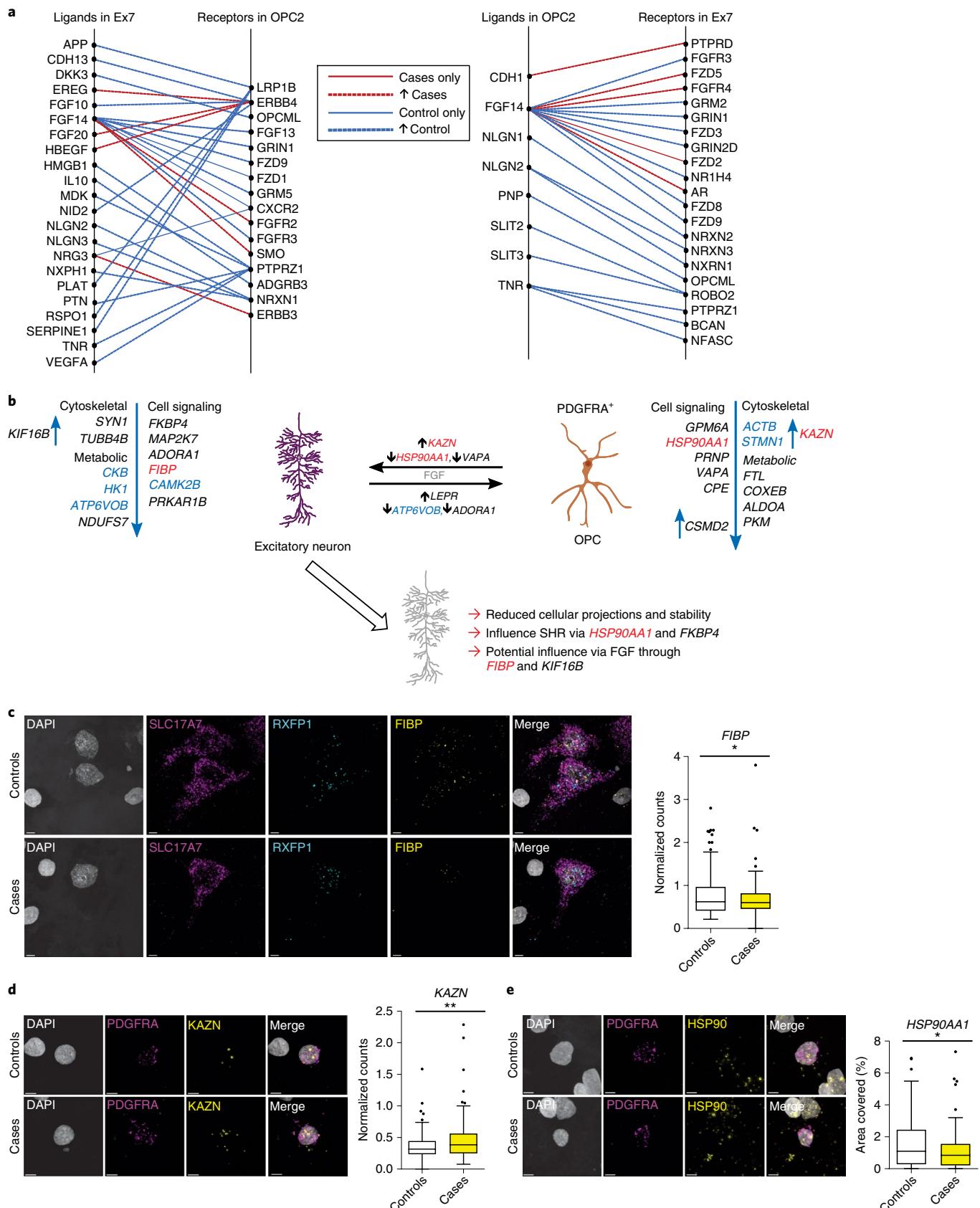
Fig. 5 | WGCNA. **a**, Venn diagram of the overlap between blue module genes and DEGs (hypergeometric test, $P=6.037692 \times 10^{-19}$). **b**, Venn diagram of the overlap between blue module hub genes and DEGs (hypergeometric test, $P=4.954172 \times 10^{-31}$). **c**, Visualization of the top 50 hub genes assessed for the blue module. DEG nodes and all edges connected to them are highlighted in teal. **d**, Box plots of the expression levels of DEGs validated using high-throughput qPCR in FANS-sorted populations that were also hub genes in the blue module. Mann-Whitney U-tests (two-sided) were performed for *PRAF2*, as the values were not normally distributed based on the Shapiro-Wilk's test for normality. All other genes were tested using unpaired two-sided *t*-tests, as their values were normally distributed. * $P<0.05$, ** $P<0.01$, *** $P<0.001$, **** $P<0.0001$. Whiskers on box plots represent maximum and minimum values. Boxes extend from the 25th to the 75th percentiles, the center lines represent the median, and dots represent all values in the dataset. *ATP6VOB*: $n=15$ cases, 11 controls, $t=3.10$, degrees of freedom (d.f.) = 12.62, $P=0.0087$; *CKB*: $n=9$ cases, 7 controls, $t=2.48$, d.f. = 16.85 $P=0.023$; *PRAF2*: $n=14$ cases, 10 controls, $U=8$, $P=6.8 \times 10^{-5}$; *TKT*: $n=16$ cases, 14 controls, $t=2.25$, d.f. = 19.83, $P=0.036$; *PLD3*: $n=15$ cases, 14 controls, $t=3.06$, d.f. = 15.83, $P=0.0075$; *OTUB1*: $n=16$ cases, 14 controls, $t=2.39$, d.f. = 20.92, $P=0.026$; *ACTB*: $n=14$ cases, 15 controls, $t=3.14$, d.f. = 19.98, $P=0.0052$; *HNRNPK*: $n=14$ cases, 13 controls, $t=2.41$, d.f. = 16.07, $P=0.028$.

glial subtypes—including multiple astrocytic, oligodendrocytic and OPC clusters—enabled us to pinpoint changes specific to OPCs, but not oligodendrocytes, and changes selective to only one subset of astrocytic cells.

In recent years, the target cell types in MDD pathophysiology have expanded from excitatory neurons to include inhibitory interneurons¹⁸ and non-neuronal cells^{4–9}. Here, we found 16 unique cell types showing evidence of differential gene expression in depression, including 4 non-neuronal clusters and 6 clusters of interneurons, which provides support for the complex interplay between multiple cell types in MDD. Previous studies have shown that SST and PVALB interneurons are dysregulated in patients with MDD¹⁸; here, we report several DEGs in three interneuron clusters that were defined by the expression of these GABAergic markers (Inhib_3_SST, Inhib_6_SST and Inhib_8_PVALB). Interestingly,

a separate cluster of PVALB interneurons (Inhib_7_PVALB) did not show differential expression, which may indicate that not all PVALB interneurons are equally affected. However, we identified DEGs in non-SST, non-PVALB interneuron clusters (Inhib_2_VIP, Inhib_1 and Inhib_5), which suggests that additional interneuron subtypes could have a role in depression and should be examined in future research.

We found ten different excitatory cell types that were annotated to specific cortical layers based on known markers. Ex10 represented a large cluster of superficial cortical layer cells, whereas there were numerous clusters representing different excitatory cell types from deeper cortical layers. The neuronal cluster with the most change was Ex7, a deep layer cluster characterized primarily by *DPP10* expression. *DPP10* encodes a dipeptidyl-peptidase-related protein that regulates neuronal excitability and has previously



been associated with a human-specific, neuron-based regulatory network. Structural variants of this gene have been implicated in neuropsychiatric diseases, including autism, schizophrenia and bipolar disorder³³.

