Cell type and sex specific transcriptomics of the dorsolateral prefrontal cortex in depression

Malosree Maitra

Integrated Program in Neuroscience

McGill University, Montreal

June 2023

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy.

Contents

Abstract

Major depressive disorder (MDD), one of the most prevalent psychiatric diagnoses, affects over 200 million people worldwide. MDD is a leading cause of disability globally and there are known sex differences, including higher rates of MDD in women than in men. In transcriptomic studies of the postmortem human brain and in rodent models of MDD, disease-associated gene expression changes in multiple brain regions differed between the sexes. Imbalance in excitatory and inhibitory neuronal signaling, impairments in astrocytic function, disruption of white matter, neuroinflammatory changes, and alterations in blood brain barrier integrity have been reported in MDD. Thus, the majority of the highly functionally specialized cell types of the human cortex have been implicated in this disease. To elucidate cell type and sex specific depression-associated gene expression changes we performed, for the first time, single-nucleus RNA-sequencing (snRNA-seq) postmortem in the dorsolateral prefrontal cortex, a cohort of male subjects and a cohort of female subjects, who were depressed and died by suicide, or were psychiatrically healthy. The combined dataset encompasses over 160,000 cells across 41 computationally defined cell type clusters. Differential expression analysis implicated microglia and parvalbumin interneurons in females with depression and deep layer excitatory neurons, oligodendrocyte precursor cells, and astrocytes in males with depression. Meta-analysis revealed concordant patterns of depression-associated transcriptomic changes across both sexes within cell types, despite the prominence of distinct cell types in each sex individually. These data represent the first unbiased survey of cell type and sex specific gene expression changes in depression at a single cell resolution.

Resumé

Trouble dépressif majeur (TDM), l'un des diagnostics psychiatriques les plus répandus, affecte plus de 200 millions de personnes dans le monde. Le TDM est l'une des principales causes d'invalidité à l'échelle mondiale, et il existe des différences connues entre les sexes, notamment des taux plus élevés de TDM chez les femmes que chez les hommes. Dans les études transcriptomiques du cerveau post-mortem humain et dans les modèles rongeurs de TDM, les modifications de l'expression génique associées à la maladie dans plusieurs régions cérébrales différaient entre les sexes. Un déséquilibre dans la signalisation neuronale excitatrice et inhibitrice, des altérations de la fonction astrocytaire, des perturbations de la substance blanche, des changements neuro-inflammatoires et des altérations de l'intégrité de la barrière hémato-encéphalique ont été signalés dans le TDM. Ainsi, la majorité des types cellulaires hautement spécialisés du cortex humain ont été impliqués dans cette maladie. Pour élucider les modifications de l'expression génique associées à la dépression spécifiques aux types cellulaires et aux sexes, nous avons réalisé, pour la première fois, des séquençages d'ARN à l'échelle du noyau unique (snRNA-seq) post-mortem dans le cortex préfrontal dorsolatéral, sur une cohorte d'hommes déprimés décédés par suicide ou en bonne santé sur le plan psychiatrique, et sur une cohorte de femmes déprimées décédées par suicide ou en bonne santé sur le plan psychiatrique. L'ensemble de données combinées comprend plus de 160 000 cellules réparties dans 41 clusters de types cellulaires définis par calcul. L'analyse de l'expression différentielle a impliqué les microglies et les interneurones à parvalbumine chez les femmes souffrant de dépression, ainsi que les neurones excitateurs de la couche profonde, les cellules précurseurs des oligodendrocytes et les astrocytes chez les hommes souffrant de dépression. Une méta-analyse a révélé des schémas concordants de modifications

transcriptomiques associées à la dépression dans les deux sexes au sein des types cellulaires, malgré la prédominance de types cellulaires distincts pour chaque sexe individuellement. Ces données représentent la première étude non biaisée des modifications de l'expression génique spécifiques aux types cellulaires et aux sexes dans la dépression à une résolution cellulaire unique.

Acknowledgements

I would like to express my gratitude towards my thesis supervisor, Dr. Gustavo Turecki, for always encouraging me to be ambitious and for supporting the growth of my career, my co-supervisor Dr. Corina Nagy, for her patient guidance, and Dr. Matthew Suderman, for his mentorship, over the past seven years. I am grateful for the constructive feedback of my thesis advisory committee members Dr. Guillaume Bourque and Dr. Luis Barreiro. The work presented here would not be possible without the help and support of our collaborators, the staff at the Douglas Brain Bank and the core facilities for sequencing, bioinformatics, and microscopy at Genome Quebec, McGill, and the Douglas, and most importantly the very knowledgeable and supportive researchers and bioinformaticians at the McGill Group for Suicide Studies, many of whom are my co-authors and from whom I have learned so much. Finally, I would like to acknowledge the individuals whose brain samples we have the privilege to study for our research, and their families, without whose cooperation we could not have conducted this research.

I am grateful for my family – my parents, my grandparents, my brother, and my uncle, who allowed me to become the person I am today. I am grateful for my friends and lab-mates who have believed in me and helped me to make it through the difficult times – especially Neelima and Saishree who were there with me from the beginning of this journey to the end. I am very grateful to my husband for his patience, kindness, and support without which I may not have made it till the end. Thank you all, from the bottom of my heart.

Preface to the Thesis

This thesis is presented in the manuscript-based format for Doctoral Theses, as described in the Thesis Preparation Guidelines by the Department of Graduate and Postdoctoral Studies at McGill University. The work described here was performed under the supervision of Dr. Gustavo Turecki and Dr. Corina Nagy, and the thesis contains four chapters. Chapter 1 is a review of the literature that provides the necessary background for this thesis. Chapter 2 was published in Nature Neuroscience and Chapter 3 was published in Nature Communications. Lastly, Chapter 4 summarizes the key findings from Chapters 2 and 3 in the context of previous work in the field and discusses the strengths, limitations, and future directions. Texts connecting the individual studies were written in accordance with the guidelines outlined by the McGill University Graduate and Postdoctoral Studies Office.

Contribution of authors

Chapter 1:

The first section of this chapter contains a review of the literature pertaining to sex differences in transcriptomic changes in the brain in depression and in animal models of depression and highlights some of the established contributions of major brain cell types to the disease. The second half contains a published review (https://doi.org/10.1007/s40473-019-00192-3) of single-cell and single-nucleus sequencing studies of the human brain as this methodology is pertinent to the research presented in the thesis. **The writing and literature review were conducted by the thesis author under the supervision of G.T. and C.N.**

Chapter 2:

This chapter contains published original results (https://doi.org/10.1038/s41593-020-0621-y) from single-nucleus RNA-seq in the postmortem dorsolateral prefrontal cortex (dlPFC) in male individuals with or without depression. **The thesis author performed experiments (sample preparation, in situ hybridization, microscopy) and bioinformatic analysis, and wrote the manuscript.** C.N. conceptualized the study and performed experiments (sample preparation, qPCR) and data analysis, and wrote the manuscript. A.T. and M.A.D. performed experiments (in situ hybridization, microscopy). A.T. additionally analyzed microscopy data. Y.C.W. performed snRNA-seq library preparation and high-throughput qPCR. V.Y. performed fluorescence assisted nuclei sorting. J-F.T. and K.P. analyzed data. M.S., S.J.T., and P.P. provided bioinformatics expertise and reviewed the manuscript. N.M. contributed to tissue procurement, results interpretation, and manuscript preparation. J.R. provided expertise in single-cell methodology and reviewed the manuscript. G.T. supervised the study, experimental design, data interpretation, and manuscript preparation.

Chapter 3:

This chapter contains published original results (https://doi.org/10.1038/s41467-023-38530-5) from single-nucleus RNA-seq in the postmortem dlPFC in female individuals with or without depression and a comparison to the results in males. **The thesis author performed sample preparation, conceptualized and performed bioinformatic analysis, and wrote the manuscript.** H.M. performed weighted gene co-expression network analysis. R.R. and K.Y. supported data interpretation and reviewed the manuscript. J.Y., L.F., and M-A.D. contributed to sample preparation. A.C. and Z.A. supported data analysis. D.C.M. aided in sample procurement. M.S. provided bioinformatic expertise and reviewed the manuscript. N.M. supported sample procurement and reviewed the manuscript. G.T. secured funding and supervised the study design, implementation, and manuscript preparation. C.N. conceptualized and performed snRNA-seq experiments, and supervised manuscript preparation.

Chapter 4:

This chapter contains an overall discussion and contextualization of the findings, strengths and limitations, conclusions, and future directions. **This chapter was written by the thesis author under the guidance of G.T. and C.N.**

Contribution to original knowledge

In the first study, published in Nature Neuroscience, we examined using a high-throughput, unbiased approach, the cell type specific transcriptomic differences between male individuals with depression who died by suicide and psychiatrically healthy individuals. Our dataset represents one of the first high-throughput interrogations of the human dlPFC at single-nucleus resolution and the first evaluation of depression-associated transcriptomic changes at single-nucleus resolution. We identified the most prominent changes in gene expression in a subtype of deep layer excitatory neurons and in immature oligodendrocyte precursor cells (OPCs). The data we generated was made publicly available and subsequently used by other groups to further understand cell type specificity of brain pathologies at single-nucleus resolution.

In the second study, published in Nature Communications, we performed the first snRNA-seq study of cell type specific transcriptomic changes in the dlPFC in female individuals with depression who died by suicide compared to psychiatrically healthy controls. Females are typically understudied in biomedical research and our dataset in females will be widely useful to the research community. We found that, unlike in males, the cell types with the most prominent depression-associated changes in gene expression were microglia and parvalbumin interneurons. A reanalysis of the previously generated male dataset in parallel, using the updated pipeline applied to the female dataset, confirmed the main features of our previous findings, and further allowed meta-analysis of results from both sexes. The meta-analysis revealed, within each cell type, more evidence for concordance, rather than discordance, of depression-associated transcriptomic differences between the sexes.

Prior to this work, no unbiased evaluation across all brain cell types of transcriptomic

contributions to depression had been performed and few studies had examined similarities and differences in depression-associated transcriptomic patterns in male and female individuals, primarily pointing to strikingly distinct gene expression changes in each sex. Together our studies highlight that distinct cell types are likely the prominent contributors to depression-associated transcriptomic changes in the dlPFC in males and females which could explain the discordance in transcriptomic changes reported thus far in bulk tissue studies. On the other hand, within each cell type depression-associated changes in gene expression showed evidence for similarity between the sexes. The extent and mechanisms of similarities and differences in gene expression alterations in depression between males and females across brain regions remain to be elucidated. The results we have generated, which are made publicly available for the research community, will generate further insights into the biological processes contributing to depression in each sex.

List of figures and tables outside manuscripts

List of abbreviations and acronyms outside manuscripts

ACC – anterior cingulate cortex
ACTH – adrenocorticotropic hormone
BDNF – brain-derived neurotrophic factor
CRH – corticotropin releasing hormone
CUMS – chronic unpredictable mild stress
CSDS – chronic social defeat stress
CVS – chronic variable stress
DEG – differentially expressed gene
dlPFC – dorsolateral prefrontal cortex
DMN – default mode network
ePRS – expression-based polygenic risk score
eQTL – expression quantitative trait loci
FANS – fluorescence-assisted nuclei sorting
GFAP – glial fibrillary acidic protein
GR – glucocorticoid receptor
GWAS – genome-wide association study
HPA – hypothalamic-pituitary-adrenal axis
MDD – major depressive disorder
mPFC – medial prefrontal cortex
NAcc – nucleus accumbens
OFC – orbitofrontal cortex
OL – oligodendrocyte lineage
OPC – oligodendrocyte precursor cell
PNN – perineuronal net
PTSD – post-traumatic stress disorder
PV – parvalbumin
PVN – paraventricular nucleus
SNRI – serotonin-norepinephrine reuptake inhibitor
SSRI – selective serotonin reuptake inhibitor
T-PRS – transcriptome-based polygenic risk score
WGCNA – weighted gene co-expression network analysis

Chapter 1: Introduction

Preface to Chapter 1

We start by discussing essential background information on depression, the phenotype of interest in the research presented here, and a brief overview of the major biological processes involved in MDD. We then examine some of the evidence from molecular studies in human individuals, and in animal models of stress and depression, for biological sex differences in MDD. Next, we discuss some of the current literature implicating specialized brain cell types in MDD pathology. Finally, we delve into the single-cell and single-nucleus technologies which have been employed to dissect the human brain biology and pathology in great detail in the past decade, which provide the necessary tools for our investigation of cell type specific transcriptomic contributions to MDD. We end with a clear statement of the objectives of the research presented which brings together open questions about the cell type and sex specificity of molecular changes in the brain in depression and then applies single-nucleus technology to answer these questions.

Depression

Major depressive disorder (MDD) is a psychiatric condition characterized by persistent low mood, anhedonia, feelings of guilt and worthlessness, and suicidal ideation among other symptoms[1]. MDD affects over 200 million people globally and is a leading cause of disability worldwide[2]. A diagnosis of depression increases the risk for suicide[3]. There are genetic contributions to depression, which has a modest heritability (around 30-40%), and recent genome-wide association studies (GWAS) have identified over 100 genetic loci that are associated with the disorder[4,5]. Brain imaging studies indicate region specific changes in brain activity and morphology in depression and brain region specific transcriptomic and epigenetic changes have been reported in post-mortem studies.

Depression is quite heterogenous in its presentation and there can be varying levels of symptom severity, recurrence, treatment resistance, and associated comorbidities such as anxiety disorders and substance use disorders. Moreover, the incidence of depression in higher in women than in men[2] and the symptomatology may also differ between the sexes. There is accumulating evidence that the molecular basis of depression may also differ to some extent between the sexes.

Biological mechanisms of depression

Several theories exist about the etiology of MDD. One theory which is well supported by the evidence models MDD as a dysregulation of the stress response system[6,7]. The hypothalamic-pituitary-adrenal (HPA) axis normally responds to stress by increasing secretion of corticotropin releasing hormone (CRH) from the paraventricular nucleus (PVN) of the hypothalamus which in turn leads to release of adrenocorticotropic hormone (ACTH) from the pituitary gland and subsequently the release of corticosteroids from the adrenal glands[6].

Negative feedback mechanisms acting through the glucocorticoid receptor (GR) then return corticosteroid levels back to baseline when the stressful conditions pass. However, evidence suggests that components of the system including the GR, CRH, and corticosteroids are not properly regulated in depression and in animal models of depression[6]. Interestingly, in addition to this hormonal stress response, the sympathetic nervous system's response to stress and restoration of homeostasis by the parasympathetic nervous system is also likely disrupted in depression[8,9]. Activation of the parasympathetic nervous system through the vagus nerve as a potential therapy for depression is an active area of research[9]. Related to the involvement of the HPA axis and the autonomic nervous system, disrupted modulation of inflammation has also been proposed as a mechanism for depression etiology[6,8,9].

Within the brain, imbalance in excitatory and inhibitory neurotransmission, glial cell and white matter dysfunction, neuroinflammation and blood brain barrier deficits have all been linked to MDD. The potential contributions of major brain cell types to MDD are discussed in greater detail in a subsequent section. Of note, the monoaminergic neurotransmitter systems – serotonergic, norepinephrinergic, and dopaminergic – given their role in regulating mood, emotions, and reward processing have been widely studied in depression and related animal models, investigating the monoaminergic hypothesis of depression[6,7]. In fact, selective serotonin reuptake inhibitors (SSRIs) and selective serotonin-norepinephrine reuptake inhibitors (SNRIs) are first line pharmacological treatments for depression, and are very effective for some patients, although their benefits are by no means universal. Lastly, gene-environment interactions, particularly in the context of exposure to early life stress, contribute to depression etiology, likely through epigenetic mechanisms[7,10].

Sex differences in depression

The most prominent sex difference in MDD is the higher incidence of MDD in women compared to men[11]. Further, women are more likely to experience comorbid anxiety and recurrent depression whereas men are more likely to experience comorbid substance use disorder[12] and to die by suicide[3]. Sex differences may also exist at the genetic level with sex-stratified GWAS implicating distinct loci in men and women[13].

Several studies which have attempted to connect brain imaging findings with MDD-associated molecular changes have also uncovered sex-specificity[14,15]. For example, one study found differences in the resting state functional connectivity specific to males and females when examining the components of the default mode network (DMN) as well as regions previously shown to have molecular sex differences in MDD. In particular, they found that oft reported observations of hyperconnectivity in the DMN in MDD were mostly driven by males. Utilizing genes pre-associated with depression from transcriptomic studies and relying on publicly available brain transcriptomic atlases, they further found the expression of distinct genes to be associated with male and female specific alterations in resting state functional connectivity[14]. Similarly, another group focused on structural imaging and exploring the relationship of cortical thickness, surface area, and gyrification as well as subcortical volumes, with MDD genetic susceptibility and transcriptomic patterns revealing sex specificity[15].

A systematic review from 2018 examined 20 studies published over 10 years interrogating biological sex differences in the pathophysiology of depression[16]. The authors mentioned differences in the levels of sex hormones including androgens, such as testosterone, as well as female sex hormone levels which vary during puberty, perimenopause, pregnancy and childbirth as potential contributors to sex differences in depression, but pointed to the

limitation that not many studies evaluated directly measured gonadal hormone levels making it difficult to draw conclusions[16]. Further the authors mentioned baseline and depression-related differences between males and females in the monoaminergic, GABA-ergic, and glutamatergic neurotransmission systems, in neurotrophic factors, and in pro- and anti-inflammatory cytokines as additional components factoring into MDD-associated sex differences[16]. The authors' conclusions are in line with brain molecular differences between the sexes in the stress response system both at baseline and in response to stress in humans and in animal models[17].

Sex specificity of MDD associated transcriptomic changes in the human brain

Not many studies have directly and specifically compared MDD-associated transcriptome-wide gene expression differences in the human brain between the sexes. Further, some studies have been biased towards one sex, mostly males given that the brain tissues used are often from those who died by suicide and thus less likely to include females. A widely cited and pioneering study, by Labonté and colleagues, not only identified little overlap of MDD-associated differentially expressed genes (DEGs) in males and females, but also identified overall lack of concordance in threshold free patterns of MDD-associated transcriptomic change between the sexes across several brain regions[18]. Gene co-expression module analysis supported the sex differences in MDD-associated gene expression changes[18]. These findings were further bolstered by genetically manipulated animal models which showed similar sex specific patterns as in the human cohorts[18].

Several studies re-examining the Labonté dataset focusing on long non-coding RNAs also found striking differences in MDD-associated changes in expression of lincRNAs in males and females across brain regions[19,20]. Manipulating the expression of these lincRNAs in rodent

models recapitulated depressive like behaviors in a cell type and sex specific manner[19,20]. These studies also detected striking baseline differences in lincRNA expression in the brain between the sexes[20]. Further, a transcriptomic study across several cortical regions focusing on post-traumatic stress disorder also examined a cohort of individuals with MDD[21]. The authors found considerably fewer DEGs in male MDD cases versus controls in each of the cortical regions, compared to the number of DEGs between female MDD cases versus controls in the same regions[21].

Additionally, several meta-analyses have specifically examined sex differences in gene expression differences in depression in the brain. Most of these studies have identified differences in depression associated changes between the sexes, even if not all studies report perfectly consistent findings.

A 2015 meta-analysis assessed MDD-associated gene expression differences in the human brain using eight microarray datasets from three brain regions (the anterior cingulate cortex (ACC), the dlPFC, and the amygdala) with separate datasets corresponding to male and female cohorts for each region[22]. The meta-analysis identified 566 genes associated with MDD using a lenient FDR of 25%, and 39 of these genes showed an effect of sex although not necessarily revealing opposite direction of effect between males and females[22].

A subsequent 2018 meta-analysis[23] specifically examined differential gene expression associated with MDD stratified by sex, using the same microarray datasets as analyzed by Ding et al., 2015 but with a methodological approach designed to examine sex differences. This meta-analysis found that most of the DEGs were sex specific and the DEGs in common between the sexes primarily showed opposite directions of effect[23]. Further, in threshold free

comparisons the authors observed evidence for discordance in patterns of MDD-associated gene expression between males and females[23].