OPC2 also showed extensive gene expression changes between the cases and the controls. OPC2 was the youngest cell type in the OL pseudotime trajectory. The use of cellular deconvolution techniques indicated that cells of the OPC1 cluster have some similar-

Fig. 6 | Contributions of OPC2 and Ex7. **a**, CCInx receptor-ligand-based cell-cell interaction network analysis for communication between the clusters Ex7 and OPC2. Given the large number of connections (Supplementary Table 47a,b), a subset is shown here. **b**, Our data point to a change in the communication between deep layer excitatory neurons (Ex7) and immature OPCs (OPC2). Altered FGF bidirectional signaling was identified via CCInx. We propose that immature OPCs have an important role in regulating plastic properties of deep layer excitatory cells, such as neuron projection outgrowth and maintenance. Lines between cell types are labeled with secreted or junction proteins that are dysregulated in the given cell type; for example, HPS90AA1 codes for the stress-inducible isoform HSP90 α (which is secreted in certain contexts), KAZN encodes an upregulated junction protein in OPCs, and ATP6V0B could represent altered ATP signaling. Arrows next to gene names indicate upregulation or downregulation. Adjacent to each cell type are the genes in given functional categories and their direction of change in the disease state. **c**, Decreased expression of *FIBP* was validated in deep layer neurons using RNAScope. *SLC17A7* (which encodes VGLUT) was used as a marker for excitatory cells, and *RXFP1* was used to identify deep layer neurons. *SLC17A7*⁺ and *RXFP1*⁺ cells were imaged (left), and *FIBP* expression was counted (right; $n=119$ (cases), 100 (controls) nuclei, unpaired two-sided t-test, $t=2.49$, d.f.=217, * $P=0.013$). **d,e**, Images (left) and quantification (right) showing increased KAZN expression (**d**; $n=95$ (cases), 95 (controls) nuclei, unpaired two-sided t-test, $t=-2.69$, d.f.=188, ** $P=0.008$) and HPS90AA1 expression (**e**; $n=86$ (cases), 102 (controls) nuclei, unpaired two-sided t-test, $t=2.23$, d.f.=186, * $P=0.027$). Expression was validated in OPCs using *PDGFRA* as a marker. For **d–e**, whiskers on box plots represent the 5th and 95th percentiles. Boxes extend from the 25th to 75th percentiles, and the center lines represent the median. Dots represent points beyond the 5th or 95th percentile. Scale bars, 5 μ m.

ity to committed OPCs, whereas cells of the OPC2 cluster showed no such correspondence, which provides support for the idea that there is functional heterogeneity among OPCs³⁴. Furthermore, compared to OPC1 cells, OPC2 cells expressed higher levels of certain glutamate and sodium receptors, which are typically lost as the cells mature³⁴.

Evidence suggests that half of the OPCs (NG2⁺) in the brain do not give rise to any other cell type³⁵ and exhibit synaptic contact with neurons³⁶. As such, OPCs are now thought to be a distinct glial cell type implicated in brain plasticity through roles such as integration of synaptic activity³⁷ and mediation of long-term potentiation³⁸. Additionally, there is evidence that directly implicates the loss of this cell type with the emergence of depressive-like behavior³¹. The data from this study support a role for OPCs in MDD independent from their role as precursor cells for oligodendrocytes.

Our STRING DB protein network analysis highlighted a number of links, including connections between three DEGs encoding kinesin-related proteins: *KIF26B*, *KLC2* and *KIF16B*. *KIF16B* (increased in Ex7) is involved in recycling receptors, including the FGF receptor (FGFR). Interestingly, *FIBP* was decreased in Ex7. FGFR transport relies, in part, on the interaction between kinesins and Rab GTPases³⁹. Notably, we found that *RAB11B* (which encodes a Rab GTPase) and *KLC2* was downregulated in Inhib_3. Taken together, these finding could point to a disruption of FGFR recycling by kinesins and Rab GTPases, as well as disrupted modulation of FGF intercellular signaling by *FIBP* in neurons in MDD.

Based on animal models and cell culture experiments, FGFs (specifically FGF2) and FGFRs seem to be affected by stress and the glucocorticoids⁴⁰. The glucocorticoid receptor (GR) has consistently been implicated in MDD⁴¹. *HSP90AA1* (decreased in OPC2) and *FKBP4* (decreased in Ex7), along with its homolog *FKPB5*, encode co-chaperones for the GR and regulate intracellular signaling functions of this receptor³⁰. *HSP90AA1* encodes the stress-inducible isoform HSP90 α and, interestingly, is secreted in certain stress contexts⁴². These changes may point to a fundamental disruption in GR signaling in deep layer excitatory cells and OPCs, which could further interact with the above-described changes in FGF signaling.

The genes related to chaperone-mediated SHP cycling overlapped with genes involved in innate immune function. This is unsurprising given the role of glucocorticoids in modulating inflammation, one of the primary responses of the immune system. Both OPC2 and Ex7 were enriched for the common genes between these pathways. Finally, both the FGF and GR system have been implicated to have roles in the plastic properties of excitatory neurons, such as projection outgrowth and stability^{43,44}.

Additionally, genes such as *PRNP* (which encodes prion protein) and *KAZN* (a gene involved in desmosome assembly, which encodes

kazrin) were strongly altered in the OPC2 cluster, and these genes are associated with mediating synaptic plasticity and cellular communication^{45,46}. The absence of *Prnp* is associated with an increased number of undifferentiated oligodendrocytes and the delayed expression of differentiation markers⁴⁷, which is intriguing given the evidence implicating a lack of mature adult oligodendrocytes in animal models of depression and anxiety⁴⁸. Meanwhile, over-expression of kazrin in keratinocytes profoundly changed the shape of cells, reduced filamentous actin and impaired the assembly of intercellular junctions⁴⁶. Interestingly, decreased desmosome length has been described in *Prnp*^{-/-} mice⁴⁹, which suggests that there is an interplay between these proteins. Furthermore, a single nucleotide polymorphism in *KAZN* showed one of the strongest associations in individuals with treatment-resistant depression²⁶.

Based on the information we derived from various bioinformatics strategies, we proposed a putative model for the bidirectional interactions between lower layer excitatory neurons and immature oligodendrocytes. We used RNAScope to validate some of the key transcriptional changes highlighted by the model. Although these results are interesting, functional follow-up studies are required to determine the role of molecules such as FGF, HSP90 α and kazrin in the communication between these two cell types.

Our study is not without limitations. All individuals included in our study were male, so our results are not necessarily generalizable to females, particularly as previous studies have suggested that brain transcriptomics changes associated with MDD are different in females⁵⁰. Nonetheless, this first screen provides important information that may help inform subsequent studies exploring both males and females with MDD. Technical limitations with droplet-based snRNA-seq of human brain samples have been previously described. We, like others^{10,11}, found a much greater proportion of neurons compared to glial cells than would be expected based on histologically determined estimates, which points to a potential limitation of the methodology for capturing non-neuronal cells. Although droplet-based snRNA-seq does not capture genes expressed at low levels, we were still able to identify differential gene expression for thousands of genes in precisely defined cell types.