Most recently, a meta-analysis examined transcriptomic differences in depression across several brain regions, combined both RNA-seq and microarray studies (same as Ding et al., 2015), and performed sex-stratified analyses[24]. This meta-analysis specifically focused on a subset of 269 genes implicated in MDD by a recent GWAS[4] and found two genes (*SPRY2* and *ITPR3*) with evidence for overall differential expression in MDD, and a handful of mutually exclusive DEGs in male (4 genes, including *ITPR3*) and female (3 genes, including *SPRY2*) specific sex-stratified analysis[24]. The study also found 6 genes with opposite direction of effect in males and females when examining sex by phenotype interactions, including *CKB* and *UBE2M* which were increased in expression in the depressed males and *MANEA* which was increased in expression in the depressed females per sex-stratified analysis. These results provide further support that in males and females, distinct genes may show differential expression related to depression in the brain.

However, it is important to note that the above meta-analysis incorporates results from Ding et al., 2015 and Labonté et al., 2017 rather than previously unexplored biological samples. Additionally, some technical caveats require consideration – the above-mentioned meta-analysis studies have combined data from cohorts which have technological differences between them – such as different microarray platforms used[22-24], and in some cases the meta-analyses combined data from separate cohorts of males and females of a necessity since suitable datasets with comparable representation of males and females are few in number.

Apart from meta-analyses, several recent reviews have discussed the molecular basis of sex differences in MDD[25,26]. These reviews have suggested a difference in the role of microglia and

excitatory neurons in depression in males and females. Seney et al., 2022, examined the evidence for the dlPFC, ACC, and basolateral amygdala (BLA) from postmortem studies, and concluded that region specific alterations in excitatory and inhibitory signaling balance involving both somatostatin and parvalbumin neurons show sex differences in MDD. Further, the authors pointed to more activated state of microglia in females at baseline compared to males and that the MDD-associated changes in microglia may be in opposite directions between the sexes[25]. Bollinger, 2021 further suggests that the difference in microglial state and dendritic spine density between the sexes in rodent models of stress may be mediated by the effect of glucocorticoids and estrogen. Moreover, Bollinger concludes that in human males with MDD neuronal transcripts are decreased and microglial transcripts are increased, whereas the opposite is true for females[26]. However, it is important to note that the research so far, especially in humans has been in bulk tissue samples, and thus the contributions of different cell types to these overall apparent sex differences remain to be established.

Sex differences in animal models of MDD

Although much of the research in animal models was restricted to males in the past, many recent studies have examined sex specific effects in rodent models of MDD. Animal models cannot recapitulate fully the complex and heterogeneous nature of MDD but different aspects of MDD, such as anhedonia or comorbid anxiety, and different causal factors, such as social isolation or early life adversity, can be captured in animal models.

Several reviews have focused on sex differences in animal models of MDD. Not all models of stress have the same effect in male and female animals. Chronic variable stress (CVS), chronic unpredictable mild stress (CUMS), and social isolation paradigms produce different responses in male and female rodents, whereas the sex-specific effects seem less pronounced for the

learned helplessness paradigm[27]. Early life stress models such as maternal separation shows effects in both males and females, whereas models such as chronic social defeat stress (CSDS) have only recently been adapted for use in female mice[27].

Mena et al., 2019 summarized the evidence from animal models for sex differences within candidate biological systems of interest in depression. Components of the HPA axis, including the GR and CRH, show evidence for sex-specific differences in response to stress at the levels of gene expression and epigenetic modulation[6]. Similarly, the brain-derived neurotrophic factor (BDNF), serotonin receptors and transporters, and GABA-ergic and glutamatergic signaling components show some evidence of sex differences in animal models of stress, and in some cases in human studies, in terms of gene expression and regulation[6]. Interestingly, in rodent models, the estrogen receptors, ESR-alpha and ESR-beta contributed to sex differences in resilience and susceptibility in stress models, and some of the effects seem to be mediated through components of the HPA axis[27].

Additional studies have further contributed to our understanding of this phenomenon. A study assessing sex differences in response to oral cortisol administration, a mouse model of stress-related disorders, using wild-type and BDNF mutant mice reported several sex differences overall and sex-specific changes in response to cortisol not only in behavior but blood cortisol levels and adrenal gland weights, reflecting sex differences in the hypothalamic-pituitary-axis function[28]. Moreover, transcriptome wide patterns of hippocampal gene expression changes related to oral cortisol administration were dependent on hippocampal sub-region, genotype, and sex[28].

Another study examining the effects of CSDS female mice compared to previous results in males, found sex-specific increases in in peripheral inflammatory markers[29]. Further, these

increases were partially reversed by a phytochemical combination therapy. Moreover, in the PFC the transcriptomic effect of combination therapy of phytochemicals, designed to decrease inflammation and modulate synaptic effects of stress, partially rescued the transcriptomic effects of CSDS in each sex[29]. However, the transcriptomic effects of CSDS stress on the PFC in each sex were distinct[29].

A recent transcriptomic study of the dorsal and ventral hippocampus in a rat model of chronic restraint stress, found concordance between the effects of stress between the two regions within each sex and a concordance between the ventral hippocampal gene expression changes in females and dorsal hippocampal gene expression changes in males[30]. Although there were some overlapping DEGs and pathways affected by stress in both sexes, most affected genes and pathways identified were sex specific[30]. The behavioral and somatic effects of stress showed sex differences. Further the authors identified sex differences, both at baseline and in stress response, in the levels of protein products related to BDNF signaling and glutamatergic neurotransmission which showed hippocampal sub-region specificity[30].

A series of studies in rodent models found sex differences in microglial states and the expression of immune function related genes at baseline and in response to acute or chronic stress[31-33]. More specifically, at baseline in females in the medial prefrontal cortex (mPFC), microglia were more shifted towards a primed state versus a ramified state compared to males, but stress exposure decreased this tendency for a more activated morphology in female microglia[31]. Similarly, in an examination of the orbitofrontal cortex (OFC), dorsal hippocampus, and the BLA, the ratio of primed to ramified microglia and the changed expression of a set of inflammatory markers showed sex and brain region specificity in response to acute or chronic stress[32]. Further, the behavioral effects of varying stress duration

(14 or 28 days) differed between male and female mice, with only males showing earlier effects at 14 days[33]. Moreover, several genes related to microglial function showed sex specific alterations in response to stress in the PFC[33].

Finally, the differential contributions of sex chromosomes and gonadal hormones to sex differences in stress response within rodent models are beginning to be dissected using the four-core genotypes mouse model where the hormonal status and sex chromosome configuration of mice are decoupled[34]. Despite the limitations of animal models in recapitulating complex human phenotypes such as depression, the ability to perform such informative genetic and molecular manipulations will strengthen our understanding of brain molecular differences between the sexes and their role in depression.

Taken together the evidence from animal models of depression, or more generally stress, suggest that there are sex differences in the behavioral response to stress, in peripheral markers of stress and inflammation, and in brain transcriptomic changes associated with stress. Moreover, these sex differences exist in interplay with gene expression differences due to brain region variations, types of stressors, genetic susceptibility, and effect of treatments.

Cell types implicated in depression

Our understanding of cell types and cell states in the human brain is evolving as the technologies and approaches for defining them evolve. Nevertheless, studies of depression and animal models of depression have supported the involvement of a wide variety of the well-established cell types of the cortex in this disease. Changes in neuronal and glial number and morphology in cortical areas, specific to cell subtypes or cortical layers, were detected in early postmortem studies of MDD[35]. More recent molecular studies have lent further support to the involvement of a broad range of cortical cell types in MDD.

Neuronal contributions

Transcriptome wide studies of depression in the human brain implicated genes related to synaptic and neuronal functions in depression. For example, an early microarray study of the postmortem dlPFC in depression identified genes related to synaptic vesicles (e.g. - *SYN1*, *CALM2*), dendritic spines (*RAB4B*) and axon growth (*TUBB4*) as showing decreased expression in MDD[36]. The results were confirmed with *in situ* hybridization, and supported by results from a rat model of depression, and pointed to GATA1 as a repressive transcription factor contributing to the decrease in expression of these genes related to neuronal function[36]. A meta-analysis of gene co-expression in 11 microarray studies of depression across several brain regions, identified a gene module overlapping GWAS-implicated MDD-genes which was enriched for glutamatergic and GABAergic gene sets as well as containing genes related to neuron growth and development (BDNF, ephrins)[37].

Among more targeted studies, a variant specific comparison of synapsin genes *SYN1*, *SYN2*, and *SYN3*, demonstrated that specific transcript variants of *SYN1* and *SYN2* were increased in expression in the PFC (BA10) in individuals with MDD, in contrast to Kang et al. (2012)[38]. The increased expression of *SYN1* was in line with an increased H3K4me3 (activating) histone modification at the *SYN1* promoter. Further, for *SYN2*, a decrease in methylation was found at several CpGs in its promoter in individuals with MDD which could epigenetically contribute to its increased expression[39].

Excitatory and inhibitory neuronal subtype contributions

Neurons are a heterogeneous population of cells. In the adult human cortex, there are not only excitatory and inhibitory neuronal populations but also great diversity within these populations. For example, excitatory neurons differ in their inputs and their projections

according to their cortical layer location and inhibitory neurons are classified into sub-categories based on their expression of neuropeptides, their morphology, and their role in the cortical microcircuitry.

Evidence from postmortem brain studies and animal models of stress suggest that there are changes in the glutamatergic neurotransmission system in the cortex[40]. Deficits in the GABAergic neurotransmission system in the cortex have also been reported in human and animal models studies, strongly implicating somatostatin expressing neurons, and to some extent parvalbumin (PV) expressing neurons[40,41]. Together, the evidence suggests an imbalance of inhibitory and excitatory signaling in the cortex in depression, which may be associated with a shift from the executive network to the DMN[41].

A targeted assessment of glutamate receptor gene expression measured a panel of 21 metabotropic and ionotropic glutamate receptors in the dlPFC in MDD individuals and controls. Five of the 21 genes, including ionotropic and metabotropic subunits, showed higher expression in cases than controls[42]. Further, a sex-stratified assessment revealed that the majority of glutamate receptor genes showed increased expression in female cases compared to female controls specifically[42]. Similarly, a screen of 200 genes related to glutamatergic function including glutamate receptors and transporters, glutamate metabolizing enzymes, and selected neurotrophic factors and signaling molecules, found four genes upregulated in MDD cases in the PFC, including an NMDAR subunit gene and a mitochondrial glutamate transporter[43].

One study examining somatostatin positive inhibitory neurons[44], using a mouse model of stress, found reductions in the number of this type of neuron in both male and female mice in response to stress, but found genes in the EIF2 signaling pathway of stress related

transcription changes were reduced in expression in males while genes related to growth factor signalling were reduced in expression in female mice. Another recent study focusing on perineuronal nets (PNNs) in the postmortem ventromedial PFC[45], found higher PNN density around neurons in deep cortical layers in depressed individuals with a history of experiencing childhood abuse. PNN intensity and the area of the neurons covered by PNNs was also increased in the same group. PNNs primarily surround PV interneurons in the human cortex and the percentage of PV interneurons surrounded by PNNs was higher in the depression individuals with a history of childhood abuse[45]. Interestingly, the expression of genes related to PNN production was increased in OPCs (*VCAN*, *TNR*, *PTPRZ1*) in the depressed individuals who had been abused in childhood. These results suggest that inhibitory PV interneurons and their surrounding extra-cellular matrix may be altered in depression, in the context of childhood abuse, and OPCs may be contributing to this process.

The above discussion barely skims the surface of the existing literature implicating various cortical neuronal subtypes in depression, but already reveals a complex molecular and cellular landscape. A complete picture of the relative contributions of the numerous neuronal types of the cortex to depression and the association of their transcriptomic state to MDD, remains to be established.

Glial contributions

It is not only neurons that have been studied or implicated in depression in the brain. Each of the major glial cell types, including astrocytes, oligodendrocytes, microglia, and endothelial cells have in turn been investigated and proposed to play a role in the brain in depression.

Astrocytes

Astrocytes are part of the “tripartite synapse” and perform essential functions in the brain

including modulating synaptic transmission, for example by uptake of neurotransmitters, and providing trophic support to neurons. Astrocytes are known to be connected to each other in a network and to communicate through the flow of calcium ions through gap junctions between the cells. Several lines of evidence suggest that astrocytes are involved in depression pathology.

Genes important for astrocytic functions, such as *GJA1* and *GJB6* which encode connexins that form gap junctions between astrocytes, have been shown to be downregulated in the PFC in suicide completers and in individuals with depression, using genome wide and targeted approaches[46,47]. The decreased expression of these genes was shown to be associated with epigenetic modulation via histone modification, specifically increased presence of the repressive mark, H3K9me3[48]. The SOX9 transcription factor, another core astrocytic gene which likely regulates the expression of the connexins, was also decreased in these studies[46,47]. The reduction of connexins was verified at the protein level[46] and in additional brain regions[48].

Further, the glial acidic fibrillary protein (GFAP), which is expressed in a subset of astrocytes and considered to be a marker for those astrocytes, was decreased in depressed individuals, both at the mRNA and protein levels in the thalamus and the caudate[49], in addition to decreased mRNA expression in the PFC[48], in depressed individuals.

Moving beyond the molecular to the cellular level, there is evidence for changes in morphology of astrocytes, with white matter astrocytes showing more elaborate branching patterns, in the ACC in individuals with depression[50]. Moreover, the density of astrocytes immunoreactive for GFAP was decreased in PFC gray matter, and in thalamus and caudate[51]. Similarly, the density of astrocytes immunoreactive for vimentin, another intermediate

filament protein similar to GFAP which is enriched in but not exclusive to astrocytes, was decreased in PFC gray and white matter and in caudate[51].

Oligodendrocyte lineage cells

An early microarray study of the temporal cortex in depressed and control individuals from the Stanley consortium (https://www.stanleyresearch.org/brain-research/) indicated a downregulation of genes related to myelin and oligodendrocyte function such as *MAG*, *MOG*, *MOBP*, and *PLP1*, and a subset of these results were validated[52]. The differentially expressed genes were not exclusively oligodendroglia related, since genes involved with synaptic function and axon growth were also implicated[52]. Nevertheless, transcriptome wide gene expression screening pointed towards oligodendroglial differences in depressed patients.

Molecular changes in the oligodendrocyte lineage, comprising oligodendrocyte precursor cells as well as myelinating and mature oligodendrocytes which are essential cellular components of the white matter, have been reported in depression related pathologies. A comprehensive study profiled DNA methylation and gene expression differences in the ACC between depressed individuals with a history of experiencing childhood abuse and psychiatrically healthy controls and found evidence both at the epigenetic and transcriptomic level for dysregulation of genes involved in myelin function, thus implicating the oligodendrocyte lineage (OL)[53]. DNA methylation differences were shown to be specific to the OL using fluorescence-assisted nuclei sorting (FANS) and gene expression changes specific to myelin genes were consistent with a rodent model of early life adversity[53]. Further, studies of the ACC white matter showed decreased density of SOX10+ OL cells in abused individuals and a decrease in axon diameter and myelin density[53].

A related study found that the presence of astrocytic connexins was decreased on

oligodendrocytic cells and on myelinated axons in depressed individuals in the ACC[54]. Further the expression of oligodendrocytic connexin genes (*GJB1* and *GJC2*) were decreased along with several genes encoding proteins that interact with and regulate connexins (*OCLDN*, *CAV1* and *CAV2*)[54]. Most of these effects were not specific to depressed individuals with childhood abuse, but applicable to the depressed group overall. Altogether the authors concluded that there may be dysregulation of gap junctions between astrocytes and oligodendrocytes, which can have a functional impact on myelin, in depressed individuals.

Further, examination of fatty acids, specifically choline glycerophospholipids, in the ACC white matter revealed increased concentration of fatty acids related to the arachidonic acid synthesis pathway specifically in depressed individuals with a history of experiencing childhood abuse, but not in depressed individuals without a history of childhood abuse[55]. These results add support for molecular differences in myelin, including lipid composition, in depression related phenotypes, further implicating the oligodendrocyte lineage cells. Moreover, the arachidonic acid synthesis pathway is involved in inflammation, a process that has been linked to depression by multiple lines of evidence, some of which are discussed below[55].

Microglia and endothelial cells – inflammation and the blood brain barrier

Many studies have examined the role of inflammation in depression, both in the periphery and in the central nervous system. While most studies indicate an involvement of inflammatory processes and molecules, both in terms of genetic variation and gene expression, the relationship does not seem to be a straightforward increase in inflammation in all individuals with depression[56]. An evaluation of the expression of seven interleukin genes in the dlPFC found evidence for decreased expression of the pro-inflammatory cytokine gene

IL1A in individuals with MDD and individuals with post-traumatic stress disorder (PTSD)[57]. On the other hand, the gene expression of *TNF-A*, another pro-inflammatory cytokine, was found to be increased in individuals who died by suicide, in individuals with MDD, and in individuals with MDD who died by suicide in the dlPFC[58]. The authors found evidence for epigenetic and post-transcriptional regulation of *TNF-A* expression[58].

While these studies were more targeted, another study performed RNA-seq in the dentate gyrus (DG) region of the hippocampus to compare MDD cases and controls, followed by specific comparisons between controls and MDD cases with recurrent or single episodes and with or without suicide, as well as sex stratified analyses[59]. The study found differences in expression of a variety of molecules related to inflammation including cytokines (*CCL2*) and molecules downstream of interferons (*ISG15*, *IFI44L*, *IFI6*) and modulators of cytokine response (*SOCS3*)[59]. Again, the pattern of these expression changes was more nuanced than a direct up or downregulation of inflammation and showed evidence for sex and depression subtype specificity.

Microglia are the resident immune cells of the brain, while endothelial cells form the capillaries of the blood brain barrier, and both cell types play an important role in regulating inflammation in the brain in response to injury and disease. Thus, the findings of dysregulation of inflammatory molecules in the brain in depression point to these cell types. The above-mentioned studies considered bulk tissues for molecular analysis. On the other hand, one study evaluated IBA1 positive microglia in the white matter of the ACC and found an increased proportion of primed to ramified microglia in depressed suicide individuals and an increase in blood vessel associated IBA1 positive macrophages[60]. Gene expression measurements, in bulk, of several microglia and macrophage markers, pro-inflammatory molecules, and cell

adhesion molecules showed increases in *IBA1*, *CD45*, *MCP-1*, *ZO-1*, in MDD[60]. Most recently, a pre-print reported gene expression differences in microglia from the occipital cortex grey matter but not the corpus callosum white matter in depressed individuals compared to controls, and evidence for differences in phagocytic pathways and immune response[61].

In mouse stress models, a recent study found sex specific and brain region specific changes in expression of endothelial cell genes such as *Cldn5*, which was decreased in female mice in the PFC, but in male mice in the nucleus accumbens (NAcc)[62]. *CLDN5* was also found to be decreased in both male and female humans with MDD in the NAcc[62], but only in females in the PFC. The authors also found sex specific transcriptome wide effects of stress paradigms on gene expression in mouse brain endothelial cells. Thus, blood brain barrier and endothelial cell molecular state is likely to be altered in MDD.

The implication of such diverse cell types in MDD and the distinct roles these cell types seem to play in the pathology of MDD necessitate more comprehensive cell type specific assessments of molecular contributions to the disease. Recent advances in single-cell and single-nucleus technologies can greatly aid in performing the requisite cell type specific studies, and are discussed in detail in the following section.

Sequencing the human brain at single-cell resolution

Malosree Maitra[1], BA; Corina Nagy[1], PhD; Gustavo Turecki[1,2*], MD, PhD

[1]McGill Group for Suicide Studies, Douglas Mental Health University Institute, [2]Department of Psychiatry, McGill University

***Correspondence to**: Dr. Gustavo Turecki (gustavo.turecki@mcgill.ca)

6875 LaSalle Blvd, Verdun, Québec, Canada H4H 1R3

Tel.: (514) 761-6131 (ext. 2370)

Keywords: human brain; single-cell RNA-seq; single-nucleus RNA-seq; post-mortem; single-cell genomics

Abstract

Purpose of review: This review summarizes some of the most recent studies on single-cell resolution sequencing of the post-mortem human brain and the application of these techniques for the study of psychiatric and neurological disorders.

Recent findings: Over the last several years, researchers have optimized single-cell transcriptome and genome sequencing in post-mortem human brain tissue. This has given us unprecedented access to cell-type specific gene expression profiles and somatic mutations unique to pathological states. Additionally, we can now measure epigenetic information from individual brain cells and with advanced statistical approaches, we can focus on the key cell-types underlying psychiatric and other brain phenotypes.

Summary: A new era of cell-type specific and single-cell resolution studies of the human brain is underway. With proper application of rapidly advancing laboratory and data analysis techniques, we should witness important advances in our understanding of molecular changes associated with psychiatric and neurological phenotypes.

Introduction

The extensive specialization and diversity of cell-types in the mammalian brain has been of great interest since the earliest days of neuroscience. Over the decades, technological advances have allowed us to interrogate various features of brain cell-types, including their morphology, electrophysiology, and molecular biology, in ever-increasing detail. Many of these techniques provide some insight into the functioning of single brain cells, such as patch-clamp recordings of individual neurons in brain slices or tracking of fluorescently labelled microglia in vivo. However, with the advent of single-cell next-generation sequencing (NGS) techniques, we are now able to study in minute detail the identity, function, and state of individual brain cells.

Studies of the human brain at single-cell resolution have become an invaluable resource for teasing apart the interactions between members of this complex cellular ecosystem. Cell-type specificity provided by single-cell approaches allows us to hone in on the individual components contributing to complex neurological and psychiatric disorders. In the context of psychiatry, there has been considerable interest in measuring cell-type specific gene expression changes associated with disease, detecting somatic mutations with single-cell genomics, and pin-pointing the cell-types which contribute most prominently to a given psychopathology.

Single-cell sequencing technologies

Currently available techniques encompass single-cell level measurements of genomes, transcriptomes, proteomes, and epigenomes, potentially combined with morphological, spatiotemporal, and electrophysiological data (Figure 1) [1,2].

Some single-cell and single-nucleus RNA-seq (scRNA-seq and snRNA-seq) technologies rely on fluorescence assisted sorting of individual cells or nuclei into the wells of a microplate [3-5] followed by preparation of libraries from the RNA extracted in each well. Many flavours of microfluidics based technologies including Drop-seq [6], inDrops [7], DroNc-seq [8], sNucDrop-seq [9], and 10X Chromium [10], among others, have also been developed for scRNA-seq and snRNA-seq. These techniques rely on microfluidic devices which encapsulate individual cells or nuclei in oil droplets as the suspension is passed through the device. Each cell or nucleus is tagged with a unique barcode and mRNA molecules are captured by poly-A priming. Library preparation is performed in bulk and after sequencing, the reads are demultiplexed using barcodes, while unique molecular identifiers (UMIs) are used to generate counts of different RNA molecules originating from individual nuclei. Plate-based methods are generally low-throughput but allow for full-length cDNA sequencing whereas droplet-based methods can be ultra-high throughput (scalable up to millions of cells) but suffer from higher dropout rates and 3' biased reads.

A suite of single-cell level NGS approaches have also been developed utilizing combinatorial indexing for measuring transcription [11,12], DNA methylation [13], chromatin states [14-16], and even multiple modalities of information [17]. These approaches are based on split-and-pool strategies and involve sorting and tagging individual cells or nuclei with barcodes, followed by sequential pooling and attachment of additional barcodes. The numbers of barcodes used and cells pooled are adjusted such that the probability of multiple cells receiving the same combination of barcodes is very low. These methods are less expensive and may require less specialized equipment than droplet-based methods, but they can result

in higher multiplet rates, when multiple cells or nuclei are tagged by the same barcode combination.

For profiling DNA sequence variations at the single-cell level, most studies thus far have relied on sorting of individual nuclei into the wells of a microplate. However, recently, commercial technology for droplet-based high-throughput single-nucleus genome sequencing has been created which holds promise for detecting both copy-number variations (CNVs) and single-nucleotide variations (SNVs) at single-cell resolution [18].

Droplet-based profiling of open chromatin in single-cells has also been developed [19]. Furthermore, several techniques for acquiring multimodal data at single-cell resolution exist, such as CITE-seq which measures cell-surface epitopes and transcriptomes [20] or Patch-seq which measures electrophysiological recordings and transcriptomes [21].

Thus, a wide variety of sequencing technologies are available for single-cell resolution studies of complex tissues. However, not all technologies are equally applicable to human brain tissue, especially archived, frozen, post-mortem brain tissue which is of primary interest when studying psychopathology. For example, scRNA-seq is at best extremely challenging, and often impossible, for frozen post-mortem brain tissue because of the difficulties of extracting intact cells from such tissue, especially neurons which have extensive and fragile processes. Fortunately, several studies have indicated that identification of cell-types based on single-nucleus gene expression profiles is comparable to single-cell transcriptomic profiles, although the RNA content of the whole cell and the nucleus are not identical [22,23]. In this review, we briefly touch upon single-cell sequencing studies of the mouse brain but focus

mainly on studies that have harnessed NGS to examine gene expression (Table 1) and genomic variation (Table 2) at the single-cell level in the human brain.

Single-cell and single-nucleus sequencing of the mouse brain

Given that rodent models of psychopathology and neurological disease are a mainstay of modern neuroscience, in the past few years a plethora of studies have been published on single-cell and single-nucleus sequencing of the mouse brain. Many different brain regions have been profiled by scRNA-seq including cortical regions [8,24-26], subcortical structures [27], and the hippocampus [8,24,28,29]. Different stages in brain development have been investigated in detail [30-32] and specific cell-types have been targeted, such as oligodendrocytes [33,34] or microglia [35-37]. In many experiments, cell-type specific signatures of a variety of experimental perturbations have been measured [26,29,38].

In fact, single-cell transcriptomics has recently produced several large scale atlases of mouse brain cellular diversity [12,39-41]. Moreover, single-cell resolution ATAC-seq studies of several mouse brain regions have recently been published [14-16]. The explosion of single-cell and single-nucleus sequencing studies of the mouse brain has previously been reviewed in detail elsewhere [42-45].

Single-cell and single-nucleus transcriptomics of the healthy human brain

Darmanis et al. (2015) performed one of the first scRNA-seq studies of human brain in surgically excised non-pathological tissue from the adult human temporal cortex of 8 epilepsy patients and on developing brain samples from 16 to 18-week old fetuses. They profiled gene expression in over 400 individual cells using Fluidigm C1 chips [5]. Unbiased (i.e. unsupervised) and biased (i.e. supervised) clustering approaches, now commonplace in scRNA-seq, and cell-

type annotation, identified excitatory and inhibitory neurons as well as major glial cell-types. Soon after, Krishnaswami et al. (2016) established one of the first protocols for snRNA-seq on human post-mortem brain tissue. They used fluorescence assisted nuclei sorting (FANS) to place individual neuronal nuclei, selected based on NeuN expression, into the wells of a microplate and used a SMART-seq approach for generating libraries [3]. Their work demonstrated on a small-scale the feasibility of snRNA-seq in archived post-mortem tissues, such as those accessible through brain banks.

The same year, Lake et al. (2016) performed Fluidigm C1 chip capture of NeuN positive FAN sorted neuronal nuclei to profile more than 3000 neurons from various cortical regions of a single healthy subject. They identified numerous inhibitory and excitatory neuronal subtypes which were in broad agreement with the cell-types described by Darmanis et al. (2015), but revealed finer subtypes powered by the larger dataset, such as layer-specific and region-specific excitatory neuron subtypes [46].

Habib et al. (2017), created an snRNA-seq method they termed DroNc-seq and applied it to several post-mortem human prefrontal cortex (PFC) and hippocampus samples, in addition to mouse brain tissue. DroNc-seq incorporated several adjustments to the Drop-seq [6] protocol, including an alteration to the dimensions of the microfluidic device to allow for capture of nuclei, which are smaller than cells, and inclusion of intronic reads in analyses due to the preponderance of pre-mRNA in the nucleus. They were the first to demonstrate the feasibility of droplet-based high-throughput snRNA-seq in archived post-mortem human brain tissue [8]. Furthermore, there was good correspondence in cell-types between the mouse and human datasets and with the findings of Lake et al. (2016).

Lake et al. (2018) independently designed an adaptation of Drop-seq for snRNA-seq. Their modifications included heat-based lysis of nuclei and incorporation of intronic reads, similar to Habib et al. (2017). Moreover, they performed an assay for single-nucleus chromatin accessibility based on combinatorial barcoding, in addition to snRNA-seq, and used their snRNA-seq findings to refine clustering of single-nuclei based on chromatin accessibility [47]. Although tissue was obtained from healthy subjects, the cell-type specific chromatin accessibility information generated was used to indirectly assess cell-type involvement in neurological and psychiatric diseases.

While most studies have focused on the cortex, Welch et al. (2019) performed high-throughput snRNA-seq on more than 40,000 nuclei derived from archived substantia nigra samples from 7 healthy donors. They developed and applied a single-cell data analysis tool called LIGER for aligning the data from multiple individuals into a consolidated dataset. Clustering driven by inter-individual variability is a recurring problem in snRNA-seq datasets, and LIGER was able to mitigate this effect. They identified the expected subtypes of glial and neuronal cells, including dopaminergic neurons. Moreover, they were able to pin-point subject specific effects: including activation of microglia in one subject who experienced traumatic brain injury (TBI) at the time of death, and distinct microglial and astrocytic signatures in another subject with histological signs of amyloid deposits upon post-mortem examination [48]. Thus, not only will their dataset serve as a reference for future snRNA-seq studies of the substantia nigra, but their software, which can effectively combine results from multiple datasets for joint analysis without losing dataset-specific components of the information, will be widely applicable in future single-cell sequencing studies.

While the massive capacity of high-throughput snRNA-seq is enticing, some questions require a more targeted approach as exemplified by a recent human brain snRNA-seq study from the Allen Institute [49]. Boldog et al., (2018) identified a new subtype of inhibitory neuron, dubbed the rosehip neuron, which seems to be uniquely found in the human cortex. The information from snRNA-seq was complemented by morphological and electrophysiological data from surgical tissue as well as corroborated with fluorescent *in situ* hybridization (ISH). The data from this study is part of a larger human mid-temporal gyrus dataset [50], generated by the Allen Institute from both post-mortem samples and surgical tissue and it provides an excellent resource for benchmarking data produced by high-throughput platforms. Uniquely, this dataset accounted for cortical layer location during the dissection and extraction of nuclei.

Single-nucleus transcriptomic studies of human brain pathology

Over the past year, a slew of single-cell sequencing studies of the human brain has exploited the rapidly developing technology to ask questions about cell-type diversity and cell-type specific gene expression changes in pathological states, including autism spectrum disorders (ASD) [51-53], Alzheimer's disease (AD) [54], multiple sclerosis (MS) [55], and depression [56].

Among studies focusing on ASD, Renthal et al. (2018) performed snRNA-seq using the inDrops approach in a mouse model of Rett syndrome and in the post-mortem occipital cortex of Rett syndrome patients. Rett syndrome is an X-linked developmental disorder in the autism spectrum. Affected females are heterozygous carriers of causal mutations in the *MECP2* gene. Since one X-chromosome carries the mutation and the other does not, random X-chromosome inactivation in individual cells results in mosaic expression of the mutant *MECP2*

allele in the brain. Using an innovative analytical approach, this group was able to utilize 3' biased snRNA-seq reads to separate cells containing an active *MECP2* mutation causal for Rett, from cells in which the mutant allele was not expressed [51]. This was achieved by genotyping the Rett syndrome patients to identify SNPs in the 3' region of genes near the *MECP2* locus, in linkage disequilibrium with the *MECP2* mutation carried by the patient. These SNPs could then be detected by 3' snRNA-seq, thus allowing the authors to determine whether the mutant or healthy *MECP2* allele was expressed in a particular cell [51]. Their approach allowed direct comparison of gene expression within specific cell-types between cells expressing mutated versus normal *MECP2,* revealing similarities in the patterns of differential gene expression (DGE) in the human patients and the mouse model.

Further applying single-cell technology to study ASD, Velmeshev et al. (2019) performed high-throughput snRNA-seq of the PFC and anterior cingulate cortex (ACC) in individuals with ASD, epilepsy, or no pathology. Unbiased identification of cell-types across the brain regions identified the major cortical cell-types and revealed an over-representation of protoplasmic astrocytes in ASD subjects. Cell-type specific DGE analysis revealed over 500 differentially expressed genes (DEGs) which were in good agreement with previous literature on ASD-associated genes [52]. The DEGs in upper layer neurons and in microglia were the best predictors for clinical autism severity. Since several of the ASD subjects also experienced seizures, the authors performed cell-type specific DGE analysis by snRNA-seq in the PFC of matched sporadic epilepsy patients, to tease apart the contributions of seizures and ASD. Only a small proportion of DEGs in epilepsy overlapped with the ASD findings, suggestion that most of the cell-type specific DEGs were specific to ASD.

Similar to the previous study, Sorrells et al. (2019) performed snRNA-seq post-mortem on the amygdala from 8 individuals between the ages of 4 and 15 years, both neurotypical controls and ASD patients. Their paper focused on identifying a subset of neurons within the paralaminar nucleus (PL) of the human amygdala which show protracted development and retain molecular and morphological features of immature neurons well into adulthood [53]. Their snRNA-seq experiment complemented the ISH and immunohistochemistry findings. Among more than 13,000 nuclei sequenced, they were in fact able to identify a small population marked by high expression of *DCX*, *BCL2*, *NR2F2,* and *ROBO1*, characteristic of the immature PL neurons identified using other techniques. Moreover, snRNA-seq allowed them to detect additional genes that were enriched in this immature neuronal population, namely *ST8SIA2*, *SOX11*, and *MAP2*. Finally, they were able to compare gene expression between ASD cases and controls in this immature PL neuron cluster and identify around 30 DEGs.

To explore Alzheimer's disease at the single-cell level, Mathys et al. (2019) performed high-throughput snRNA-seq on the PFC of 48 individuals from the ROSMAP [57] cohort. They clustered cells into the major neuronal and non-neuronal cell-types of the PFC and measured DGE between individuals with detectable Alzheimer's pathology and individuals without pathology. Over a thousand DEGs were identified, the majority of which were downregulated, with the largest contribution from excitatory neurons. A small subset of DEGs were validated in NeuN positive and negative populations separated by FANS and by ISH [54]. They also examined progressive changes in gene expression with increase in pathological burden by grouping individuals according to clinico-pathological measures. Changes in gene expression were more cell-type specific between individuals with no pathology versus early pathology, compared to individuals with early pathology versus late pathology. Sub-clustering of the

major cell-types identified specific sub-clusters with over-representation of cells from pathological or healthy states. Interestingly, some AD-associated sub-clusters had an over-representation of female cells. Further exploration of sex-differences revealed a global pathology-associated upregulation of genes in oligodendrocytes in males and a global pathology-associated downregulation of genes in neurons in females. An interesting feature of this dataset is the advanced age of the donors (> 70 years).

In a landmark study of oligodendroglial heterogeneity in the human brain, Jäkel et al. (2019) performed snRNA-seq on white matter obtained post-mortem from multiple sclerosis (MS) patients and controls. In addition to neurons and other glia, they identified many different oligodendrocyte and oligodendrocyte precursor (OPC) clusters, including an oligodendrocyte cluster with immune features. Characteristic gene expression features of these oligodendroglial clusters were verified by ISH and immunohistochemistry [55]. Pseudotime analysis indicated that end state oligodendrocytes most highly express genes involved in cell-signalling and adhesion, whereas myelinating oligodendrocytes most highly express genes involved in myelination. On combined clustering of control and MS datasets, there was a higher representation of immune cells, including macrophages, derived from MS tissue, indicating immune infiltration. Overall, the same cell-types were present in MS versus control white matter, but OPCs and intermediate oligodendrocytes were under-represented in MS and the distribution of nuclei among the mature oligodendrocyte clusters was also altered. DGE analysis revealed increased expression of genes for myelination in multiple oligodendrocytic clusters in MS, as well as distinct changes in gene expression between lesioned, non-lesioned, and remyelinating regions of white matter from patients.

In our lab, we sequenced around 80,000 nuclei from the PFC of 17 individuals who were depressed and died by suicide and 17 individuals who were psychiatrically healthy. The 26 quality-controlled cell-type clusters we identified were in good correspondence with the cell-types identified by Habib et al. (2017). We were able to detect cluster-specific DEGs associated with depression in more than half of these cell-types [56]. Many of these genes have previously been implicated in bulk gene expression studies of MDD.

Single-cell resolution studies of sequence variation in the human brain

The potential impact of somatic mutations, including SNVs and CNVs, which may accumulate in the post mitotic cells of the central nervous system over the course of development and aging, has long intrigued neuroscientists in the context of neurological and psychiatric conditions [58]. Experimental techniques exist for studying somatic mutations and their contributions to brain disorders using tissue homogenates, but single-nucleus genome sequencing has an undeniable advantage in this context. Although limitations, such as errors introduced during whole genome amplification (WGA) and the astronomical cost of whole genome sequencing (WGS) for large numbers of cells, still remain to be addressed, several groups have succeeded in sequencing the genomes of single-cells from post-mortem brain tissue and the results have yielded some intriguing insights.

McConnell et al. (2013) measured CNVs in single neurons from post-mortem human frontal cortex and induced pluripotent stem cell (iPSC) derived neurons using WGA followed by DNA microarrays or sequencing. Cultured neurons had higher incidence of CNVs compared to NPCs or fibroblasts, suggesting that accumulation of somatic mutations may be integral to the development of neuronal identity. Non-germline small and large CNVs distributed

throughout the genome were detected in > 40% of the brain-derived neurons, although only a small number of neurons showed extensive CNV burden [59].

Cai et al. (2014) measured CNVs in post-mortem brain tissue from 3 healthy individuals and 1 subject with hemimegaencephaly (HMG) and established that aneuploidy is rare but sub-chromosomal CNVs are common. They confirmed an expected CNV at chromosome 1q in the HMG brain, although they identified a tetrasomy rather than the predicted trisomy [60]. Some CNVs were shared by multiple neurons, providing evidence that they are not artefacts of the technology.

LINE1 (L1) retrotransposon insertion, a subtype of CNV, is of special interest in psychiatry. These mobile DNA elements are capable of "'jumping" in the genome, i.e. inserting a copy of themselves into a new part of the genome via an RNA intermediate, and are thought to be especially active during neurogenesis. Most of the newly formed L1 insertions in the genome are not-capable of jumping but they can create variation in the genomes of individual neurons, even within the same individual, which may have gene regulatory consequences [58]. Moreover, studying somatic L1 retrotransposition with single-cell genomics is of interest in psychiatric research because changes in the rates of L1 retrotransposition have been linked to schizophrenia [61,62], autism spectrum disorders [63], and major depressive disorder [64].

Evrony et al. (2012) examined the rates of L1 retrotransposition in 300 neuronal nuclei from the cortex and caudate nucleus in three neurologically healthy individuals by WGA, L1 insertion profiling (L1 IP), and sequencing. They detected hundreds of known and tens of novel L1 insertions in these single-nucleus genomes, but on average each neuron had less

than one somatic L1 insertion, suggesting that such insertions are generally rare [65]. In contrast, Upton et al. (2015) reported much higher rates of somatic L1 insertions in hippocampal neurons (~13 on average) and glia (~6 on average) and cortical neurons (~16 on average) using single-cell retrotransposon capture sequencing (RC-seq). Somatic L1 insertions were identified based on their absence in bulk tissue RC-seq with brain and liver samples from the same individuals and seemed to be enriched in hippocampally transcribed genes in both neurons and glia from the hippocampus [66]. However, Evrony et al. (2016) later reanalysed these data and estimated that the true rates were closer to less than one L1 somatic insertion per cell [67], more consistent with their earlier paper.

In a 2015 study (Lodato et al.), 36 neurons from the cortex of three individuals were sequenced to detect somatic SNVs. On average, 1500 L1 variants were identified and a considerable proportion was found in transcriptionally active neuronal gene regions. Interestingly, certain SNVs seem to be caused by deamination of methyl cytosines to thymines, suggesting that they were produced post-mitotically rather than during DNA replication in development [68]. Using one subject to detect patterns of shared somatic mutations diverging over time, Lodato et al., were able to trace the developmental lineage of a subset of neurons, identifying clades of related neurons. Of note, some of the more frequently detected brain SNVs were present in non-brain tissue, indicating that they arose early in development.