Last, we believe the consistency across dissections was not sufficient for estimating cell-type proportions. For example, even a small over-representation of one cortical layer versus another during dissection can give misleading results regarding the proportion of cell types. Other groups have attempted to extract nuclei from cryo-sectioned samples to address these inconsistencies¹⁰.

Our study elucidated gene expression changes specific to numerous independent cell types in MDD. We identified a potentially important link between OPCs and deep layer excitatory neurons, which implicates fundamental pathways including FGF signaling, glucocorticoid receptor regulation and synaptic plasticity in the

brains of depressed individuals. The generalizability of these data will rely on independent validation in other MDD cohorts; nonetheless, this work provides an exciting starting point for understanding the complex interplay of cells in the brain and a platform for future functional research to assess these potential interactions. Future single-cell studies of MDD should aim to relate cell types with symptomology and severity, as has been done in recent papers^{16,17}.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41593-020-0621-y>.

Received: 19 July 2019; Accepted: 12 March 2020;

Published online: 27 April 2020

References

- Depression and Other Common Mental Disorders: Global Health Estimates (World Health Organization, 2017).
- Wray, N. R. et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50**, 668–681 (2018).
- Jansen, R. et al. Gene expression in major depressive disorder. *Mol. Psychiatry* **21**, 339–347 (2016).
- Sequeira, A. et al. Global brain gene expression analysis links glutamatergic and GABAergic alterations to suicide and major depression. *PLoS ONE* **4**, e6585 (2009).
- 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).
- 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 (2016).
- Edgar, N. & Sible, E. A putative functional role for oligodendrocytes in mood regulation. *Transl Psychiatry* **2**, e109 (2012).
- Nagy, C. et al. Astrocytic abnormalities and global DNA methylation patterns in depression and suicide. *Mol. Psychiatry* **20**, 320–328 (2015).
- Duman, R. S., Aghajanian, G. K., Sanacora, G. & Krystal, J. H. Synaptic plasticity and depression: new insights from stress and rapid-acting antidepressants. *Nat. Med.* **22**, 238–249 (2016).
- Lake, B. B. et al. Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat. Biotechnol.* **36**, 70–80 (2018).
- Habib, N. et al. Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods* **14**, 955–958 (2017).
- Mancarci, B. O. et al. Cross-laboratory analysis of brain cell type transcriptomes with applications to interpretation of bulk tissue data. *eNeuro* **4**, ENEURO.0212-17.2017 (2017).
- Zheng, G. X. Y. et al. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
- Lake, B. B. et al. A comparative strategy for single-nucleus and single-cell transcriptomes confirms accuracy in predicted cell-type expression from nuclear RNA. *Sci. Rep.* **7**, 6031 (2017).
- Renthal, W. et al. Characterization of human mosaic Rett syndrome brain tissue by single-nucleus RNA sequencing. *Nat. Neurosci.* **21**, 1670–1679 (2018).
- Mathys, H. et al. Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332–337 (2019).
- Velmeshev, D. et al. Single-cell genomics identifies cell type-specific molecular changes in autism. *Science* **364**, 685–689 (2019).
- Northoff, G. & Sible, E. Why are cortical GABA neurons relevant to internal focus in depression? A cross-level model linking cellular, biochemical and neural network findings. *Mol. Psychiatry* **19**, 966–977 (2014).
- Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
- Sofroniew, M. & Vinters, H. Astrocytes: biology and pathology. *Acta Neuropathol.* **119**, 7–35 (2010).
- Anderson, M. A., Ao, Y. & Sofroniew, M. V. Heterogeneity of reactive astrocytes. *Neurosci. Lett.* **565**, 23–29 (2014).
- Trapnell, C. et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).
- Thomas, P. D. et al. Applications for protein sequence-function evolution data: mRNA/protein expression analysis and coding SNP scoring tools. *Nucleic Acids Res.* **34**, W645–W650 (2006).
- Butts, B. D., Houde, C. & Mehmet, H. Maturation-dependent sensitivity of oligodendrocyte lineage cells to apoptosis: implications for normal development and disease. *Cell Death Differ.* **15**, 1178–1186 (2008).
- Jakel, S. et al. Altered human oligodendrocyte heterogeneity in multiple sclerosis. *Nature* **566**, 543–547 (2019).
- Li, Q. S., Tian, C., Seabrook, G. R., Drevets, W. C. & Narayan, V. A. Analysis of 23andMe antidepressant efficacy survey data: implication of circadian rhythm and neuroplasticity in bupropion response. *Transl Psychiatry* **6**, e889 (2016).
- Gutierrez-Sacristan, A. et al. PsyGeNET: a knowledge platform on psychiatric disorders and their genes. *Bioinformatics* **31**, 3075–3077 (2015).
- Pinero, J. et al. DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res.* **45**, D833–D839 (2017).
- Szklarczyk, D. et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).
- Wochnik, G. M. et al. FK506-binding proteins 51 and 52 differentially regulate dynein interaction and nuclear translocation of the glucocorticoid receptor in mammalian cells. *J. Biol. Chem.* **280**, 4609–4616 (2005).
- Birey, F. et al. Genetic and stress-induced loss of NG2 glia triggers emergence of depressive-like behaviors through reduced secretion of FGF2. *Neuron* **88**, 941–956 (2015).
- Mason, J. L. & Goldman, J. E. A2B5⁺ and O4⁺ cycling progenitors in the adult forebrain white matter respond differentially to PDGF-AA, FGF-2, and IGF-1. *Mol. Cell. Neurosci.* **20**, 30–42 (2002).
- Shulha, H. P. et al. Human-specific histone methylation signatures at transcription start sites in prefrontal neurons. *PLoS Biol.* **10**, e1001427 (2012).
- Spitzer, S. O. et al. Oligodendrocyte progenitor cells become regionally diverse and heterogeneous with age. *Neuron* **101**, 459–471.e5 (2019).
- Psachoulia, K., Jamen, F., Young, K. M. & Richardson, W. D. Cell cycle dynamics of NG2 cells in the postnatal and ageing brain. *Neuron Glia Biol.* **5**, 57–67 (2009).
- Bergles, D. E., Jabs, R. & Steinhauser, C. Neuron–glia synapses in the brain. *Brain Res. Rev.* **63**, 130–137 (2010).
- Birey, F., Kokkoski, A. G. & Aguirre, A. Oligodendroglia-lineage cells in brain plasticity, homeostasis and psychiatric disorders. *Curr. Opin. Neurobiol.* **47**, 93–103 (2017).
- Ge, W. P. et al. Long-term potentiation of neuron–glia synapses mediated by Ca²⁺-permeable AMPA receptors. *Science* **312**, 1533–1537 (2006).
- Ueno, H., Huang, X., Tanaka, Y. & Hirokawa, N. KIF16B/Rab14 molecular motor complex is critical for early embryonic development by transporting FGF receptor. *Dev. Cell* **20**, 60–71 (2011).
- Turner, C. A., Watson, S. J. & Akil, H. The fibroblast growth factor family: neuromodulation of affective behavior. *Neuron* **76**, 160–174 (2012).
- Turecki, G. & Meaney, M. J. Effects of the social environment and stress on glucocorticoid receptor gene methylation: a systematic review. *Biol. Psychiatry* **79**, 87–96 (2016).
- Zuehlke, A. D., Beebe, K., Neckers, L. & Prince, T. Regulation and function of the human HSP90AA1 gene. *Gene* **570**, 8–16 (2015).
- Huang, J. Y., Lynn Miskus, M. & Lu, H. C. FGF–FGFR mediates the activity-dependent dendritogenesis of layer IV neurons during barrel formation. *J. Neurosci.* **37**, 12094–12105 (2017).
- Pittenger, C. & Duman, R. S. Stress, depression, and neuroplasticity: a convergence of mechanisms. *Neuropharmacology* **33**, 88–109 (2008).
- Caiati, M. D. et al. PrPC controls via protein kinase A the direction of synaptic plasticity in the immature hippocampus. *J. Neurosci.* **33**, 2973–2983 (2013).
- Sevilla, L. M., Nachat, R., Groot, K. R. & Watt, F. M. Kazrin regulates keratinocyte cytoskeletal networks, intercellular junctions and differentiation. *J. Cell Sci.* **121**, 3561–3569 (2008).
- Bribian, A. et al. Role of the cellular prion protein in oligodendrocyte precursor cell proliferation and differentiation in the developing and adult mouse CNS. *PLoS ONE* **7**, e33872 (2012).
- Liu, J. et al. Impaired adult myelination in the prefrontal cortex of socially isolated mice. *Nat. Neurosci.* **15**, 1621–1623 (2012).
- Morel, E. et al. The cellular prion protein PrP(c) is involved in the proliferation of epithelial cells and in the distribution of junction-associated proteins. *PLoS ONE* **3**, e3000 (2008).
- Labonte, B. et al. Sex-specific transcriptional signatures in human depression. *Nat. Med.* **23**, 1102–1111 (2017).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2020