In a follow-up study, Lodato et al. (2018) measured SNVs in the hippocampus and PFC of 24 individuals with ages spanning 4 months to 82 years. Nine subjects were diagnosed with either Cockayne syndrome (CS) or xeroderma pigmentosa (XP), neurodegenerative diseases caused by deficiencies in the DNA repair mechanism, and the remaining 15 were free of

pathology. In line with previous studies, somatic SNVs were found to accumulate with age and be enriched in neuronally expressed genes. However, SNVs seemed to accumulate at a higher rate in the hippocampus [69]. As expected, somatic SNVs were more frequent in CS and XP subjects than in controls.

Most recently, a study looking at 1000 cells from brain-healthy individuals found, on average, that neurons harbour more CNVs than non-neuronal or non-neural cells (Chronister et al., 2019). Furthermore, these CNVs tended to affect a larger portion of the genomes [70]. Neuronally expressed transcripts are generally longer and are reported to possess more somatic mutations within neurons [68]. However, in contrast to the reported neuronal increase of somatic SNVs overtime, this study [69] found a decreased prevalence with age of neurons with CNVs in their genomes. The authors suggest that cells with more CNVs may be more susceptible to aging-related loss.

Single-nucleus methylomics in the post-mortem human brain

Whole-genome bisulphite sequencing (WGBS) of single cells from archived tissue is extremely challenging as bisulphite-conversion leads to loss of material and is very limiting when starting with extremely small amounts of DNA derived from single cells. Furthermore, the process is expensive as each cell needs to be sequenced at sufficient coverage. Nevertheless, Luo et al. (2017) performed single-nucleus WGBS and produced a single-cell resolution map of DNA methylation in the human frontal cortex. FAN sorting into microplates, followed by bisulphite-conversion and sequencing, produced single-nucleus DNA methylation profiles for almost 3000 nuclei from the frontal cortex of a single subject. Despite data sparsity, clustering of cells using single-nucleus DNA-methylation signatures resulted in

separation of the cortical excitatory and inhibitory neuronal subtypes with a resolution comparable to snRNA-seq [71]. Non-CG methylation was found to be more cell-type specific than CG methylation and overall patterns of cell-type specific methylation were highly conserved from mouse brain to human brain. This dataset is a valuable reference for cell-type specific DNA methylation in the human brain and enables deconvolution of bulk DNA methylation data to estimate constituent cell-types.

Insights, challenges, and future applications of human brain single-cell sequencing

Some limitations of snRNA-seq in the human brain may be inherent to the technology or the underlying biology, such as underrepresentation of glial cells [8,47,52] and consistently lower numbers of RNA molecules detected in glial cells compared to neurons [5,8,47]. Other limitations may be overcome using computational methods such as imputation for addressing high gene dropout and sparse data [72-74] and dataset alignment algorithms for addressing inter-individual variability [48,75].

As our knowledge of the strengths and limitations of these techniques increases, so does our ability to better design experiments. Using cryosections from histological dissections for extracting nuclei can ensure more even input from different microanatomical regions [47,52]. Combining two subjects, differing in sex or in known SNVs, for nuclei capture on a microfluidic device, followed by deconvolution using sex-specific genes or based on known SNVs, can help account for technical variability between captures [76].

In addition to direct identification of cell-types and comparison of cell-type specific features between biological groups, different modalities of single-cell data can indirectly inform our understanding of disease states. Deconvolution algorithms [77,78] can elucidate cell-

type contributions to observed disease-related changes in gene expression or DNA methylation in bulk tissue studies. Findings from genome-wide association studies (GWAS) and bulk gene expression studies can in turn help pin-point disease-relevant cell-types from single-cell datasets [79-81]. Future work will likely involve integration of multimodal data [16,48,75], and use of complementary approaches other than NGS, such as high-throughput ISH [82,83], for single-cell resolution studies.

Conclusion

Promising initiatives such as the Human Cell Atlas [84] and the Brain Somatic Mosaicism Network [85] are currently underway, and should greatly enhance our understanding of diversity in the transcriptome and genome of individual cells in the brain. The findings will contribute to furthering our knowledge of complex diseases which affect the brain. The continued rapid advancement of single-cell technology is creating ample opportunities for applying these technologies to the study of the human brain in health and disease. Soon we can hope to unravel the intricate complexity of the multitude of cell-types that compose the human brain and to better explain how they contribute to the development of disease.

Human and Animal Rights

This article does not contain any studies with human or animal subjects performed by any of the authors.

Figures

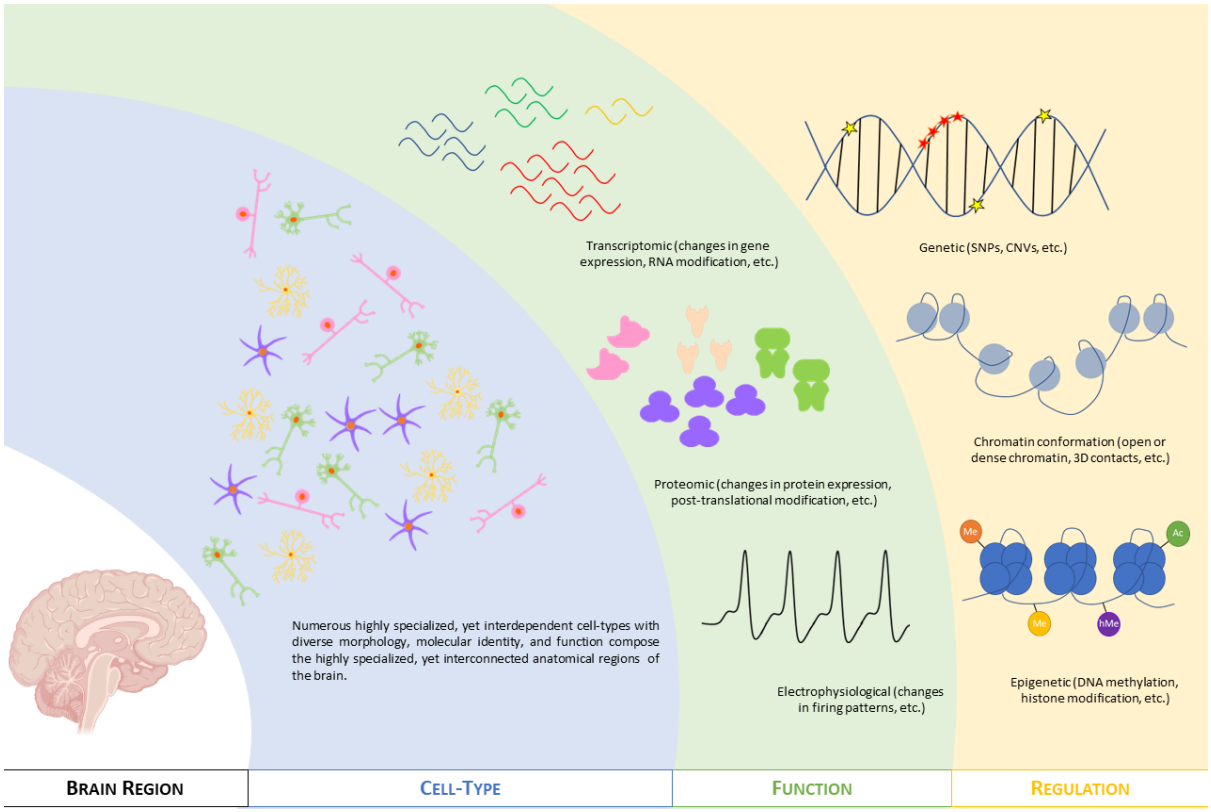


Figure 1: The human brain is composed of billions of individual cells which belong to an amazingly diverse array of cell-types. These cells and their mutual interactions give rise to the myriad complex functions of the brain, as well as dysfunctions such as psychiatric diseases. Each cell in turn is defined by its genome, epigenome, transcriptome, proteome, and other measurable properties. How do the specific states, functions, and interactions of a billion cells relate to the state and functioning of an entire human brain? A rapidly evolving suite of technologies which apply next-generation sequencing at the single-cell level may help us find the answer. We are now able to measure changes in the genomic sequences, gene expression, epigenetic modifications, chromatin states and sometimes combinations of these variables within individual cells of diverse cell-types in the brains of healthy and diseased individuals. The detailed picture of cell-types and states produced by these datasets will aid us in teasing

apart the specific and complementary roles of diverse brain-cell types in psychiatric and neurological diseases.

Tables

Table 1: Recent single-cell and single-nucleus RNA-sequencing studies of the human brain

Publication	Subjects	Sample type	Brain regions	PMI/RIN	Through-put	Technique	Goal of study
Darmanis et al., 2015	8 adult epilepsy patients, 4 fetal (16-18 week)	Surgical	Temporal cortex, fetal cortex	NA	400-500 cells	Fluidigm C1 chips, full length cDNA (SMART-seq)	Identify major cell types in adult human cortical tissue at single-cell resolution
Krishnaswami et al., 2016	NA	Post-mortem	Prefrontal cortex – neurons and non-neurons	RIN>=7, PMI up to several hours	10-30 cells	FANS, plate-based, full-length cDNA (SMART-seq)	To establish feasibility of snRNA-seq from post-mortem archived human brain
Lake et al., 2016	1 subject, 51 yr old female	Post-mortem	Cortical regions (BA41, 17, 10, 8, 22, 21) – only neurons	PMI 22 hrs	3000-4000 cells	FANS, Fluidigm C1 chips, full length cDNA (SMART-seq)	Establish snRNA-seq in human brain, characterize cell types
Habib et al., 2017	5 subjects, 40-65 yrs, male	Post-mortem	Hippocampus & prefrontal cortex, neurons & glia	RIN > 6.9, PMI average 12.4 hrs	15,000 cells	Modified Drop-seq, poly-A capture, 3' sequencing	Establish droplet-based snRNA-seq in human brain, characterize cell types
Lake et al., 2018	6 subjects, 20-49 yrs, males and females	Post-mortem	Cortex & cerebellum, neurons & glia	PMI < 24 hours	35,000 cells	Modified Drop-seq, ploy-A capture, 3' sequencing	Characterize different cell types, relate to regions of open chromatin
Renthal et al., 2018	3 subjects, 8-24 years, females, Rett syndrome	Post-mortem	Occipital cortex	PMI < 16 hours	30,000 cells	inDrops	Comparing cells expressing mutated *MECP2* to cells expressing normal *MECP2*
Boldog et al., 2018	2 subjects, 50 & 54 yrs old, male	Post-mortem	Mid-temporal gyrus, cortical layer 1	RIN>=7, PMI 24 hours	800-900 cells	FANS, plate-based, SMART-seq, full-length cDNA	Identification of novel inhibitory neuron subtypes
Jäkel et al., 2019	5 controls, 4 MS, males and females	Post-mortem	White matter areas	NA	17,000 cells	10X Genomics Chromium	Characterizing oligodendrocyte heterogeneity in MS and healthy white matter

Mathys et al., 2019	24 controls, 24 AD pathology, balanced males and females	Post-mortem	Prefrontal cortex	NA	80,000 cells	10X Genomics Chromium	Cell-type specific differential gene expression in Alzheimer's disease
Velmeshev et al., 2019	15 ASD, 8 epilepsy, 23 controls, 4-54 yrs	Post-mortem	Prefrontal and anterior cingulate cortex	3-42 hours	120,000 cells	10X Genomics Chromium	Cell-type specific differential gene expression in ASD and epilepsy
Sorrells et al., 2019	4 ASD, 4 control, 4-15 yrs	Post-mortem	Amygdala	3-27 hours	13000 cells	10X Genomics Chromium	Identification of immature or newly formed neurons
Welch et al., 2019	5 males, 3 females, controls, 18-75 yrs	Post-mortem	Substantia nigra	< 24 hours	44,000 cells	10X Genomics Chromium	Cell-type identification, cross donor and cross species comparisons

MS – multiple sclerosis, ASD – autism spectrum disorders

Table 2: Recent and single-nucleus whole-genome sequencing studies of the human brain

Publication	Subjects	Sample type	Brain regions	Throughput	Technique	Goal of study
Evrony et al., 2012	3 controls, 1 fetus, 1 HMG	Surgical and post-mortem	Cortex and caudate nucleus, only neurons	300 cells	FACS, MDA, L1 IP sequencing	Assess somatic L1 retrotransposon variability in healthy brains
McConnell et al., 2013	3 subjects, 20-26 yrs, 1 male	Post-mortem	Frontal cortex, only neurons	>100 cells	FACS, GenomePlex, sequencing	Detecting somatic CNVs in human brain neurons
Cai et al., 2014	2 controls, 1 fetal, 1 HMG	Post-mortem	Cortex, neurons and non-neurons	>200 cells	FACS, GenomePlex or MDA, sequencing	Detecting somatic CNVs in normal and HMG brain
Upton et al., 2015	4 subjects, 1 AGS individual	Post-mortem	Hippocampus and cortex, neurons and glia	100-200 cells	FACS, MALBAC, RC-seq	Assess somatic L1 retrotransposon variability in healthy brains
Lodato et al., 2015	3 subjects, 15-42 yrs	Post-mortem	PFC, neurons	~250 cells	FACS, MDA, sequencing	Assess somatic SNVs in single neurons from a healthy brain
Lodato et al., 2018	24 subjects, 4-83 years, 9 subjects CS or XP	Post-mortem	Cortex and hippocampus, only neurons	~160 cells	FACS, MDS, sequencing	Assess effect of age and neurodegeneration on somatic SNVs
Chronister et al., 2019	5 subjects, 4 months - 95 years	Port-mortem	Frontal cortex, neurons and non-neurons	>800 cells	PicoPlex, sequencing	Assess somatic CNV diversity of the human brain cells with age

Notes: FACS – fluorescence assisted cell sorting, MDA – multiple displacement amplification, AGS – Aicardi-Goutières Syndrome, MALBAC – Multiple Annealing and Looping Based Amplification Cycles

References

1. Stuart, T. & Satija, R. Integrative single-cell analysis. *Nature Reviews Genetics* (2019).
2. Kulkarni, A., Anderson, A.G., Merullo, D.P. & Konopka, G. Beyond bulk: a review of single cell transcriptomics methodologies and applications. *Current Opinion in Biotechnology* **58**, 129-136 (2019).
3. Krishnaswami, S.R.*, et al.* Using single nuclei for RNA-seq to capture the transcriptome of postmortem neurons. *Nature protocols* **11**, 499-524 (2016).
4. Johnson, M.B.*, et al.* Single-cell analysis reveals transcriptional heterogeneity of neural progenitors in human cortex. *Nature Neuroscience* **18**, 637 (2015).
5. Darmanis, S.*, et al.* A survey of human brain transcriptome diversity at the single cell level. *Proceedings of the National Academy of Sciences* **112**, 7285-7290 (2015).
6. Macosko, Evan Z.*, et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202-1214 (2015).
7. Zilionis, R.*, et al.* Single-cell barcoding and sequencing using droplet microfluidics. *Nature protocols* **12**, 44 (2016).
8. Habib, N.*, et al.* Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nature Methods* **14**, 955 (2017).
9. Hu, P.*, et al.* Dissecting Cell-Type Composition and Activity-Dependent Transcriptional State in Mammalian Brains by Massively Parallel Single-Nucleus RNA-Seq. *Molecular Cell* **68**, 1006-1015.e1007 (2017).
10. Zheng, G.X.Y.*, et al.* Massively parallel digital transcriptional profiling of single cells. **8**, 14049 (2017).
11. Cao, J.*, et al.* Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* **357**, 661-667 (2017).
12. Rosenberg, A.B.*, et al.* Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science* **360**, 176-182 (2018).
13. Mulqueen, R.M.*, et al.* Highly scalable generation of DNA methylation profiles in single cells. *Nature Biotechnology* **36**, 428 (2018).
14. Cusanovich, D.A.*, et al.* A Single-Cell Atlas of In Vivo Mammalian Chromatin Accessibility. *Cell* **174**, 1309-1324.e1318 (2018).
15. Preissl, S.*, et al.* Single-nucleus analysis of accessible chromatin in developing mouse forebrain reveals cell-type-specific transcriptional regulation. *Nature Neuroscience* **21**, 432-439 (2018).
16. Sinnamon, J.R.*, et al.* The accessible chromatin landscape of the murine hippocampus at single-cell resolution. *Genome research* **29**, 857-869 (2019).
17. Cao, J.*, et al.* Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science* **361**, 1380 (2018).
18. Pellegrino, M.*, et al.* High-throughput single-cell DNA sequencing of acute myeloid leukemia tumors with droplet microfluidics. *Genome research* **28**, 1345-1352 (2018).
19. Lareau, C.A.*, et al.* Droplet-based combinatorial indexing for massive-scale single-cell chromatin accessibility. *Nature Biotechnology* (2019).
20. Stoeckius, M.*, et al.* Simultaneous epitope and transcriptome measurement in single cells. *Nature Methods* **14**, 865 (2017).
21. Cadwell, C.R.*, et al.* Electrophysiological, transcriptomic and morphologic profiling of single neurons using Patch-seq. *Nature Biotechnology* **34**, 199 (2015).

22. 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. *Scientific reports* **7**, 6031 (2017).
23. Grindberg, R.V.*, et al.* RNA-sequencing from single nuclei. *Proceedings of the National Academy of Sciences* **110**, 19802-19807 (2013).
24. Zeisel, A.*, et al.* Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**, 1138-1142 (2015).
25. Tasic, B.*, et al.* Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nature Neuroscience* **19**, 335 (2016).
26. Hrvatin, S.*, et al.* Single-cell analysis of experience-dependent transcriptomic states in the mouse visual cortex. *Nat Neurosci* **21**, 120-129 (2018).
27. Kalish, B.T.*, et al.* Single-cell transcriptomics of the developing lateral geniculate nucleus reveals insights into circuit assembly and refinement. *Proceedings of the National Academy of Sciences* **115**, E1051 (2018).
28. Habib, N.*, et al.* Div-Seq: Single-nucleus RNA-Seq reveals dynamics of rare adult newborn neurons. *Science* **353**, 925 (2016).
29. Arneson, D.*, et al.* Single cell molecular alterations reveal target cells and pathways of concussive brain injury. *Nature Communications* **9**, 3894 (2018).
30. Hochgerner, H., Zeisel, A., Lönnerberg, P. & Linnarsson, S. Conserved properties of dentate gyrus neurogenesis across postnatal development revealed by single-cell RNA sequencing. *Nature Neuroscience* **21**, 290-299 (2018).
31. Tiklová, K.*, et al.* Single-cell RNA sequencing reveals midbrain dopamine neuron diversity emerging during mouse brain development. *Nature Communications* **10**, 581 (2019).
32. La Manno, G.*, et al.* Molecular Diversity of Midbrain Development in Mouse, Human, and Stem Cells. *Cell* **167**, 566-580.e519 (2016).
33. Marques, S.*, et al.* Oligodendrocyte heterogeneity in the mouse juvenile and adult central nervous system. *Science* **352**, 1326-1329 (2016).
34. Marques, S.*, et al.* Single-cell transcriptomic profiling of progenitors of the oligodendrocyte lineage reveals transcriptional convergence during development. *bioRxiv* (2017).
35. Mathys, H.*, et al.* Temporal Tracking of Microglia Activation in Neurodegeneration at Single-Cell Resolution. *Cell Reports* **21**, 366-380 (2017).
36. Hammond, T.R.*, et al.* Single-Cell RNA Sequencing of Microglia throughout the Mouse Lifespan and in the Injured Brain Reveals Complex Cell-State Changes. *Immunity* **50**, 253-271.e256 (2019).
37. Sousa, C.*, et al.* Single-cell transcriptomics reveals distinct inflammation-induced microglia signatures. *EMBO reports* **19**, e46171 (2018).
38. Lacar, B.*, et al.* Nuclear RNA-seq of single neurons reveals molecular signatures of activation. *Nature Communications* **7**, 11022 (2016).
39. Zeisel, A.*, et al.* Molecular Architecture of the Mouse Nervous System. *Cell* **174**, 999-1014.e1022 (2018).
40. Saunders, A.*, et al.* Molecular Diversity and Specializations among the Cells of the Adult Mouse Brain. *Cell* **174**, 1015-1030.e1016 (2018).
41. Häring, M.*, et al.* Neuronal atlas of the dorsal horn defines its architecture and links sensory input to transcriptional cell types. *Nature Neuroscience* **21**, 869-880 (2018).

42. Cuevas-Diaz Duran, R., Wei, H. & Wu, J.Q. Single-cell RNA-sequencing of the brain. *Clinical and Translational Medicine* **6**, 20 (2017).
43. Ofengeim, D., Giagtzoglou, N., Huh, D., Zou, C. & Yuan, J. Single-Cell RNA Sequencing: Unraveling the Brain One Cell at a Time. *Trends in Molecular Medicine* **23**, 563-576 (2017).
44. Guillaumet-Adkins, A. & Heyn, H. Single-Cell Genomics Unravels Brain Cell-Type Complexity. in *Neuroepigenomics in Aging and Disease* (ed. Delgado-Morales, R.) 393-407 (Springer International Publishing, Cham, 2017).
45. Cembrowski, M.S. Single-cell transcriptomics as a framework and roadmap for understanding the brain. *Journal of Neuroscience Methods* **326**, 108353 (2019).
46. Lake, B.B.*, et al.* Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science* **352**, 1586-1590 (2016).
47. 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).
48. Welch, J.D.*, et al.* Single-Cell Multi-omic Integration Compares and Contrasts Features of Brain Cell Identity. *Cell* **177**, 1873-1887.e1817 (2019).
49. Boldog, E.*, et al.* Transcriptomic and morphophysiological evidence for a specialized human cortical GABAergic cell type. *Nature Neuroscience* **21**, 1185-1195 (2018).
50. Hodge, R.D.*, et al.* Conserved cell types with divergent features between human and mouse cortex. *bioRxiv*, 384826 (2018).
51. Renthal, W.*, et al.* Characterization of human mosaic Rett syndrome brain tissue by single-nucleus RNA sequencing. *Nature Neuroscience* **21**, 1670-1679 (2018).
52. Velmeshev, D.*, et al.* Single-cell genomics identifies cell type–specific molecular changes in autism. *Science* **364**, 685 (2019).
53. Sorrells, S.F.*, et al.* Immature excitatory neurons develop during adolescence in the human amygdala. *Nature Communications* **10**, 2748 (2019).
54. Mathys, H.*, et al.* Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332-337 (2019).
55. Jäkel, S.*, et al.* Altered human oligodendrocyte heterogeneity in multiple sclerosis. *Nature* **566**, 543-547 (2019).
56. Nagy, C.*, et al.* Single-nucleus RNA sequencing shows convergent evidence from different cell types for altered synaptic plasticity in major depressive disorder. *bioRxiv* (2018).
57. Bennett, D.A.*, et al.* Religious Orders Study and Rush Memory and Aging Project. *Journal of Alzheimer's disease : JAD* **64**, S161-s189 (2018).
58. Singer, T., McConnell, M.J., Marchetto, M.C.N., Coufal, N.G. & Gage, F.H. LINE-1 retrotransposons: mediators of somatic variation in neuronal genomes? *Trends in Neurosciences* **33**, 345-354 (2010).
59. McConnell, M.J.*, et al.* Mosaic Copy Number Variation in Human Neurons. *Science* **342**, 632 (2013).
60. Cai, X.*, et al.* Single-Cell, Genome-wide Sequencing Identifies Clonal Somatic Copy-Number Variation in the Human Brain. *Cell Reports* **8**, 1280-1289 (2014).
61. Doyle, G.A.*, et al.* Analysis of LINE-1 Elements in DNA from Postmortem Brains of Individuals with Schizophrenia. *Neuropsychopharmacology* **42**, 2602 (2017).

62. Guffanti, G.*, et al.* LINE1 insertions as a genomic risk factor for schizophrenia: Preliminary evidence from an affected family. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics* **171**, 534-545 (2016).
63. Shpyleva, S., Melnyk, S., Pavliv, O., Pogribny, I. & Jill James, S. Overexpression of LINE-1 Retrotransposons in Autism Brain. *Molecular Neurobiology* **55**, 1740-1749 (2018).
64. Liu, S.*, et al.* Inverse changes in L1 retrotransposons between blood and brain in major depressive disorder. *Scientific reports* **6**, 37530 (2016).
65. Evrony, Gilad D.*, et al.* Single-Neuron Sequencing Analysis of L1 Retrotransposition and Somatic Mutation in the Human Brain. *Cell* **151**, 483-496 (2012).
66. Upton, Kyle R.*, et al.* Ubiquitous L1 Mosaicism in Hippocampal Neurons. *Cell* **161**, 228-239 (2015).
67. Evrony, G.D., Lee, E., Park, P.J. & Walsh, C.A. Resolving rates of mutation in the brain using single-neuron genomics. *eLife* **5**, e12966 (2016).
68. Lodato, M.A.*, et al.* Somatic mutation in single human neurons tracks developmental and transcriptional history. *Science* **350**, 94 (2015).
69. Lodato, M.A.*, et al.* Aging and neurodegeneration are associated with increased mutations in single human neurons. *Science* **359**, 555 (2018).
70. Chronister, W.D.*, et al.* Neurons with Complex Karyotypes Are Rare in Aged Human Neocortex. *Cell Reports* **26**, 825-835.e827 (2019).
71. Luo, C.*, et al.* Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex. *Science* **357**, 600-604 (2017).
72. van Dijk, D.*, et al.* Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. *Cell* **174**, 716-729.e727 (2018).
73. Li, W.V. & Li, J.J. An accurate and robust imputation method scImpute for single-cell RNA-seq data. *Nature Communications* **9**, 997 (2018).
74. Huang, M.*, et al.* SAVER: gene expression recovery for single-cell RNA sequencing. *Nature Methods* **15**, 539-542 (2018).
75. Stuart, T.*, et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e1821 (2019).
76. Skelly, D.A.*, et al.* Single-Cell Transcriptional Profiling Reveals Cellular Diversity and Intercommunication in the Mouse Heart. *Cell Reports* **22**, 600-610 (2018).
77. Gong, T. & Szustakowski, J.D. DeconRNASeq: a statistical framework for deconvolution of heterogeneous tissue samples based on mRNA-Seq data. *Bioinformatics (Oxford, England)* **29**, 1083-1085 (2013).
78. Newman, A.M.*, et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nature Methods* **12**, 453 (2015).
79. Skene, N.G.*, et al.* Genetic identification of brain cell types underlying schizophrenia. *Nature Genetics* **50**, 825-833 (2018).
80. Skene, N.G. & Grant, S.G.N. Identification of Vulnerable Cell Types in Major Brain Disorders Using Single Cell Transcriptomes and Expression Weighted Cell Type Enrichment. *Frontiers in Neuroscience* **10**(2016).
81. Calderon, D.*, et al.* Inferring Relevant Cell Types for Complex Traits by Using Single-Cell Gene Expression. *American journal of human genetics* **101**, 686-699 (2017).
82. Codeluppi, S.*, et al.* Spatial organization of the somatosensory cortex revealed by osmFISH. *Nature Methods* **15**, 932-935 (2018).

83. Moffitt, J.R.*, et al.* Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science* **362**, eaau5324 (2018).
84. Regev, A.*, et al.* The Human Cell Atlas. *eLife* **6**, e27041 (2017).
85. McConnell, M.J.*, et al.* Intersection of diverse neuronal genomes and neuropsychiatric disease: The Brain Somatic Mosaicism Network. *Science* **356**, eaal1641 (2017).

Recent advances in single-cell sequencing of the human brain

Since the advent of high-throughput snRNA-seq interrogation of human postmortem brain in disease contexts in 2019[63-65] the technologies have further evolved, and the field has progressed in leaps and bounds. Multimodal profiling of the transcriptome and epigenome (open-chromatin regions) has become commonplace as has spatial gene expression profiling in tissue sections at near single-cell resolution. Moreover, profiling histone modifications at a single-nucleus level at a high-throughput has become feasible[66] and, most recently, spatially resolved simultaneous measurement of the transcriptome and epigenome in tissue sections has been achieved[67]. Keeping pace with the development of experimental techniques, analysis methods have developed at a rapid pace. From cell filtering[68] and sample multiplexing[69] to multiplet removal[70], from batch correction[71] to clustering assessment[72,73], and frameworks for facilitating comparison between phenotypes[74,75] or linking findings in genetics to single-cell sequencing results[76], the possibilities are numerous. Add to that the specialized tools for analysis of spatial[77] or epigenetic datasets[78] or for combining spatially resolved data with single-nucleus data, and we have a formidable arsenal of methods at our disposal to analytically dissect the identity, state, function, and dysfunction of individual cells in the human brain. Interestingly, new suites of tools have begun to explore the potential ligand receptor interactions between cell types while also accounting for the spatial distance between them[79,80]. Identifying not only cell types relevant to a phenotype, but the disrupted cell-cell communication between them, could provide promising therapeutic avenues for future research. Figure 1 below summarizes some of the key steps in

handling snRNA-seq data and Table 1 presents a sampling of tools that can be used for achieving each processing step.

Many scRNA-seq and snRNA-seq studies of postmortem tissue in disease contexts have added to our understanding of several different neurological and psychiatric phenotypes over the past few years. Researchers have extensively explored the molecular underpinnings of Alzheimer's disease at single-nucleus resolution[81,82]. Studies have interrogated postmortem tissue in Huntington's disease[83], Parkinson's disease[84] and schizophrenia[85,86] using single-nucleus techniques. Several of these studies have combined snRNA-seq with snATAC-seq[81] or spatial transcriptomics[87]. Additionally, outside of disease contexts, our understanding of brain cell types across brain regions, development, and evolution has steadily increased[88,89]. Further, targeted studies have been addressing the earlier drawbacks of single-nucleus sequencing methods for profiling glial populations[90,91]. In summary, the field has made tremendous progress in the past few years and this growth trend seems likely to continue in the near future bringing not just more exciting technologies and impressive datasets, but also more importantly, deeper biological insights to the research community.

Figures

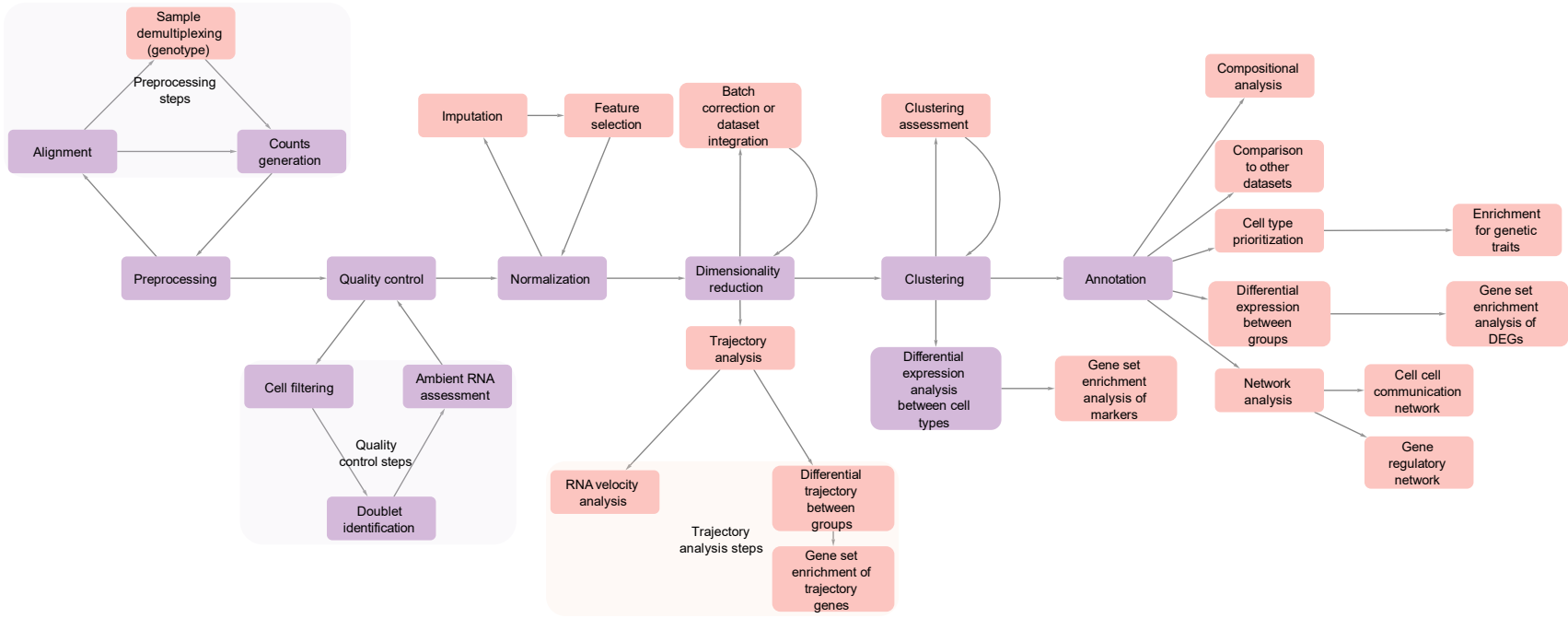


Figure 1: Single-cell and single-nucleus RNA-seq data analysis overview. Required steps are shown in purple and optional steps are shown in pink.

Tables

Table 1: A partial list of recently published tools for scRNA-seq and snRNA-seq data analysis. A more comprehensive list[92], which is continually growing, is available here: https://www.scrna-tools.org/

Step	Tools
Clustering assessment	clustree, scclusteval
Gene set enrichment of trajectory genes	enrichR, fGSEA, ClusterProfiler
RNA velocity analysis	scVelo, velocyto
Differential trajectory between groups	tradeSeq
Counts generation	CellRanger, alevin-fry, kallisto-bustools, DropSeq tools
Alignment	CellRanger, alevin-fry, kallisto-bustools, DropSeq tools
Enrichment for genetic traits	EWCE, scDRS
Cell type prioritization	Augur
Trajectory analysis	monocle3, slingshot
Gene set enrichment analysis of markers	enrichR, fGSEA, ClusterProfiler
Gene regulatory network	SCENIC, SCENIC+, hdWGCNA
Cell-cell communication network	CellChat, CellPhoneDB, iTALK, LIANA
Network analysis	Various depending on type
Gene set enrichment analysis of DEGs	enrichR, fGSEA, ClusterProfiler
Differential expression between groups	muscat, Libra, edgeR, limma, DESeq2

Compositional analysis	propeller, scCODA
Comparison to other datasets	MetaNeighbor, FR-Match
Differential expression analysis between cell types	MAST, Wilcoxon tests, presto
Annotation	scType, BRETIGEA, BrainInABlender, Azimuth, scArches
Clustering	Louvain, Leiden, k-nearest neighbors
Batch correction or dataset integration	Seurat integration, Harmony, LIGER, Scanorama, ComBat
Dimensionality reduction	PCA, UMAP, tSNE
Feature selection	Seurat vst, marker based
Imputation	MAGIC
Normalization	Log normalization, scran, sctransform
Quality control	SampleQC
Preprocessing	Various depending on substeps
Sample demultiplexing (genotype)	CellSNP, demuxafy

Objectives

As outlined in the above review of the relevant literature, the extent of similarity or differences in depression-associated molecular changes in the brain between males and females is still under investigation. The evidence from animal models and human studies suggests sex specificity in the molecular basis of depression but the information available is far from comprehensive. Molecular and cellular studies have implicated disruptions in different classes of brain cell types, both neuronal and glial, in depression pathophysiology. The relative importance of these cell type specific contributions is as yet unknown and previous studies have examined specific target cell types rather than screening, at the molecular level, all cell types of the brain. Finally, limited research has examined the sex specificity of cell types implicated in depression.

Thus, the research presented in this thesis sought to explore the cell type and sex specificity of molecular differences in the brain between individuals with depression and individuals without, focusing on gene expression. We leveraged novel snRNA-seq technology, the strengths and limitations of which have been discussed above, to attempt to answer the research question. Our objectives were to profile cell type specific gene expression in the postmortem human dlPFC across all brain cell types with snRNA-seq, to identify cell types with differentially expressed genes between individuals with depression and those without, and to compare whether the cell types and genes implicated are similar or different in males and females.

Chapter 2

Preface to Chapter 2

There is ample molecular evidence for the contribution of both neuronal cell types – including excitatory and inhibitory neurons, and glial cell types – including astrocytes, oligodendrocyte lineage cells, and microglia, to MDD-associated changes in the human brain and in animal models of MDD. With the advent of methods for profiling gene expression in individual nuclei from the human brain in an unbiased, high-throughput manner, we set out to assess differences associated with depression across all cortical cell types in the dlPFC. First, we optimized an approach to extract nuclei for snRNA-seq from archived post-mortem brain samples (Appendix 1). Then, to apply this very novel snRNA-seq technology within the constraints of ensuring sufficient power and managing experimental costs, we initially only profiled a cohort of male individuals with or without depression, thus simplifying statistical analysis and including more individuals per group than would be possible with a sex-stratified cohort. While studies examining a single sex, of a necessity, provide an incomplete picture, they can serve as a starting point for understanding the sex specific contributions to a condition with established sex differences, such as depression. Our findings are presented below.

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

Corina Nagy[1,2], Malosree Maitra[1,2,3], Arnaud Tanti[2], Matthew Suderman[6,7], Jean-Francois Théroux[2], Maria Antonietta Davoli[2], Kelly Perlman[2,3], Volodymyr Yerko[2], Yu Chang Wang[4,5] Shreejoy Tripathy[8,9], Paul Pavlidis[8], Naguib Mechawar[2,3], Jiannis Ragoussis[4,5], Gustavo Turecki[2,3,4]*

[1]These authors contributed equally

[2]McGill Group for Suicide Studies, Douglas Mental Health University Institute, [3]Department of Psychiatry, McGill University [4]Department of Human Genetics, McGill University, [5]McGill University and Genome Quebec Innovation Centre, [6]MRC Integrative Epidemiology Unit, University of Bristol, UK, [7]Population health Sciences, Bristol Medical School, Bristol, UK.[8]Michael Smith Laboratories and Department of Psychiatry, University of British Columbia, Vancouver, BC, Canada. [9]Present address: Krembil Centre for Neuroinformatics, Centre for Addiction and Mental Health, Toronto ON, Canada and Department of Psychiatry, University of Toronto, Toronto ON, Canada.

***Correspondence to**: Dr. Gustavo Turecki (gustavo.turecki@mcgill.ca)

6875 LaSalle Blvd, Verdun, Québec, Canada H4H 1R3

Tel.: (514) 761-6131 (ext. 2370)

Abstract: 142
Words: 4,462
Figures: 6
Tables:1

Abstract

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 post-mortem brain tissue, obscuring 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 healthy controls (n=17). We identified 26 cellular clusters, and over 60% of these showed differential gene expression between groups. We found the greatest dysregulation in deep layer excitatory neurons and immature oligodendrocyte precursor cells (OPCs), contributing 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.

Major depressive disorder (MDD) is a complex and heterogeneous disorder that affects an estimated 300 million people worldwide[1]. Genetic factors underlying the risk for MDD have been investigated using including genome-wide association studies, among other approaches[2]. Although some genetic associations have been detected, it remains a challenge to extract causal disease mechanisms from these findings [3]. It has been positing that MDD results from dysregulation of monoaminergic transmission, largely implicating the serotonergic and noradrenergic systems, 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 integrity[6], and inflammation[9]. Given the wide variety of cell types in the brain and their complex interactions, investigative approaches with 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 even 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 cell type composition between samples[12] Recently developed techniques for high-throughput single-cell and single-nucleus RNA-sequencing provide a solution for addressing this inherent drawback to bulk tissue experiments[11, 13].

High-throughput droplet-based single-nucleus RNA-sequencing (snRNA-seq) allows the profiling of thousands of nuclear transcriptomes, by utilizing nucleus-specific barcodes and unique molecular identifiers (UMI) to tag individual RNA molecules. snRNA-seq yields comparable, albeit distinct, information[14] from single-cell RNA-seq (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 has yet 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. These results point to gene expression changes in predominantly two cell types, oligodendrocyte precursor cells and deep layer excitatory neurons. The relationships between and functions of the differentially expressed genes from these two cell clusters suggest impairments to FGF signalling, steroid hormone receptor cycling, immune function, 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 major depressive disorder[18]. 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, 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 have consistently been found to 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 (see Methods, Supplementary Fig. 1a-e, Supplementary Table 4). In an initial subset of 20 subjects, applying our custom filtering increased the total number of cells 1.8–fold but increased the number of non-neuronal cells by almost 6-fold (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 2144 in neurons to 1144 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), median UMI count per nucleus (t test p=0.14), and number of cells detected per individual (t test=0.07) (Supplementary Table 1).