Methods

Participants: postmortem brain samples. This study was approved by the Douglas Hospital Research Ethics Board, and written informed consent from next of kin was obtained for each individual. Postmortem brain samples were provided by the Douglas–Bell Canada Brain Bank (www.douglasbrainbank.ca). Frozen gray matter samples were dissected from Brodmann area 9 (dlPFC). Brains were dissected by trained neuroanatomists and stored at -80°C . For each individual, the cause of death was determined by the Quebec Coroner's office, and psychological autopsies were performed using proxy-based interviews, as previously described³¹. Cases met criteria for MDD and died by suicide, whereas controls were individuals who died suddenly and did not have evidence of any axis I disorders (Table 1). The post mortem interval (PMI) represents the delay between an individual's death and collection and processing of the brain. To assess RNA quality, we measured the RNA integrity number (RIN) obtained for our samples using tissue homogenates. An unpaired, two-tailed, Student's *t*-test revealed no significant difference ($P=0.15$) in RIN between cases (mean RIN of 6.74) and controls (mean RIN of 6.16). A total of 17 cases and 17 controls were included in the snRNA-seq experiment, and the full cohort of participants (except 25) was used for follow-up validation of DEGs by FANS and high-throughput qPCR. RNAscope experiments were performed on representative subsets of samples using five cases and five matched controls. Detailed information on experimental design and reagents can also be found in the Nature Research Reporting Summary.

Nuclei isolation and capture. Frozen tissue (50 mg) was dounced in 3 ml of lysis buffer, 10 times with a loose pestle and an additional 5 times with the tight pestle. The lysis buffer contained 10 mM Tris (pH 7.4), 10 mM NaCl, 3 mM MgCl₂ and 0.05% (v/v) NP-40 detergent. The sample was left to lyse in a total of 5 ml of buffer for 5 min, after which 5 ml of wash buffer was added and swirled. The sample was passed through a 30-μm cell strainer and spun for 5 min at 500 × g. This step was repeated for a total of two filtering steps. After pelleting, the nuclei were resuspended in 5–10 ml of wash buffer by pipetting up and down 8–10 times. After 3 washes, the nuclei were resuspended in 1 ml of wash buffer and mixed with 25% Optiprep and layered on a 29% Optiprep cushion and spun for 30 min at 10,000 × g. Nuclei were resuspended in wash buffer to achieve a concentration of $\sim 5 \times 10^5$ nuclei per ml. Representative images of extracted nuclei are presented in Supplementary Fig. 12.

We used the 10x Genomics Chromium controller for single-cell gene expression to isolate single nuclei for downstream bulk-RNA library preparation. We strictly followed the protocol as outlined by the user guide (CG00052_SingleCell3_ReagentKitv2UserGuide_RevD.pdf; latest version at <https://bit.ly/3dUNOLZ>), with the exception of loading concentration, which we increased by 30% as we assessed the capture of nuclei to be slightly less efficient than cell encapsulation. We aimed to capture $\sim 3,000$ nuclei per sample. So, for example, if our sample concentration was 390 nuclei per μl, (~ 400 nuclei per μl) according to page 10 of Protocol Step 1 of the user guide, we were required to load 13.1 μl of the stock to capture 3,000 cells. Instead, we recalculated our stock concentration to be 70% of 390 = 273 nuclei per μl and therefore load 17.4 μl (the recommended amount for 300 nuclei per μl). This system only allows for a maximum of eight samples per capture run. As such, we required multiple batches to collect the individual nuclei for all 34 samples (6 batches). Samples 24 and 25 performed poorly; therefore, we carried out the capture on two separate chips and sequenced twice, combining the data from both runs for the final analysis.

Sequence alignment and UMI counting. A pre-mRNA transcriptome was built using the cellranger mkref (Cellranger v.2.0.1) command and default parameters starting with the refdata-cellranger-GRCh38-1.2.0 transcriptome and as per the instructions provided on the 10x Genomics website. Reads were demultiplexed by the sample index using the cellranger mkfastq command (Cellranger v.2.1.0). Fastq files were aligned to the custom transcriptome, cell barcodes were demultiplexed and UMIs corresponding to genes were counted using the cellranger count command and default parameters.

Data transformation for secondary analysis. The unfiltered gene barcode matrices for each sample were loaded into R using the Read10X function in the Seurat R package (v.2.2.0, 2.3.0)¹⁹. Cell names were modified such that the participant name, batch and biological condition were added to them. Seurat objects were created corresponding to each sample using the CreateSeuratObject function with the imported unfiltered gene–barcode matrices provided as the raw data. Individual Seurat objects for each sample were sequentially combined into one object using the MergeSeurat function. No filtering or normalization was performed up to this step. Since this is a single-nucleus dataset, all mitochondrial genes that were transcribed from the mitochondrial genome were removed, along with genes not detected in any cell.

Barcode and gene filtering. Based on the distribution of nGene (the total number of genes detected in each cell) for the total dataset (assessed by summary and hist R³² functions), barcodes that were associated with fewer than 110 detected genes were removed. Based on the distribution of nUMI (the total number of UMIs detected in each cell), the top 0.5% of barcodes were also excluded as most likely

being multiplets rather than single nuclei, as there was a very sharp increase of nUMI from 16,393 at the 99.5th percentile to 102,583 at the maximum.

Next, the distribution of nUMI for the remaining barcodes was fitted to three normal distributions using the normalmixEM function from the mixtools³³ package (Supplementary Fig. 1c). The rationale for this was that the filtered barcodes contain a population of low-quality ‘noise’ barcodes that have a very low nUMI on average, a population of non-neuronal cells that have an intermediate nUMI and a population of neuronal cells that have a high nUMI. Based on the fitting of the normal distributions, only the barcodes with a high probability (>0.95) of belonging to either the putative ‘non-neuronal’ or putative ‘neuronal’ distributions, and a low probability (<0.05) of belonging to the ‘noise’ distribution were retained for further analysis (Supplementary Fig. 1c,d). A total of 78,886 cells and 30,062 genes were retained.