Identification of 26 distinct cell types in the dlPFC

In order to identify different cell types present in the brain samples, we applied unsupervised graph-based clustering[19] using the first 50 principal components derived from the 2135 most variable genes across individual nuclei (Methods, 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, oligodendrocyte precursor cells (OPCs), endothelial cells, and microglia (see Methods for full list of markers, Supplementary Table 6, Supplementary Fig. 3a-p). Gene expression patterns specific to cell type clusters were visualised using a DotPlot (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 are consistent with those previously reported in snRNA-seq of human PFC (Supplementary Fig. 5)[11]. Gene expression patterns previously linked to specific cortical layers (see Methods) coincide with our clustering of excitatory cells. In Fig. 2c, the genes are arranged from top to bottom in order of their expression across the cortical layers (first 17 rows, from the layer I/II to layer VI). There is a gradient of expression of these genes across the excitatory clusters. For example, clusters Ex1, Ex4, and Ex7-9 had high expression of *TLE4* (layer VI specific). 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 layer V specific, 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 demonstrated subtype specific gene expression patterns. For example, In7 was classified as inhibitory-parvalbumin because it expressed *GAD1* and *PVALB*, and lacked *VIP* and *SST* (Fig. 2d). Multiple astrocytic clusters were also identified, and while the typical sub-classification of astrocytes is based on their morphology within grey or white matter[20], we used only grey 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 astrocytes[21] (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, whereas other lineage markers like*, OLIG2* or *SOX10*, are present in both mature and immature cells. Given these developmental stage specific markers it was possible to plot a pseudotime trajectory[22] using gene expression for OPC1, OPC2, Oligos1, Oligos2 and Oligos3. Our result 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 cases and controls (Supplementary Fig. 6a). However, among the genes exclusively

associated with pseudotime in cases, there was a 2.7–fold enrichment of apoptosis signalling in PANTHER[23] pathway analysis (FDR $p<9.01x10^{-3}$), while no enrichment was observed in controls. Given that certain stages of oligodendrocyte differentiation are associated with heightened susceptibility to apoptosis, this may indicate differences in OL development between cases and controls[24]. To assess the individual profiles of important developmental gene markers, we plotted their expression across pseudotime (Supplementary Fig. 6b-i), revealing their expected pattern of expression.

To compare our oligodendrocyte lineage (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 from Jäkel et al. (2019)[25]. The OPC1 profile also primarily corresponded to the OPCs, but consistent with this cluster being further along the pseudotime trajectory, it showed a small correspondence to the COPs (committed oligodendrocyte precursors). Our oligodendrocyte clusters showed varying degrees of correspondence to the published data, with decreasing overlap to the published "OPCs" expression profile with increasing maturity of the cell type (ranging from 70-11% correspondence). Interestingly, among our oligodendrocytes, Oligos3 showed the highest correspondence to the ImOlGs (immune oligodendroglia), as defined by Jäkel et al[25]. The "immune gene expression" feature of Oligos3 is highlighted in our hierarchical clustering dendrogram (Fig. 1b), in which Oligos3 is located closer to the Micro/Macro cluster compared to the other OL clusters.

Cell type-specific patterns of altered gene expression in MDD

We set out to assess gene expression differences between cases and controls within each cluster. However, one limitation of droplet based single-nucleus technology is the possibility of capturing doublet or multiplet nuclei, which we have estimated to be 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. Using only these purified clusters and filtered genes (median 5212 per cluster), we performed a differential gene expression analysis (Supplementary Tables 8-31).

A total of 96 genes (FDR <0.10) were differentially expressed in 16 of the 25 clusters analyzed (Fig. 3a) and 45 of those remained significant at FDR<0.05 (12 of 25 clusters). FDR correction considering all clusters together yields 41 significant genes (FDR < 0.10) in 16 clusters (Supplementary Table 32).This further supports that our statistical analyses are in fact able to detect differences in gene expression between the groups. To retain a larger set of genes in order to better capture functional enrichments within individual cell types, we considered all genes which passed FDR < 0.10, corrected per cluster. The majority, 83% (80 genes), were downregulated in line with findings from previous transcriptomic studies in MDD[3, 4]. Differential

expression analysis treated each cell as a sample (Supplementary Fig. 8), but per subject contributions were visualized using heatmaps of average gene expression to assess biases in subject contributions. Patterns of gene expression averaged by subject reflected the expected differences between cases and controls (Supplementary Fig. 9a-p). Thirty-nine of the 96 differentially expressed genes were found in excitatory cell clusters and, of those, 34 were downregulated (Fig. 3a, insert). Some neuronal clusters contained both upregulated and downregulated genes, but it was more common for affected neuronal clusters to contain only downregulated genes (8/12, 67%). All but one inhibitory cluster showed altered gene expression and non-neuronal clusters tended to have both up- and downregulated genes (Fig. 3b).

Of particular interest, two clusters – one composed of immature oligodendrocyte precursor cells (OPC2) and one composed of deep layer excitatory neurons (Ex7) – accounted for almost half (47%) of the dysregulated genes (Fig. 3c). Finally, two genes were differentially expressed in more than one cluster: *PRKAR1B* showed decreased expression in excitatory clusters Ex7 (FDR=0.087, 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 GWAS of MDD[2, 26].To further compare our DEGs with previously reported findings in MDD we took advantage of publically available databases PsyGeNET[27] and DisGeNET[28]. Using PsyGeNET we found that 26 of our DEGs have previously been linked to mental illness in the literature. The highest number of associations (22/54 associations) were for depressive disorders, followed by associations for

schizophrenia spectrum and other psychotic disorders (20/54; Fig. 3d). Using DisGeNET we found 15 genes associated with MDD related terms (hypergeometric test, p-value = 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 ($p=5.7x10^{-4}$, 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 Table 33-35.

Functional implications of cell type specific DEGs

We used Gene Ontology and Reactome Pathway enrichment analysis 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 are 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 Pathways enrichments included *Kinesins* (21.74-fold enrichment; FDR = $6.24x10^{-4}$), *HSP90 chaperone cycle for steroid hormone receptors* (15.79-fold enrichment; FDR = $3.4x10^{-2}$), and *Innate Immune System* (3.01-fold enrichment, FDR=$3.29x10^{-2}$). A full list of all enrichment analyses performed is provided in Supplementary Table 36-41.

The majority (excluding three*: AC133680.1, MEG3, FAM66C*) of the DEGs were protein-coding. We used STRING network analysis [29] to plot the interactions between these proteins 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-value = 3.64×10^{-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 cells types (In2 VIP and In3 SST respectively), belong to both the kinesin pathway and cytoskeletal function (Fig. 4c). Of note, *KAZN*, a gene previously associated with MDD[26], interacts with the *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 sparsity of snRNA-seq data, we performed WGCNA on the average gene expression profile for each subject across all cell types and included the percentage contribution of different cell types as a correlate. Our results indicated that 5 modules were significantly associated with MDD (Supplementary Table 42).

Four of the 5 modules were also strongly associated with Ex7, representing the highest cluster-

phenotype overlap. We chose to focus on the largest module (blue), which included 2699 genes and significantly overlapped with our identified DEGs (Fig. 5a, 44%, p-value = $6.04x10^{-19}$, hypergeometric test for overlap). To identify the most connected genes within the blue module, we performed a hub gene analysis resulting in 285 hub genes (Fig.5b, see 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=$7.21\ x10^{-3}$), suggesting a disruption of intercellular communication between neural cells. Furthermore, we found that 26 of the 41 DEGs that overlapped with the blue module were also hub genes (p-value = 4.95×10^{-31}, hypergeometric test for overlap).

Validation of gene expression changes

We preformed validation of our DEGs using fluorescence-assisted nuclei sorting (FANS) to separate broad cell types followed by high-throughput 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 Fig. 10-11, Supplementary tables 43-46). Figure 5 (d) highlights validated genes that overlap with the WGCNA results.

Intercommunication between lower layer excitatory neurons and oligodendrocyte precursor cells

Next, in order 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, Supplementary Table 47a) and 32 OPC2 ligand to Ex7 receptor interactions were altered between cases and controls (Fig. 6a right, Supplementary Table 47b). We found significant changes to FGF signalling 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 particularly, changes in FGF signalling 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 indicating the class of protein and change in expression of the gene (Fig. 6b). To add support to the model we selected genes to further study with RNAScope® fluorescence *in situ* hybridization. Given the important change in FGF signalling we chose to investigate *FIBP* (FGF1 intercellular binding protein), *KAZN* a potential junction protein and HPS90AA1 a co-chaperone involved in stress hormone receptor cycling. We found *FIBP* was downregulated, as expected, in deep layer excitatory neurons (Fig. 6c, Unpaired t test, $t_{217}=2.5$, $p=0.013$, n= 95 nuclei for cases and controls) while *KAZN* was upregulated in OPCs (Fig. 6d, Unpaired t test, $t_{188}=2.7$, $p=0.007$, n=100 nuclei for controls, n=119 nuclei for cases) and *HSP90AA1* was downregulated, also in OPCs (Fig. 6e, Unpaired t test, $t_{192}=2.0$, $p=0.026$, n= 107 nuclei for controls, n= 87 nuclei for cases).

Discussion

Our examination of single-nucleus transcriptomes from the dlPFC in MDD revealed dysregulation of gene expression in almost 60% of the cell types identified, with a total of 96 differentially expressed genes. There were prominent gene expression changes in immature oligodendrocyte precursor cells (OPC2) and in deep layer excitatory neurons (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 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 interneurons[18] 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 supporting the complex interplay between multiple cell types in MDD. Previous studies have shown that SST and PVALB interneurons are dysregulated in MDD patients[18], and here we report several DEGs in 3 interneuron clusters that are 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 find differentially expressed genes 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 10 different excitatory cell types which 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 disorder[33].

OPC2 also showed extensive gene expression changes between cases and controls. OPC2 was the youngest cell type in the OL pseudotime trajectory. The use of cellular deconvolution techniques indicated that OPC1 have some similarity to committed OPCs whereas OPC2 showed no such correspondence, supporting the idea of functional heterogeneity among OPCs [34]. Furthermore, compared to OPC1, OPC2 expressed higher levels of certain glutamate and sodium receptors, which are typically lost as the cells mature [34].

Evidence suggests that half of the OPCs ($NG2^+$) in the brain do not give rise to any other cell type[35], and exhibit synaptic contact with neurons [36]. As such, OPCs are now thought to be a distinct glial cell type implicated in brain plasticity through roles such as integration of synaptic

activity[37] and mediation of long term potentiation[38]. Additionally, there is evidence directly implicating the loss of this cell type with emergence of depressive-like behaviour[31]. The data from this study support a role for OPCS in MDD independent from their role as precursor cells for oligodendrocytes.

STRING DB protein network analysis highlighted a number of links including connections between three differentially expressed genes encoding kinesin-related proteins: *KIF26B, KLC2* and *KIF16B*. *KIF16B* (increased in Ex7) is involved in recycling receptors including the fibroblast growth factor receptor (FGFR). Interestingly, *FIBP,* encoding acidic FGF1 intracellular-binding protein, was decreased in Ex7. FGFR transport relies, in part, on the interaction between kinesins and Rab GTPases [39]. Notably, we found *RAB11B* (encoding a Rab GTPase) and *KLC2* to be downregulated in In3. Taken together, these findings could point to a disruption of FGFR recycling by kinesins and Rab GTPases, as well as disrupted modulation FGF intercellular signalling by *FIBP* in neurons in MDD.

Based on animal models and in cell culture, FGFs (specifically FGF2) and FGFRs seem to be affected by stress and the glucocorticoids[40]. The glucocorticoid receptor (GR) has consistently been implicated in MDD [41]. *HSP90AA1* (decreased in OPC2) and *FKBP4* (decreased in Ex7), along with its homolog *FKPB5*, encode cochaperones for the GR and regulate intracellular signalling functions of this receptor[30]. *HSP90AA1* codes for the stress inducible isoform HSP90α and interestingly, is known to be secreted in certain stress contexts [42]. 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 signalling.

The genes related to chaperone mediated steroid hormone receptor 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 implications in the plastic properties of excitatory neurons such as projection outgrowth and stability [43, 44].

Additionally, genes such as *PRNP* (the prion protein gene) and *KAZN* (a gene involved in desmosome assembly), were strongly altered in the OPC2 cluster and are associated with mediating synaptic plasticity and cellular communication[45, 46] . The absence of *Prnp* has been associated with an increased number of undifferentiated oligodendrocytes and the delayed expression of differentiation markers[47], which is intriguing given the evidence implicating a lack of mature adult oligodendrocytes in animal models of depression and anxiety[48]. On the other hand, overexpression of kazrin in keratinocytes profoundly changed cell shape, reduced filamentous actin, and impaired assembly of intercellular junctions [46]. Interestingly, decreased desmosome length has been described in $Prnp^{-/-}$ mice[49] suggesting an interplay between these proteins. Further, a SNP in *KAZN* showed one of the strongest associations in individuals with treatment resistant depression [26].

Based on the information we derived from various bioinformatics strategies we have 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. Though these results are interesting, functional follow up

studies will be required to determine the role of molecules like FGF, HSP90α 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 women, particularly as previous studies have suggested that brain transcriptomic changes associated with MDD are different in females[50]. Nonetheless, this first screen provides important information that may help inform subsequent studies exploring both men and women with MDD. Technical limitations with droplet-based snRNA-seq of human brain 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, pointing to a potential limitation of the methodology for capturing non-neuronal cells. Although droplet-based snRNA-seq does not capture lowly expressed genes, nevertheless, we were able to perform differential gene expression for thousands of genes in precisely defined cell types.

Lastly, 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 inconsistencies [10].

Our study has elucidated gene expression changes specific to numerous independent cell types in MDD. We have identified a potentially important link between OPCs and deep layer excitatory

neurons, which implicates fundamental pathways including FGF signalling, 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 start 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] .

Acknowledgements

GT 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 (Josée Prud'homme, Maâmar Bouchouka and Annie Baccichet), and the technology development team at the McGill University and Genome Quebec Innovation Centre. 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

CN conceptualized, performed experiments and wrote the manuscript, MM performed experiments, bioinformatics and wrote manuscript. AT, VJ and MAD performed experiments and wrote the manuscript. MS, JFT, ST and PP contributed to data analysis and reviewed the manuscript. NM contributed to tissue processing, data interpretation and manuscript preparation. JR provided technical single-cell expertise and experimental support, aided in manuscript preparation. GT provided general oversight of the study, including in experimental design, data interpretation and manuscript preparation.

Competing Interests statement

The authors declare no competing interests.

References

1. Organization, W.H. Depression and Other Common Mental Disorders: Global Health Estimates. 24 (Geneva, Switzerland, 2017).
2. Wray, N.R.*, et al.* Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature Genetics* **50**, 668-681 (2018).
3. Jansen, R.*, et al.* Gene expression in major depressive disorder. *Molecular psychiatry* **21**, 339-347 (2016).
4. Sequeira, A.*, et al.* Global brain gene expression analysis links glutamatergic and GABAergic alterations to suicide and major depression. *PLoS One* **4**, e6585 (2009).
5. 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* (2018).
6. 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. *Molecular psychiatry* **22**, 760 (2016).
7. Edgar, N. & Sibille, E. A putative functional role for oligodendrocytes in mood regulation. *Translational Psychiatry* **2**, e109 (2012).
8. Nagy, C.*, et al.* Astrocytic abnormalities and global DNA methylation patterns in depression and suicide. *Molecular psychiatry* **20**, 320-328 (2015).

9. 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).
10. 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).
11. Habib, N*., et al.* Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat Methods* **14**, 955-958 (2017).
12. Mancarci, B.O*., et al.* Cross-Laboratory Analysis of Brain Cell Type Transcriptomes with Applications to Interpretation of Bulk Tissue Data. *eneuro* **4** (2017).
13. Zheng, G.X.Y*., et al.* Massively parallel digital transcriptional profiling of single cells. *Nature Communications* **8**, 14049 (2017).
14. 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).
15. Renthal, W*., et al.* Characterization of human mosaic Rett syndrome brain tissue by single-nucleus RNA sequencing. *Nature Neuroscience* **21**, 1670-1679 (2018).
16. Mathys, H*., et al.* Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332-337 (2019).
17. Velmeshev, D*., et al.* Single-cell genomics identifies cell type-specific molecular changes in autism. *Science* **364**, 685-689 (2019).
18. Northoff, G. & Sibille, E. Why are cortical GABA neurons relevant to internal focus in depression? A cross-level model linking cellular, biochemical and neural network findings. *Molecular psychiatry* **19**, 966-977 (2014).
19. 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).
20. Sofroniew, M. & Vinters, H. Astrocytes: biology and pathology. *Acta neuropathologica* **119**, 7-35 (2010).
21. Anderson, M.A., Ao, Y. & Sofroniew, M.V. Heterogeneity of reactive astrocytes. *Neurosci Lett* **565**, 23-29 (2014).
22. 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).
23. 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**, W645-650 (2006).
24. 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).
25. Jakel, S*., et al.* Altered human oligodendrocyte heterogeneity in multiple sclerosis. *Nature* **566**, 543-547 (2019).
26. 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).

27. Gutierrez-Sacristan, A.*, et al.* PsyGeNET: a knowledge platform on psychiatric disorders and their genes. *Bioinformatics* **31**, 3075-3077 (2015).
28. Pinero, J.*, et al.* DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res* **45**, D833-D839 (2017).
29. 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).
30. 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).
31. 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).
32. 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).
33. Shulha, H.P.*, et al.* Human-specific histone methylation signatures at transcription start sites in prefrontal neurons. *PLoS Biol* **10**, e1001427 (2012).
34. Spitzer, S.O.*, et al.* Oligodendrocyte Progenitor Cells Become Regionally Diverse and Heterogeneous with Age. *Neuron* **101**, 459-471 e455 (2019).
35. 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).
36. Bergles, D.E., Jabs, R. & Steinhauser, C. Neuron-glia synapses in the brain. *Brain Res Rev* **63**, 130-137 (2010).
37. Birey, F., Kokkosis, A.G. & Aguirre, A. Oligodendroglia-lineage cells in brain plasticity, homeostasis and psychiatric disorders. *Current Opinion in Neurobiology* **47**, 93-103 (2017).
38. Ge, W.P.*, et al.* Long-term potentiation of neuron-glia synapses mediated by Ca2+-permeable AMPA receptors. *Science* **312**, 1533-1537 (2006).
39. 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).
40. Turner, C.A., Watson, S.J. & Akil, H. The fibroblast growth factor family: neuromodulation of affective behavior. *Neuron* **76**, 160-174 (2012).
41. 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).
42. Zuehlke, A.D., Beebe, K., Neckers, L. & Prince, T. Regulation and function of the human HSP90AA1 gene. *Gene* **570**, 8-16 (2015).
43. 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).
44. Pittenger, C. & Duman, R.S. Stress, depression, and neuroplasticity: a convergence of mechanisms. *Neuropsychopharmacology* **33**, 88-109 (2008).

45. 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).
46. 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).
47. 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).
48. Liu, J*., et al.* Impaired adult myelination in the prefrontal cortex of socially isolated mice. *Nat Neurosci* **15**, 1621-1623 (2012).
49. 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).
50. Labonte, B*., et al.* Sex-specific transcriptional signatures in human depression. *Nat Med* **23**, 1102-1111 (2017).

Figures and Figure Legends

Figure 1

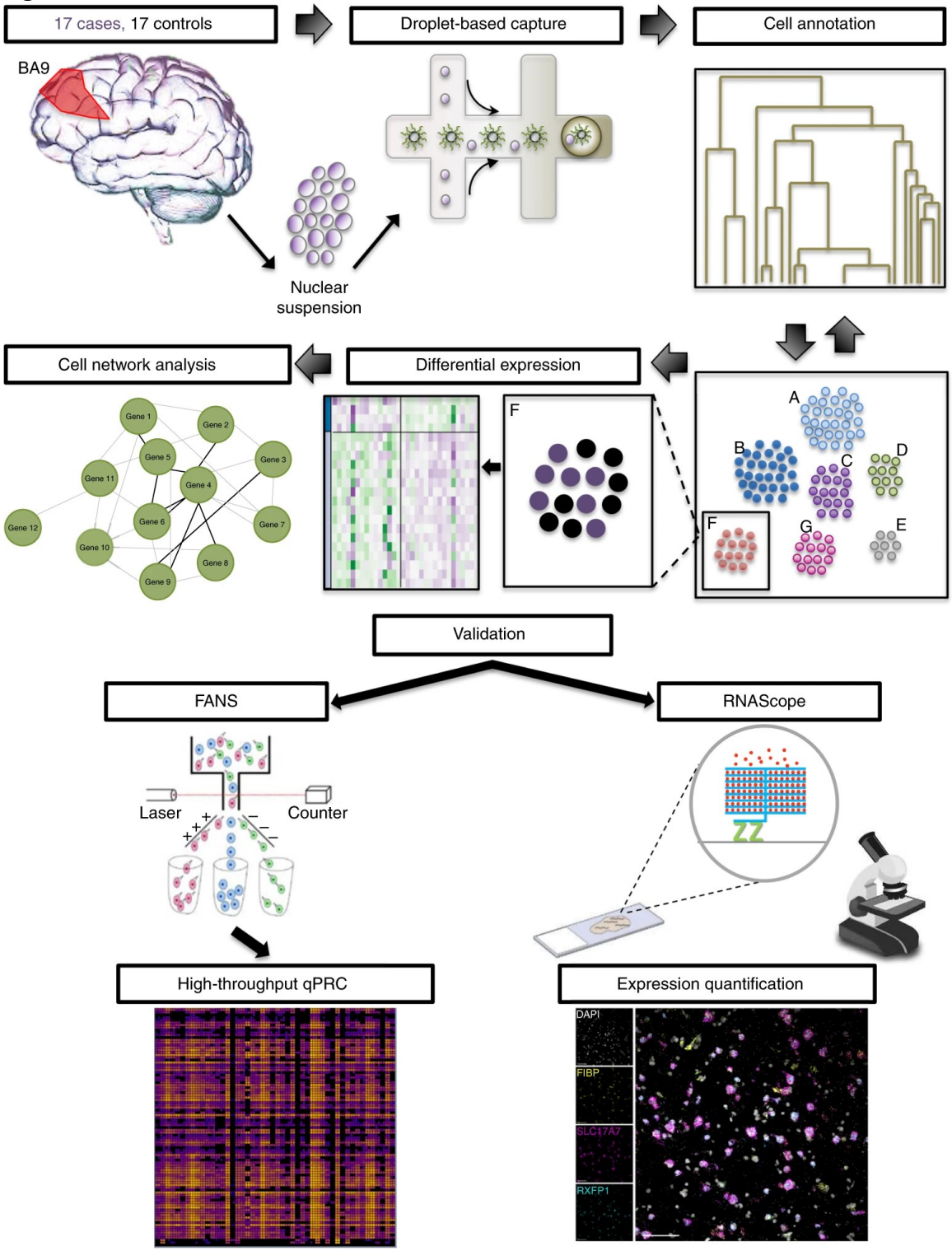


Figure 1: Experimental Flow. Schematic representation of experimental procedures. Nuclei were extracted from Brodmann area 9 (BA9) in the dlPFC of 17 cases and 17 controls, single nuclei were captured in droplets for RNA-seq. Unsupervised clustering and cell type annotation were followed by differential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

Figure 2

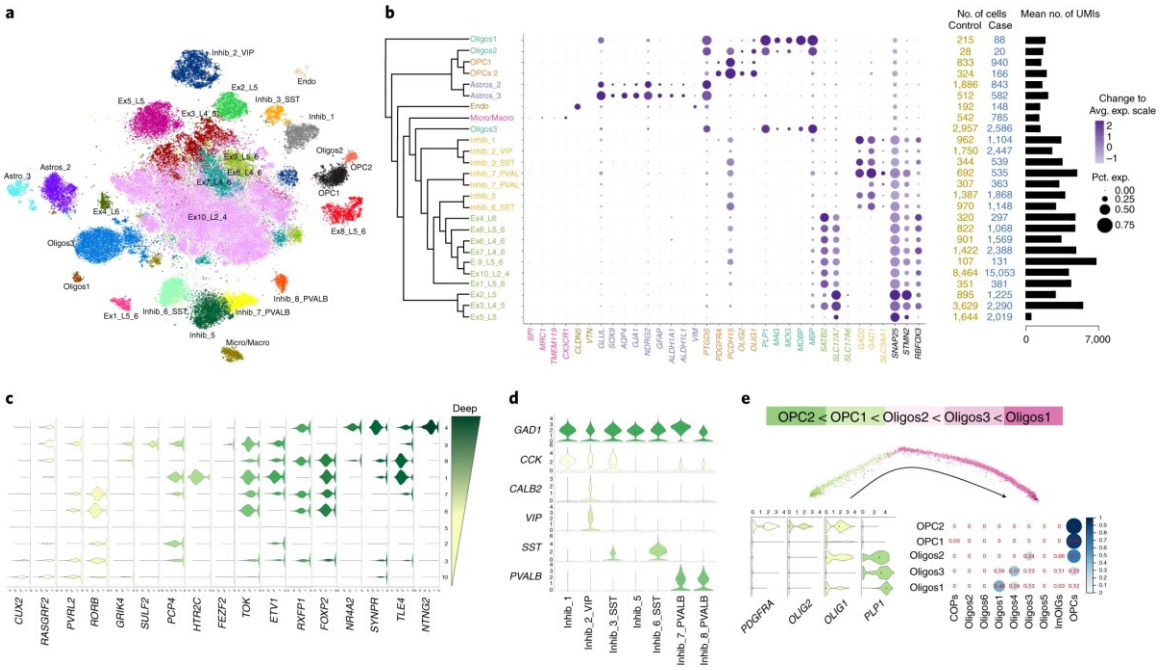


Figure 2: Identification of cell types a) TSNE plot depicting the ~73,000 cells in 26 clusters identified after strict quality control of initial clusters. **b)** Cell type annotation was performed based on expression of well-established marker genes. (Left) Dendrogram representing relationship between identified cell type clusters based on gene expression.

(Middle) DotPlot depicting expression of known marker genes in the 26 clusters of interest. Marker genes are colour 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 whereas the colour intensity represents the average expression level. (Right) Columns listing the number of cells per group and the bar plot depicting 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 going from the more superficial layers (I/II) on the left to the deeper layers (V/VI) on the right. **d)** Known classes of inhibitory neurons are identifiable based on the expression pattern of peptide genes (*VIP, SST, CCK*) and calcium binding protein genes (*PVALB*). **e)** (Left, violin plots) Cells belonging to the oligodendrocyte lineage expressed the expected markers**.** (Top) The oligodendrocyte lineage cells from 5 clusters were analysed to produce a pseudotime trajectory to gauge their developmental stages. . (Right) The location of these clusters along the trajectory was consistent with deconvolution (Jäkel et al., 2019). The numbers represent the percentage contribution of each of the previously published cluster signatures to the corresponding clusters in our dataset. For violin plots in figures 2c-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 2b. Nuclei were derived from 34 subjects.

Figure 3

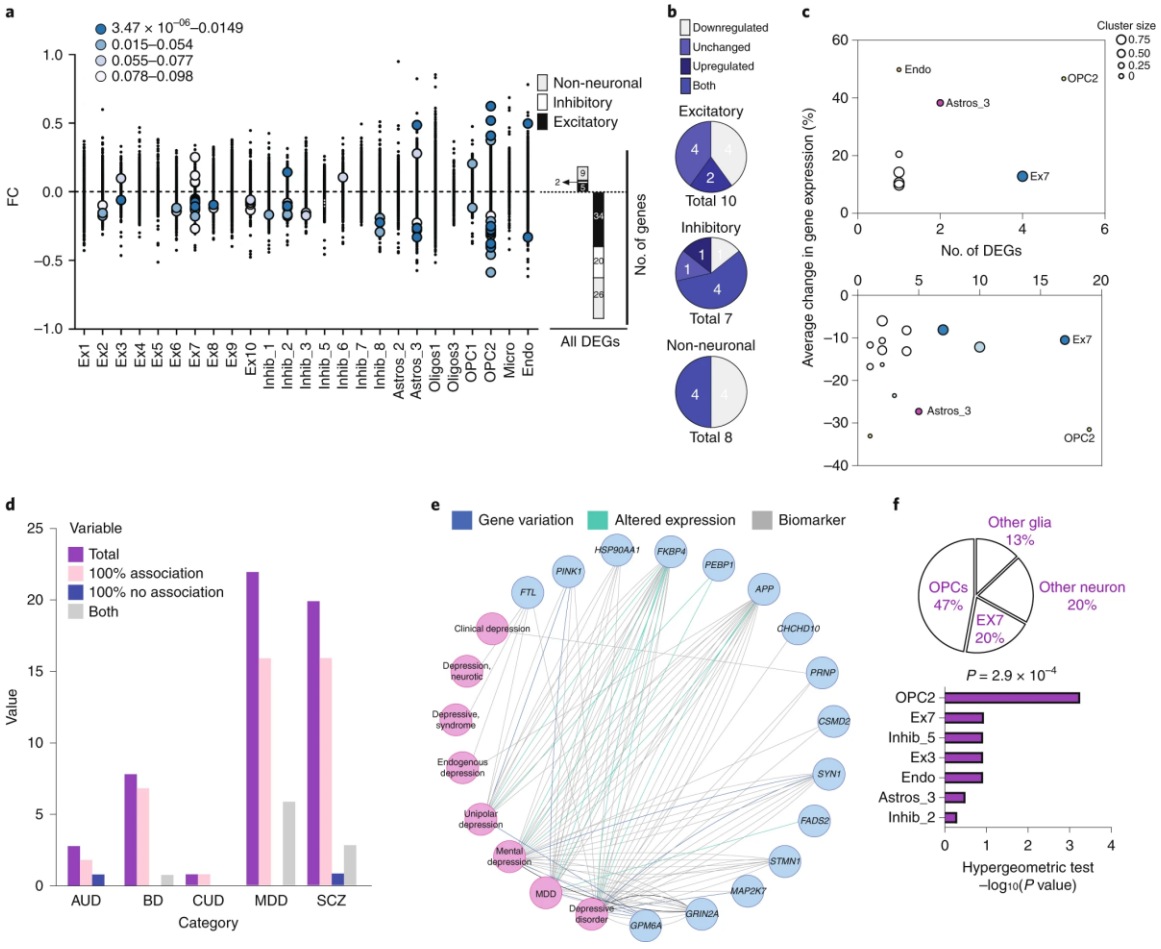


Figure 3: Differentially expressed genes. a) For each cluster the percentage change in expression between cases and controls of all detected genes are plotted with decreased expression to the bottom of the midline and increased expression to the top. Ninety-six significantly changed genes (16 were up-and 80 down-regulated) are marked in colour, based on their corrected FDRs as shown in the legend. The numbers of nuclei from cases and 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

subjects. Sixteen out of the 26 clusters contained significantly differentially expressed genes. (Insert) Stacked bar-graph shows contribution of different cell type clusters to differentially expressed genes. **b)** Number of clusters in each broad category showing up and downregulated genes in MDD cases. **c)** The scatter plots represent 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. Upper graph depicts upregulated genes, lower graph depicts downregulated genes. OPC2 and Ex7 show the highest level of both up and down regulated genes. **d-e)** The number of genes with known relationship to psychiatric phenotypes using available databases PsyGeNET and DisGeNET. **d)** 26 of the 96 dysregulated genes were found in PsyGeNET and showed an enrichment for MDD (Total, all the genes which overlap database for a given disorder; 100% association, the genes positively associated with the disease; 100% no association, the genes negatively associated with the disease; both, mixed findings (positive and negative) for a given gene related to the disease. **e)** (Left) 15 genes were found to be associated with depression related terms in DisGeNET. (Right) The percentage of genes per cluster associated with MDD from DisGeNET, along with cluster specific enrichment of DisGeNET MDD associated genes. For hypergeometric tests, the number of depression-associated genes in DisGeNET was 1199 and the number of unique genes in DisGeNET was 17545 for all tests. The number of DEGs in DisGeNET (k) and the number of depression-associated DEGs (x) are listed: All clusters: k=85, x=15; OPC2: k=24, x=7; Ex7: k=19, x=3; Endo: k=2, x=1; Astro3: k=6, x=1; Ex3: k=2, x=1; In2: k=11, x=1; In5: k=2, x=1.

Figure 4

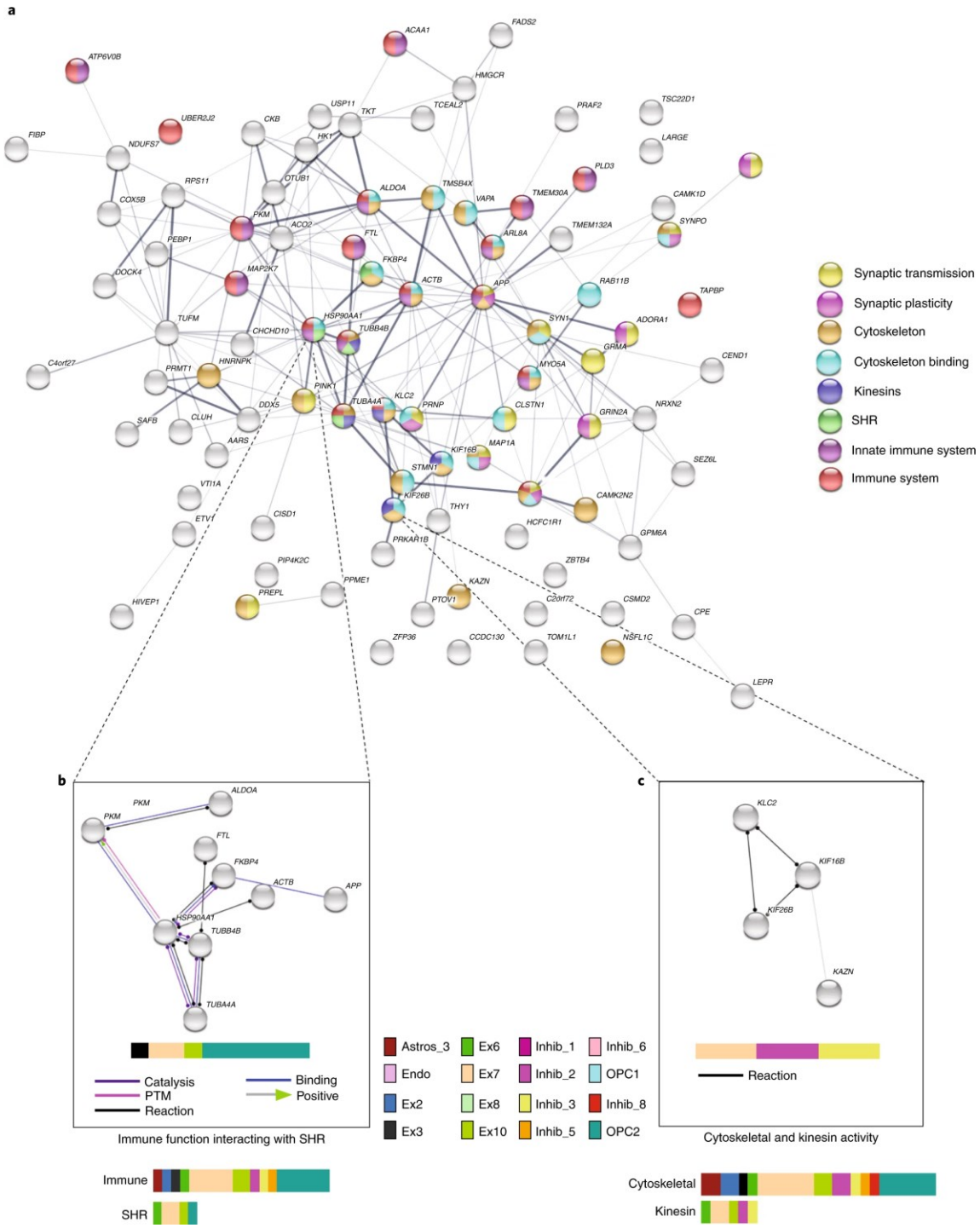


Figure 4: Differential expression and biological associations. a) String DB network for all

DEGs with nodes corresponding to a set of biological processes and pathways highlighted (legend on right). **b)** Subset of genes shared between the immune function related terms and the steroid hormone receptor cycling pathway. **c)** Subset of genes involved in cytoskeletal function and kinesin activity. Colour strips beneath networks give a proportional representation of the contributing clusters.

Figure 5

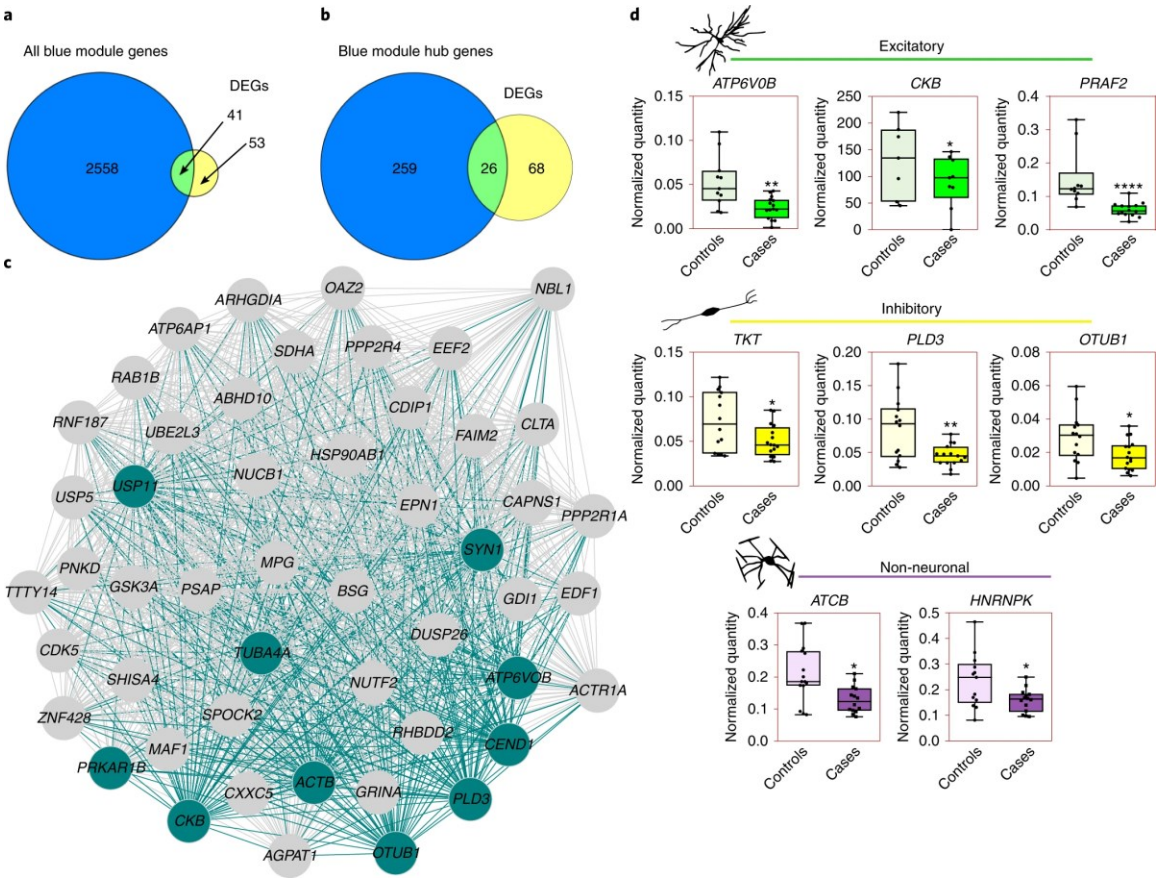


Figure 5: Weighted gene co-expression network analysis. a) Venn diagram of overlap between blue module genes and DEGs (hypergeometric test, p-value = 6.037692e-19). **b)**

Venn diagram for overlap between blue module hub genes and DEGs (hypergeometric test, p-value = 4.954172e-31). **c)** Visualization of the top 50 hub genes assessed for the blue module. DEG nodes and all edges connected to them are colored teal. **d)** Boxplots represent expression levels of DEGs validated with high-throughput qPCR in FAN sorted populations which 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 with unpaired two-sided t-tests as their values were normally distributed. P-values: * < 0.05, ** < 0.01, *** < 0.001, **** < 0.0001. Whiskers on box plot represent maximum and minimum values. Box extends from the 25th percentile to 75th percentile, the center line represents the median, and dots represent all values in the dataset. *ATP6V0B*: n=15 cases, 11 controls, t=3.10, df=12.62, p-value=0.0087; *CKB*: n=9 cases, 7 controls, t= 2.48, df=16.85 p-value= 0.023 ; *PRAF2*: n=14 cases, 10 controls, U= 8, p-value=6.8 x 10^{-5}; *TKT*: n= 16 cases, 14 controls, t= 2.25, df=19.83, p-value=0.036; *PLD3*: n=15 cases, 14 controls, t= 3.06, df= 15.83, p-value=0.0075; *OTUB1*: n=16 cases, 14 controls, t= 2.39, df=20.92, p-value=0.026; *ACTB*: n=14 cases, 15 controls, t=3.14, df= 19.98, p-value=0.0052; *HNRNPK*: n= 14 cases, 13 controls, t=2.41, df=16.07, p-value=0.028.