Our custom filtering (Supplementary Fig. 1a–e and Supplementary Table 4) helped increase the number of glial cells recovered. With an initial subset of 20 participants, applying our custom filtering criteria increased the total number of cells 1.8-fold, but it also increased the number of non-neuronal cells by almost sixfold (data not shown). After custom filtering, the minimum numbers of genes and UMIs per nucleus were 254 and 340, respectively.

Once nuclei were filtered, the percentages of mitochondrial reads associated with the retained barcodes were calculated, although for quality control purposes, those reads were not used during the filtering or downstream analysis (Supplementary Fig. 1f). Although the percentage of reads mapping to mitochondrially expressed genes is a more pertinent quality control parameter for single-cell rather than single-nucleus approaches, contaminating mitochondrial reads often present a problem in single-nucleus protocols (B. B. Lake, personal communication). However, our optimized approach was able to minimize this technical issue.

Data processing and dimensionality reduction. The UMI counts were normalized to 10,000 counts per cell and converted to log scale (Seurat function NormalizeData). The batch, condition and personal identifying information were added as metadata to the final Seurat object; nUMI and batch were regressed out using the ScaleData function. The Seurat FindVariableGenes function was used with default selections and cut-offs as follows: x.low.cutoff = 0.003, x.high.cutoff = 2 and y.cutoff = 1. This resulted in a list of 2,135 highly variable genes, which excluded lowly expressed genes (below the 25th percentile), and selected only the top 10% of genes in terms of the scaled dispersion. These highly variable genes were used to calculate 100 PCs. Based on the PC elbow plot of the standard deviation of the PCs (Supplementary Fig. 2a), the first 50 PCs were retained for use in the downstream analysis.

Clustering by gene expression. The FindClusters function was applied with a resolution of 2.5 and produced 44 initial clusters. The goal of clustering is to sort nuclei by cell type so that all remaining gene expression variation within clusters is not related to cell differentiation processes. Before the advent of single-nuclei expression profiling, cell types were identified by observing differences in cell morphology, behavior and anatomic location. It is fairly straightforward to sort single-nuclei expression profiles into known cell types according to the expression levels of marker genes that differentiate between these cell types. However, it is very unlikely that all cell types have been identified, so we must rely on nuclei clustering to uncover as-yet unknown cell types. Unfortunately, the number of clusters obtained from the clustering algorithm is somewhat arbitrary because clustering depends on the settings of several parameters, and there is no consensus on how they should be set. Although clusters obtained using reasonable default settings usually correspond to known biological cell types, some clusters may appear to potentially identify entirely new cell types or splinter existing cell types into multiple subtypes. Deciding whether the clusters really do identify new cell types can be difficult or may even be impossible from available data.

To address this issue, we used tools in the Seurat package to sequentially combine any clusters that were not sufficiently distinct from each other. In particular, after performing initial hierarchical clustering of the graph-based clusters (BuildClusterTree), we assessed the nodes of the dendrogram using a random forest classifier (AssessNodes) and then merged together any nodes that were in the bottom 25% of the dendrogram (using the branching.times function from the ape R package³⁴) and had an out-of-bag error of more than 5%. We then repeated this clustering and merging process for the nuclei within each terminal node until none of the remaining nodes fulfilled our cut-off criteria (Supplementary Fig. 2b). The resulting set of 30 clusters were then characterized in terms of known marker genes of all major well-defined brain cell types (Supplementary Fig. 2c–d). To refine the identification of excitatory neuron types, we combined and reclustered a set of excitatory clusters with highly correlated gene expression profiles ($R > 0.95$) (Supplementary Fig. 13a–c) using similar parameters for clustering as the whole dataset. This included 7 clusters of $\sim 40,000$ cells. Reclustering yielded 33 final clusters for downstream analysis. Finally, the clusters were manually curated to eliminate potential biases; for example, clusters were removed if mainly one sample contributed to the cells contained within the cluster (Supplementary Tables 48–51 and Supplementary Fig. 14a–e).

Cluster annotation. Genes used as markers for major cell types and layer specificity are listed below. Inhibitory neuron subtypes were annotated based on the expression of the canonical inhibitory interneuron markers *SST*, *PVALB* and *VIP* where possible. Excitatory neuron subtypes were annotated with some level of layer specificity based on the expression of layer-specific markers^{11,55,56}. We also characterized clusters in terms of all genes differentially expressed between clusters (FindAllMarkers function, bimodal test, logfc.threshold of log(2), other parameters set to default) (Supplementary Table 6).

Major cell-type markers are presented in Supplementary Fig. 3a–p.

The following markers were used for cell types: macrophage/microglia: *SPI1*, *MRC1*, *TMEM119* and *CX3CR1*; endothelial: *CLDN5* and *VTN*; astrocytes: *GLUL*, *SOX9*, *AQP4*, *GJA1*, *NDRG2*, *GFAP*, *ALDH1A1*, *ALDH1L1* and *VIM*; OPCs: *PTGDS*, *PDGFRA*, *PCDH15*, *OLIG2* and *OLIG1*; oligodendrocytes: *PLP1*, *MAG*, *MOGP* and *MBP*; excitatory neurons: *SATB2*, *SLC17A7* and *SLC17A6*; inhibitory neurons: *GAD1*, *GAD2* and *SLC32A1*; neurons: *SNAP25*, *STMN2* and *RBFOX3*.

The following layer-specific markers were used: layer II: *GLRA3*; layer II/III: *LAMP5* and *CARTPT*; layers II–IV: *CUX2* and *THSD7A*; layers II–VI: *RASGRF2* and *PVRL3*; layer III/IV: *PRSS12*; layer IV/V: *RORB*; layers IV–VI: *GRK4*; layer V: *KCNK2*, *SULF2*, *PCP4*, *HTR2C* and *FEZF2*; layer V/VI: *TOX*, *ETV1*, *RPRM*, *RXFP1* and *FOXP2*; layer VI: *SYT6*, *OPRK1*, *NR4A2*, *SYNPR*, *TLE*, *NTNG2* and *ADRA2A*.

Pseudotime trajectory using Monocle. For oligodendrocyte developmental trajectory assessment, the data for cells belonging to the five clusters in the OL (Oligos_1, Oligos_2, Oligos_3, OPCs_1, OPCs_2) were used to create a separate Seurat object using the SubsetData function. The most variable genes for these clusters alone were identified using the FindVariableGenes function and the following parameters: x.low.cutoff = 0.003, x.high.cutoff = 3 and y.cutoff = 1 (giving a total of 895). The Seurat object was imported into a CDS (CellDataSet) object using the Monocle²² function importCDS.

Estimation of size factors and dispersions was performed (using the estimateSizeFactors and estimateDispersions Monocle functions) on the CDS object using default parameters. Dimensionality reduction was then performed using reduceDimension, with reduction_method set to DDRTree. The 895 variable genes identified above were used for ordering the cells into a trajectory with the orderCells function. The pseudotime trajectory was then plotted with plot_cell_trajectory (Fig. 2e), and the change in expression of genes known to be involved in oligodendrocyte development were plotted using plot_genes_in_pseudotime (Supplementary Fig. 6b–i). differentialGeneTest was applied separately to OL cells from the controls and the MDD cases with fullModelFormulaStr = “sm.ns(Pseudotime)”. This allowed us to model the expression of each gene as a function of pseudotime. All genes detected in at least one cell in the respective group were compared and their changes across pseudotime were assessed. A *q* value cut-off of <0.01 was used to identify genes associated with pseudotime. The overlapping and non-overlapping genes were identified by comparing the lists obtained for the two groups (Supplementary Fig. 6a).

Purification of clusters for differential expression. Our doublet-removal approach consisted of calculating a median gene expression profile for all our clusters, calculating the correlation of the gene expression of each cell, with the median profile of its cluster (considering only the top 865 genes whose median expression was highly variable; that is, had a variance of >0.25 across the different clusters), and selecting cells with high correlation. This was done by fitting bimodal normal distributions to the total distribution of correlations in the cluster to identify low and high correlation peaks. Cells were retained only if they had a low probability of falling in the low correlation peak ($P < 0.25$) and a high probability ($P > 0.75$) of falling in the high correlation peaks (Supplementary Fig. 7).

Differential gene expression analysis. Differential expression analysis between the cases and the controls was performed using linear mixed models implemented in the lme4 (ref. ⁵⁷) and lmerTest⁵⁸ R packages. Mixed models were necessary to account for dependencies between nuclei obtained from the same brain. Biological condition and number of UMIs were included in the models as fixed effects, and the individual and batch were included as random effects. The inclusion of participant as a random effect should account for person-specific effects such as age and PMI, as well as technical effects of capture and library preparation, which were separately performed for each brain. A FDR of 0.1 was used to detect DEGs within each cell type.

WGCNA. The average cell expression for each sample across every cluster was calculated. These average counts were converted to log + 1 counts to reduce dispersion. WGCNA was carried out in R with the WGCNA package (v.1.68) by Langfelder and Horvath. Genes with insufficient variance were excluded, as well as outlier samples. After some tests, a soft-thresholding power of 7 and a minimum module size of 30 genes were selected for the gene network construction. Resulting modules were correlated with the phenotype information (MDD versus control), as well as each sample’s respective composition of each of the 26 single-cell type clusters they were composed of.

We performed hub gene analysis on the blue module, which was the largest module (2,699 genes) that was correlated to phenotype. Potential hub genes were identified in the module of interest by selecting genes with a module membership larger than 0.80 and a gene significance larger than 0.20 with a $P < 0.05$. The top 50 potential hub genes were extracted alongside any weighted interaction of more than 0.2. The resulting network was visualized using Cytoscape (3.7.1).

FANS. Nuclear suspensions were prepared from 80–100 mg of postmortem brain tissue from Brodmann area 9 as previously described¹⁵⁹, but with the following modifications: homogenized tissue was centrifuged on the sucrose layer at 800 $\times g$ for 20 min at 4 °C, followed by another centrifugation in nuclei extraction buffer. Resuspended nuclei were stained with the following primary antibodies in 600 μl of blocking buffer, incubating at room temperature, away from light, with rotation for 2 h: mouse anti-CUTL2-PerCP conjugated (1:100, Novus, catalog number (cat. no.) H00023316-M03, clone 2H8, conjugated to PerCP using the Novus Lightning Link Labeling kit, cat. no. 718-0010); goat anti-SOX10 (1:100, R&D Systems, cat. no. AF2864); and mouse anti-NeuN-A700 (1:300, Novus, cat. no. NBP1-92693AF700, clone-1B7). Secondary antibody (donkey anti-goat Alexa Fluor 488, 1:1,000, JacksonImmuno, 705-545-147) was added and incubated for 1 h at room temperature with rotation. All antibodies were purchased from Cedarlane. Nuclei were washed with PBS, and the DNA was stained with Hoechst 33342 (Invitrogen, H1399).

FACSAria Fusion (BD Biosciences) was used for sorting four cell populations—SOX10 positive, SOX10 negative, CUTL2 positive and CUTL2 negative. The gating strategy used for sorting is shown in Supplementary Fig. 11 and was as follows. Doublet discrimination was achieved by gating Hoechst-33342-stained singlets in FSC-A versus Hoechst-A plot using a 350-nm ultraviolet laser and a 450/50 filter. Subsequent SOX10-positive, SOX10-negative and NeuN-positive populations were gated in an Alexa Fluor 700-A versus Alexa Fluor 488-A plot utilizing a red 640-nm laser in combination with a 730/45 filter and a blue 488-nm laser in combination with a 530/30 filter, respectively. CUTL2-positive and CUTL2-negative populations, the derivatives of the NeuN-positive gate, were defined in a Alexa 488-A versus PerCP-A (blue 488-nm laser, 695/40 filter) plot with interval gates. The CUTL2-positive population was identified as 30–40% of the NeuN-positive population with the highest CUTL2-PerCP fluorescence. For gating of the CUTL2-negative population, the SOX10-negative and the SOX10-positive populations were displayed in a Alexa 488-A versus PerCP-A plot, and the CUTL2-negative population was gated within the PerCP intensities of the SOX10 populations. The CUTL2-negative population constituted nearly 10% of the NeuN-positive population.

Validation information for antibodies is as follows: Novus H00023316-M03 was validated via western blots and ELISAs, used in one publication in human brain tissue (PMID: 29126813); R&D Systems AF2864 was validated via western blots against human SOX10 protein, ELISAs, immunocytochemistry, with 19 citations; Novus NBP1-92693AF700 was validated via immunocytochemistry, immunohistochemistry, western blots, used in one publication for flow cytometry in human brain tissue (PMID: 28750583).

High-throughput qPCR. RNA was extracted from FANS-sorted nuclei populations using a Norgen RNA/DNA Purification kit (catalog no. 48700). Complementary DNA was synthesized using a modified SMART-seq procedure as previously described⁶⁰. The Fluidigm Biomark system was used for performing high-throughput qPCR as per the manufacturer’s protocol as previously described⁶¹. Fluidigm Delta Gene primer designs were used for the 93 targets (all differentially expressed transcripts, excluding AC133680.1) and 3 endogenous controls (GAPDH, POLR2A and UBC).

Cell–cell interaction measurement. To assess cell–cell communication, we calculated predicted ligand–receptor interactions between Ex7 and OPC2 clusters using CCInx⁶² (<https://github.com/BaderLab/CCInx>), in which the connection between each ligand and receptor is quantified as an edge weight. We chose a gene expression threshold of 2.75 and above to limit our search to relatively highly expressed ligands and receptors and for ease of visualization. To test whether the edge weights were significantly different between the cases and the controls, we randomly permuted our participants into two groups 100 times and formed normal distributions of the edge weight differences between groups for each ligand–receptor pair. We then calculated a *P* value for the case–control edge weight difference for each ligand–receptor pair based on its position in the distribution. Edge weight difference values of $P < 0.01$ were considered significant. A sample script used for assessing the significance of edges has been provided.

Cell deconvolution for all clusters. Expression data from dbGaP (phs000424.v8.p2 and phs000424.v8.p1, available at <https://www.gtexportal.org/home/datasets>)¹¹ were used as reference signatures for annotated cell types. UMI counts for each cell were converted to transcripts per million (TPMs) to account for the varying sequencing depth of each cell and sample. Average expression levels were calculated for each cell-type-specific cluster defined in the paper.

Cluster-specific gene expression profiles were obtained by summing the UMI values of all 24,301 genes common to our dataset and the reference for

each nucleus in each cluster and converting the sums to TPMs. The R package DeconRNASeq (v.1.18.0)²⁵ was used to deconvolute these cluster-specific profiles. Using the data from ref. ¹¹ as reference, we were able to estimate the cell-type composition of our clusters.

Cell deconvolution for OL cells. The average expression from every control sample from the Jäkel et al. dataset²⁵ were calculated and used as cell signatures for the deconvolution of our oligodendrocytic clusters (the average cell expression of every cell in the cluster was considered as bulk) using the R package DeconRNASeq (v.1.26.0).

RNAscope fluorescent in situ hybridization. Frozen blocks of samples from Brodmann area 9 were serially cut using a cryostat (10 µm in thickness) on superfrost-charged slides and kept at -80°C until further processing. In situ hybridization was performed using RNAscope probes and reagents from Advanced Cell Diagnostics and according to the manufacturer's instructions in five matched participants per group. Briefly, sections were first fixed in chilled 10% neutral buffered formalin for 15 min at 4°C, dehydrated in increasing gradients of ethanol baths and left to air dry for 5 min. Endogenous peroxidase activity was quenched with hydrogen peroxide reagent for 10 min, followed by protease digestion for 30 min at room temperature. The following sets of probes were then hybridized for 2 h at 40°C in a humidity-controlled oven (HybEZ II, ACDbio): Hs-RXFP1 (cat. no. 422821), Hs-FIBP (cat. no. 569781-C2) and Hs-SLC17A7 (cat. no. 415611-C3) to quantify *FIBP* expression in excitatory (*SLC17A7*⁺) layer V/VI (*RXFP1*⁺) neurons; and KAZN (cat. no. 569791), *PDGFRA* (cat. no. 604481-C3) and *HSP90AA1* (cat. no. 477061) to quantify KAZN and HSP90AA1 expression in OPCs (*PDGFRA*⁺). Successive addition of amplifiers was performed using the proprietary AMP reagents, and the signal visualized through probe-specific horseradish-peroxidase-based detection by tyramide signal amplification with Opal dyes (Opal 520, Opal 570 and Opal 690; Perkin Elmer) diluted 1:300. Slides were then coverslipped with Vectashield mounting medium with 4,6-diamidino-2-phenylindole (DAPI) for nuclear staining (Vector Laboratories) and kept at 4°C until imaging.

Imaging and analysis of in situ RNA expression. Image acquisition was performed on a laser scanning confocal microscope (FV1200) equipped with a motorized stage. For each experiment and brain, around 10 stack images were taken to capture at least 20 cells of interest per brain: OPCs (*PDGFRA*⁺) and excitatory neurons (*SLC17A7*⁺) from cortical layers V/VI (*RXFP1*⁺). Images were taken using a ×60 objective (numerical aperture of 1.42) with a x-y pixel width of 0.3 µm and z spacing of 0.4 µm. Laser power and detection parameters were kept consistent between brains for each set of experiments. Because tyramide signal amplification with Opal dyes yields a high signal-to-noise ratio, parameters were set so that autofluorescence from lipofuscin and cellular debris was filtered out of the image. Positivity for cell-defining markers was determined by bright clustered puncta-like signals present within the nucleus and cytoplasm of the cells. Expression of genes of interest was quantified using the Analyze Particles function in Fiji²⁶. Stacks were first converted to z projections, and for each image, cell nuclei of cells of interest were manually contoured based on DAPI staining. Single-labeled molecules of RNA were automatically counted in each channel using the find maxima function with a noise tolerance of 350 for *FIBP* and *RXFP1*, and 400 for KAZN and *PDGFRA*. Normalized *FIBP* and KAZN expression per cell was calculated by dividing *FIBP* and KAZN raw counts by *RXFP1* and *PDGFRA* raw counts, respectively. *HSP90AA1* expression was manually quantified by thresholding the signal per image and measuring the percentage of area of the nucleus covered by the resulting mask.

Statistical analysis. No statistical methods were used to predetermine sample sizes. Sample sizes were determined based on sample sizes used in previous similar studies. Individuals were assigned to groups based on diagnosis and not by random assignment. All participants were male, and groups were matched for age (18–87 years), postmortem interval (12–93 h) and brain pH (6–7.01). Clinicians were blinded to the final psychological autopsy-based diagnosis of MDD case or control. Clustering of single-nuclei gene expression profiles was performed in an unbiased blinded manner. Cluster annotations were assigned after generation of clusters.

Clusters were excluded from downstream analysis if they did not show even contribution from individuals, as these clusters are likely to reflect sample-specific artifacts rather than biological variability of interest. Single nuclei were excluded from cell-type clusters based on their level of correlation to the median expression profile of the cluster (nuclei that exhibited low correlation were removed) as detailed above to ensure that differential gene expression analysis was performed using similar nuclei populations from the cases and the controls. The exclusion criteria were not pre-established and were chosen based on a preliminary analysis of the data.

Differential expression analysis between the cases and the controls in the snRNA-seq data was performed using linear mixed models implemented in the lme4 (ref. ²⁷) and lmerTest²⁸ R packages with biological condition and number of UMIs as fixed effects, the participant and batch as random effects, and a FDR of 0.1 for significance. All DEGs from this analysis were also significantly differentially

expressed between the cases and the controls when retested using two-tailed Wilcoxon tests (Supplementary Table S2). For the analysis of the RNAscope results, two-tailed *t*-tests were performed with a significance threshold of *P* < 0.05; data distribution was assumed to be normal, but this was not formally tested. For the analysis of high-throughput qPCR data, two-tailed *t*-tests or two-tailed Wilcoxon rank sum tests (that is, Mann–Whitney *U*-tests) were performed, both at a significance threshold of *P* < 0.05 and depending on data normality as measured using the Shapiro–Wilk's test. The results in Supplementary Tables 43–46 are for genes reliably detected in more than nine participants per group.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Raw sequencing data, annotated gene–barcode matrix and lists of cells used for differential gene expression analysis are accessible on GEO using the accession number GSE144136. RNAscope and high-throughput qPCR data are available upon request.

Code availability

A sample custom R script (Supplementary_R_Script_1.R) used for analyzing high-throughput qPCR data is provided and an R script used to test the statistical significance of CCInx interactions is provided (Supplementary_R_Script_2.R) along with this paper.

References

51. 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).
52. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2017).
53. Benaglia, T., Chauveau, D., Hunter, D. R. & Young, D. S. mixtools: an R Package for analyzing mixture models. *J. Stat. Soft.* **32**, 29 (2009).
54. Paradis, E., Claude, J. & Strimmer, K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289–290 (2004).
55. Hawrylycz, M. J. et al. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**, 391–399 (2012).
56. Lake, B. B. et al. Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science* **352**, 1586–1590 (2016).
57. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Soft.* **67**, 48 (2015).
58. Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. lmerTest Package: tests in linear mixed effects models. *J. Stat. Soft.* **82**, 26 (2017).
59. Lutz, P. E. et al. Association of a history of child abuse with impaired myelination in the anterior cingulate cortex: convergent epigenetic, transcriptional, and morphological evidence. *Am. J. Psychiatry* **174**, 1185–1194 (2017).
60. Bayega, A. et al. in *Gene Expression Analysis: Methods and Protocols* (eds Raghavachari, N. & Garcia-Reyero, N.) 121–147 (Springer New York, 2018).
61. Spurgeon, S. L., Jones, R. C. & Ramakrishnan, R. High throughput gene expression measurement with real time PCR in a microfluidic dynamic array. *PLoS ONE* **3**, e1662 (2008).
62. Ximerakis, M. et al. Single-cell transcriptomic profiling of the aging mouse brain. *Nat. Neurosci.* **22**, 1696–1708 (2019).
63. Gong, T. & Szustakowski, J. D. DeconRNASeq: a statistical framework for deconvolution of heterogeneous tissue samples based on mRNA-Seq data. *Bioinformatics* **29**, 1083–1085 (2013).
64. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).

Acknowledgements

G.T. holds a Canada Research Chair (Tier 1) and a NARSAD Distinguished Investigator Award. He is supported by grants from the Canadian Institute of Health Research (CIHR) (FDN148374 and EGM141899). We acknowledge the expert help of the Douglas–Bell Canada Brain Bank staff (J. Prud'homme, M. Bouchouka and A. Baccichet), and H. Djambazian at the MUGQIC. The work was also supported by CIHR projects 32557 and 33408 (to J.R.). The Douglas–Bell Canada Brain Bank is supported in part by funding from the Canada First Research Excellence Fund, awarded to McGill University for the Healthy Brains for Healthy Lives project, and from the Fonds de recherche du Québec–Santé (FRQS) through the Quebec Network on Suicide, Mood Disorders and Related Disorders. The present study used the services of the Molecular and Cellular Microscopy Platform (MCMP) at the Douglas Institute.

Author contributions

C.N. conceptualized, performed experiments and wrote the manuscript. M.M. performed experiments, bioinformatics and wrote the manuscript. A.T., V.Y., M.A.D.

and Y.C.W. performed experiments and wrote the manuscript. M.S., K.P., J.-FT., S.J.T. and P.P. contributed to data analyses and reviewed the manuscript. N.M. contributed to tissue processing, data interpretation and manuscript preparation. J.R. provided technical single-cell expertise and experimental support, and aided in manuscript preparation. G.T. provided general oversight of the study, including in experimental design, data interpretation and manuscript preparation.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41593-020-0621-y>.

Correspondence and requests for materials should be addressed to G.T.

Peer review information *Nature Neuroscience* thanks Ronald Duman (deceased), Matthew Girgenti, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Corresponding author(s): Gustavo Turecki

Last updated by author(s): 2020/02/27

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

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

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

We used the following softwares for data collection: Cellranger version 2.0.1, Cellranger version 2.1.0, BD FACSDiva v.8.0

Data analysis

We used the following R packages for data analysis: Seurat (version 2.2.0, 2.3.0), mixtools (1.1.0), ape (5.1), lme4 (1.1.17), limTest (3.0.1), monocle (2.6.4), psych (1.8.4), DeconRNASeq (1.18.0), DESeq2 (1.22.2), CCIInx (0.4.0), BD FACSDiva (8.0), Cytoscape (3.7.1), DisGeNET App for Cytoscape, and custom code provided in R files along with manuscript.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Raw sequencing data, annotated gene-barcode matrix, and lists of cells used for differential gene expression analysis are accessible on GEO using the accession number GSE144136 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE144136>).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to predetermine sample size. Sample size was determined based on sample sizes used in previous similar studies.
Data exclusions	Single-nuclei were excluded from cell-type clusters based on their level of correlation to the median expression profile of the cluster (lowly correlated nuclei were removed) as outlined in the methods section to ensure that differential gene expression analysis was performed using similar nuclei populations from cases and controls. Clusters were excluded if they did not show even contribution from subjects as these clusters are likely to reflect sample specific artifacts rather than biological variability of interest. The exclusion criteria were not pre-established and were chosen based on preliminary analysis of the data.
Replication	Our design did not include a replication study of our findings in an independent sample.
Randomization	Subjects were assigned to groups based on diagnosis and not by random assignment. All subjects were male, and groups were matched for age, post mortem interval and brain pH.
Blinding	Clustering of single nuclei was performed in an unbiased blinded manner. Cluster names were assigned after generation of clusters. Clinicians were blinded for final psychiatry autopsy diagnosis of MDD case or control.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems		Methods	
n/a	Involved in the study	n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies	<input type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines	<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology	<input type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms		
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data		

Antibodies

Antibodies used	Primary: Mouse monoclonal anti-CUTL2-PerCP conjugated (dilution: 1:100, supplier: Novus, catalog no.: H00023316-M03, clone: 2H8, conjugated to PerCP using the Novus Lightning Link Labeling kit, catalog no.: 718-0010), goat polyclonal anti-SOX10 (dilution: 1:100, supplier: R&D Systems, catalog no.: AF2864), mouse monoclonal anti-NeuN-A700 (dilution: 1:300, supplier: Novus, catalog no.: NBP1-92693AF700, clone: 1B7) Secondary: Donkey polyclonal anti-goat Alexa Fluor 488 (dilution: 1:1000, supplier: JacksonImmuno, catalog no.: 705-545-147) All antibodies were purchased through Cedarlane.
-----------------	--

Validation	Novus H00023316-M03: Validated in Western blot and ELISA, used in one publication in human brain tissue (PMID: 29126813) R&D Systems AF2864: Validated in Western blot against human SOX10 protein, ELISAs, immunocytochemistry, 19 citations Novus NBP1-92693AF700: Validated in immunocytochemistry, immunohistochemistry, Western blot, one publication for flow cytometry in human brain tissue (PMID: 28750583)
------------	--

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	Age (18-87 years), post-mortem interval (12-93 hours), and brain pH (6.7-7.01) were used as covariates in statistical analysis. Samples were group based on history of major depressive disorder. All cases died during an episode of MDD.
----------------------------	---

Recruitment

Brains are recruited by the Douglas-Bell Canada Brain Bank in collaboration with the Quebec coroner's office.

Ethics oversight

Douglas Hospital Research Ethics Board

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Flow Cytometry

Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

Nuclear suspensions were prepared from 80-100 mg of post-mortem brain tissue from BA9 as described previously (PMID: 28750583) with the following modifications: homogenized tissue was centrifuged on the sucrose layer at 800g for 20 minutes at 4°C, followed by another centrifugation in nuclei extraction buffer.

Instrument

FACSAria Fusion (BD Biosciences, San Jose, CA)

Software

BD FACSDiva v.8.0

Cell population abundance

Purity of samples was determined by qPCR, as shown in supplementary materials. Approximately 50% of the nuclei were NeuN+. Of these approximately 10% were CUTL2- and 40% were CUTL2+. Among NeuN- nuclei, 30% were Sox10+, and 20% were Sox10-.

Gating strategy

Doublet discrimination was achieved by gating of Hoechst 33342 stained singlets in FSC-A vs Hoechst-A plot using 350 nm UV laser and 450/50 filter. Subsequent SOX10 positive, SOX10 negative and NeuN positive populations were gated in Alexa Fluor 700-A vs Alexa Fluor 488-A plot utilizing red 640 nm laser in combination with 730/45 filter and blue 488 laser in combination with 530/30 filter, respectively. CUTL2 positive and negative populations, the derivatives of NeuN positive gate, were defined in Alexa 488-A vs PerCP-A (blue 488 laser, 695/40 filter) plot with interval gates. CUTL2 positive population was identified as 30-40% of NeuN positive population with highest CUTL2-PerCP fluorescence. For gating of CUTL2 negative population the SOX10 negative and SOX10 positive populations were displayed in Alexa 488-A vs PerCP-A plot and the CUTL2 negative population was gated within PerCP intensities of SOX10 populations. CUTL2 negative population comprised near 10% of NeuN positive population.

- Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.