Figure 6

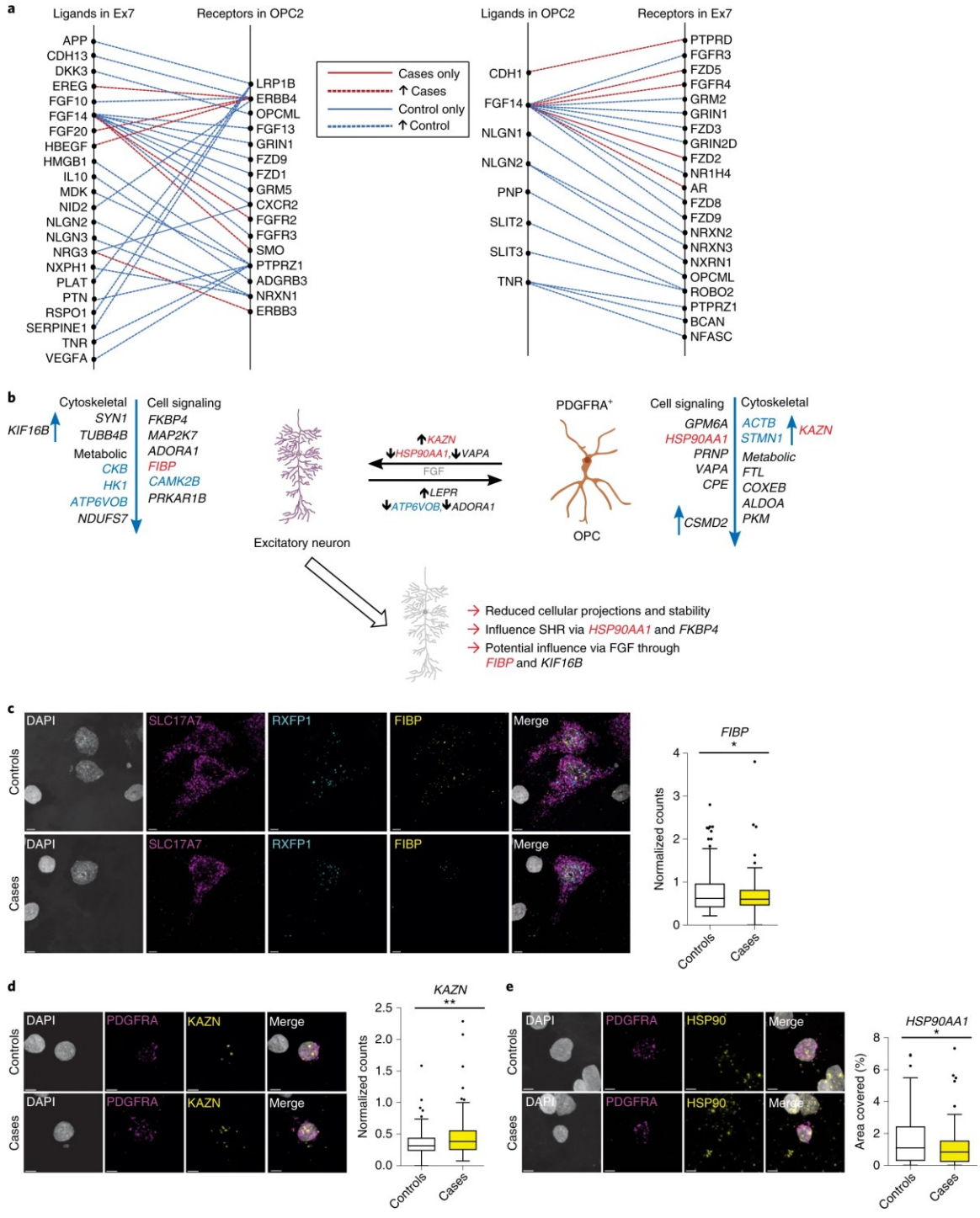


Figure 6: Contributions of OPC2 and Ex7. **a)** CCInx receptor ligand based cell-cell interaction network analysis for communication between Ex7 and OPC2. Given the large

number of connections (Supplementary Tables 47a,b), a subset are shown. **b)** Our data points to a change in the communication between deep layer excitatory neurons (Ex7) and immature OPCs (OPC2). Altered FGF bidirectional signalling was identified via CCInx. We propose that immature OPCs have a very 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 found to be dysregulated in the given cell type for example *HPS90AA1* codes for the stress inducible isoform HSP90α, known to be secreted in certain contexts, *KAZN* is an upregulated junction protein *i*n OPCs and *ATP6V0B* could represent altered ATP signaling. Arrows beside gene names indicate up or downregulation. Beside each cell type are the genes in given functional categories and their direction of change in the disease state. **c)** Decreased expression of the gene encoding FGF1 intercellular Binding Protein (*FIBP*) was validated in deep layer neurons using RNAScope®. *SLC17A7* (encoding VGLUT) was used as a marker for excitatory cells and *RXFP1* was used to identify deep layer neurons. *SLC17A7*$^{+}$*, RXFP1*$^{+}$ cells were imaged and *FIPB* expression was counted (Cases: n=119 nuclei, controls: n=100 nuclei, unpaired two-sided t-test, t = 2.49, df= 217, p = 0.013). **d)** Increased *KAZN* (cases: n=95 nuclei, controls: n=95 nuclei, unpaired two-sided t-test, t = -2.69, df= 188, p = 0.008) and **e)** decreased *HSP90AA1* (cases: n = 87 nuclei, controls: n = 107 nuclei, unpaired two-sided t-test, t = 2.23, df= 186, p= 0.027 expression were validated in OPCs using *PDGFRA* as a marker for oligodendrocyte precursor cells Whiskers on box plot represent the 5th and 95th percentile. Box extends from the 25th percentile to

75th percentile and the center line represents the median. Dots represent points beyond the 5th or 95th percentile.

Tables

Table 1: Sample information

	Controls (n=17)	Cases (n=17)	p value
Age (years)	38.71 ± 4.32	41.06 ± 4.66	p=0.714
Gender	17M	17M	-
PMI (hrs)	34.01 ± 4.94	41.69 ± 4.76	ǂp=0.190
pH	6.49 ± 0.06	6.60 ± 0.07	p=0.212
Storage Time (years)	14.71± 1.44	12.47± 1.46	ǂp=0.543
Cause of death	Accident (6), Natural (11)	Suicide (17)	
Substance dependence	None	None	
Comorbid diagnoses	None	None	
Toxicology	EtOH (2), Cannabinoids (1),	EtOH (6), BZ (1), AD (2), Cannabinoids (1), Cocaine (1),	
Antidepressant Treatment	None	3	

Mean ± SEM
ǂMann Whitney test
NA – not applicable, EtOH – ethanol, BZ – benzodiazepines, AD – antidepressants, AC – anticonvulsants

Materials and Methods

Subjects: 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 subject. Postmortem brain samples were provided by the Douglas-Bell Canada Brain Bank (www.douglasbrainbank.ca). Frozen grey 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 by proxy-based interviews, as described previously[51]. 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). Post mortem interval (PMI) represents the delay between a subject's death and collection and processing of the brain. To assess RNA quality, we measured the 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). 17 cases and 17 controls were included in the snRNA-seq experiment and the full cohort of subjects (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 5 cases and 5 matched controls. Detailed information on experimental design and reagents can also be found in the Life Sciences Reporting Summary.

Nuclei isolation and capture

50 mg of frozen tissue 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 $MgCl_2$, 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 are 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 ~$1x10^6$ nuclei/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_RevE.pdf***), with the exception of loading concentration, which we increase by 30% as we assessed the capture of nuclei to be slightly less efficient than cell encapsulation. We aimed to capture ~3000 nuclei per sample. So, for example, if our sample concentration was 390 nuclei/µL (~ 400 nuclei/ µL) according to page 10 of Protocol Step 1 we are required to load 13.1 µL of the stock to capture 3000 cells. But instead, we would recalculate our stock concentration to be 70% of 390 = 273 nuclei/ µL and load 17.4 µL (the recommended amount for 300 nuclei/ µL) instead. This system only allows for a maximum of 8 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, we therefore, 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 version 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 sample index using the cellranger mkfastq command (Cellranger v2.1.0). 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 (version 2.2.0, 2.3.0)[19]. Cell names were modified such that the subject 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 combined into one object using the MergeSeurat function sequentially. No filtering or normalization was performed up to this step. Since this is a single nucleus dataset, all mitochondrial genes that are 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 (total number of genes detected in each cell) for the total dataset (assessed by summary and hist R[52] functions), barcodes that were associated with less than 110 detected genes were removed. Based on the distribution of nUMI (total numbers 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 fit with three normal distributions using the normalmixEM function from the mixtools[53] package (Supplementary Fig. 1c). The rationale 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). 78,886 cells and 30,062 genes were retained.

Our custom filtering (Supplementary Fig. 1a-e, Supplementary Table 4) helped to increase the number of glial cells recovered. With an initial subset of 20 subjects, applying our custom filtering increased the total number of cells 1.8–fold but increased the number of non-neuronal cells by almost 6-fold (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 (pers. comm., Lake, B.B.). 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 subject information was 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, y.cutoff = 1. This resulted in a list of 2135 highly variable genes, which excludes lowly expressed genes (below 25^{th} percentile), very highly expressed genes, and selects only the top 10 % of genes in terms of the scaled dispersion. These highly variable genes were used to calculate 100 principal components. 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 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. Prior to the advent of single nuclei expression profiling, cell types were identified by observing differences in cell morphology, behaviour, and anatomic location. It is fairly straight-forward 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 if 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 which were in the bottom 25 % of the dendrogram (using the branching.times function from the ape R package[54]) 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 markers genes of all major, well-defined brain

cell types (Supplementary Fig. 2c-d). For refining identification of excitatory neuron types, we combined and re-clustered 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 ~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, 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 expression of canonical inhibitory interneuron markers *SST, PVALB,* and *VIP* where possible. Excitatory neuron subtypes were annotated with some level of layer specificity based on expression of layer specific markers[11, 55, 56]. We also characterised 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 (Supplementary Fig. 3a-p)

Macrophage/ Microglia: *SPI1, MRC1, TMEM119, CX3CR1*; **Endothelial:** *CLDN5, VTN;* **Astrocytes:** *GLUL, SOX9, AQP4, GJA1, NDRG2, GFAP, ALDH1A1, ALDH1L1, VIM;* **OPCs***: PTGDS, PDGFRA, PCDH15, OLIG2, OLIG1;* **Oligodendrocytes:** *PLP1, MAG,MOG, MOBP, MBP;* **Excitatory neurons:** *SATB2, SLC17A7, SLC17A6;* **Inhibitory neurons:** *GAD1,GAD2, SLC32A1;* **Neurons:** *SNAP25,STMN2, RBFOX3.*

Layer-specific markers:

L2: *GLRA3;* **L2-3:** *LAMP5, CARTPT*; **L2-4:** *CUX2, THSD7A*; **L2-6:** *RASGRF2, PVRL3*; **L3-4**: *PRSS12;* **L4-5**: *RORB*; **L4-6:** *GRIK4*; **L5**: *KCNK2, SULF2, PCP4, HTR2C, FEZF2*; **L5-6**: *TOX, ETV1, RPRM, RXFP1, FOXP2;* **L6:** *SYT6, OPRK1, NR4A2, SYNPR, TLE, NTNG2, ADRA2A*

Pseudotime trajectory using Monocle

For oligodendrocyte developmental trajectory assessment, the data for cells belonging to the five clusters in the oligodendrocyte lineage (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, y.cutoff = 1 (giving a total of 895). The Seurat object was imported into a CDS (CellDataSet) object using the Monocle[22] 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 as 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 oligodendrocyte lineage cells from control subjects and MDD cases with fullModelFormulaStr = "~sm.ns(Pseudotime)". This allows 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 comprised 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 cluster) 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 controls was performed using linear mixed models implemented in the lme4[57] and lmerTest[58] R packages. Mixed models were necessary in order to account for dependencies between nuclei obtained from the same subject. Biological condition and number of UMIs were included in models as fixed effects and the subject and batch as random effects. The inclusion of subject as a random effect should account for subject specific effects such as age and PMI as well as technical effects of capture and library

preparation which was performed separately for each subject. A false discovery rate (FDR) of 0.1 was used to detect differentially expressed genes within each cell type.

Weighted Gene Co-expression Network Analysis (WGCNA)

Average cell-expression for each sample across every cluster was calculated. These average counts were converted to log + 1 counts to reduce dispersion. WGCNA analysis was carried out in R with the WGCNA package (version 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 vs Control), as well as each sample's respective composition of each of the 26 single-cell type clusters they're composed of.

We performed hub gene analysis on the blue module, which was the largest module (2699 genes) which was correlated to phenotype. Potential hub genes were identified in the module of interest my selecting genes with a module membership larger than 0.80 and a gene significance larger than 0.20 with a p-value of less than 0.05. The top 50 potential hub genes were extracted alongside any weighted interaction of more than 0.2. The resulting network was visualized in Cytoscape (3.7.1).

Fluorescence-assisted nuclei sorting (FANS)

Nuclear suspensions were prepared from 80-100 mg of post-mortem brain tissue from BA9 as described previously [59] with the following modifications: homogenized tissue was centrifuged on the sucrose layer at 800g for 20 minutes at 4^{o}C, followed by another centrifugation in nuclei

extraction buffer. Resuspended nuclei were stained with the following primary antibodies in 600 µL of blocking buffer: mouse anti-CUTL2-PerCP conjugated (1:100, Novus catalog no. H00023316-M03, clone 2H8, lot 080618-112618 conjugated to PerCP using the Novus Lightning Link Labeling kit, catalog no. 718-0010), goat anti-SOX10 (1:100, R&D Systems catalog no. AF2864), mouse anti-NeuN-A700 (1:300, Novus catalog no. NBP1-92693AF700, clone- 1B7) by incubating at room temperature, away from light, with rotation for 2 hours. Secondary antibody (donkey anti-goat Alexa Fluor 488, 1:1000, JacksonImmuno 705-545-147) was added and incubated for 1 hour at room temperature with rotation. All antibodies were purchased from Cedarlane. Nuclei were washed with PBS and the DNA was stained by Hoechst 33342 (Invitrogen, H1399).

FACSAria Fusion (BD Biosciences, San Jose, CA) was used for sorting of four populations – SOX10 positive, SOX10 negative, CUTL2 positive and CUTL2 negative. Gating strategy for the sorts is shown in (Supplementary Fig. 11) and was as follows. 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 CULT2 negative population was gated within PerCP intensities of SOX10 populations. CUTL2 negative population comprised near 10% of NeuN

positive population.

Validation information for antibodies is as follows: 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).

High-throughput qPCR

RNA was extracted from FANS sorted nuclei population using the Norgen RNA/DNA Purification Kit (Cat. 48700). cDNA was synthesized using a modified SMART-seq procedure as described previously [60]. The Fluidigm Biomark system was used for performing high-throughput qPCR as per manufacturer protocol as previously described[61]. Fludigim Delta Gene™ primer designs were used for the 93 targets (all differentially expressed transcripts excluding *AC133680.1*) and 3 endogenous controls (*GAPDH, POLR2A, UBC*).

Cell-cell interaction measurement

To assess cell-cell communication, we calculated predicted ligand-receptor interactions between Ex7 and OPC2 using CCInx[62] (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 research to relatively highly expressed ligands and receptors and for ease of visualization. To test if the edge weights were significantly different between cases and controls, we randomly permuted our subjects 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 p-values <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.p1)[11] was used as reference signatures for annotated cell types. UMI counts for each cell were converted to transcripts per million (TPMs) in order 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 24301 genes common to our dataset and the reference for each nucleus in each cluster and converting the sums to TPMs. R package, DeconRNASeq v1.18.0[63] was used to deconvolute these cluster-specific profiles. Using the data from[11]as reference, we were able to estimate the cell type composition of our clusters.

Cell deconvolution for oligodendrocyte lineage

Average expression from every control samples from the Jäkel et al. dataset were calculated and used as cell signatures for the deconvolution of our oligodendrocytic clusters (average cell expression of every cell in the cluster considered as bulk) using the R package DeconRNASeq (v 1.26.0).

RNA-Scope Fluorescent *In Situ* Hybridization

Frozen BA9 blocks were cut serially with a cryostat (10µm thickness) on superfrost charged slides and kept at -80^{o}C until further processed. *In situ* hybridization was performed using Advanced Cell Diagnostics RNAscope® probes and reagents according to the manufacturer instructions in 5 matched subjects per group. Briefly, sections were first fixed in chilled 10% neutral buffered formalin for 15 mins at 4^{o}C, dehydrated by increasing gradient of ethanol bathes and left to air dry for 5 minutes. Endogenous peroxidase activity was quenched with hydrogen peroxide reagent for 10 minutes, followed by protease digestion for 30 minutes at room temperature. The following sets of probes were then hybridized for 2 hours at 40^{o}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 5-6 (RXFP1+) neurons; KAZN (cat. no. 569791) and PDGFRA (cat. no. 604481-C3), and HSP90AA1 (cat. no. 477061) to quantify KAZN expression in OPCs (PDGFRA+). Successive addition of amplifiers was performed using the proprietary AMP reagents, and the signal visualized through probe-specific HRP-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 DAPI for nuclear staining (Vector Laboratories) and kept at 4^{o}C until imaging.

Imaging and analysis of *in situ* RNA expression

Image acquisitions was performed on a FV1200 laser scanning confocal microscope (FV1200) equipped with a motorized stage. For each experiment and subject, around 10 stack images were taken to capture at least 20 cells of interest per subject: excitatory neurons (SLC17A7+) from cortical layers 5-6 (RXFP1+), and OPCs (PDGFRA+). Images were taken using a x60 objective (NA

= 1.42) with a XY pixel width of 0.3µm and Z spacing of 0.4µm. Laser power and detection parameters were kept consistent between subjects for each set of experiment. Because TSA 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 signal present within the nucleus and cytoplasm of the cells. Expression of genes of interest was quantified using the "Analyze Particles" function in Fiji [64]. Stacks were first converted to Z-projections, and for each image cell nuclei of cells of interests were manually contoured based on DAPI expression. 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 to RXFP1 and PDGFRA raw counts respectively. HSP90AA1 expression was quantified by manually 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 size. Sample size was determined based on sample sizes used in previous similar studies. Subjects were assigned to groups based on diagnosis and not by random assignment. All subjects were male, and groups were matched for age (18-87 years), post-mortem interval (12-93 hours), and brain pH (6-7.01). Clinicians were blinded for final psychiatry autopsy 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 subjects 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 (lowly correlated nuclei were removed) as detailed above to ensure that differential gene expression analysis was performed using similar nuclei populations from cases and controls. The exclusion criteria were not pre-established and were chosen based on preliminary analysis of the data.

Differential expression analysis between the cases and controls in the snRNA-seq data was performed using linear mixed models implemented in the lme4[57] and lmerTest[58] R packages with biological condition and number of UMIs as fixed effects, the subject and batch as random effects, and a false discovery rate of 0.1 for significance. For analysis of RNAScope results, two-tailed t-tests were performed with a significance threshold of $p < 0.05$ and data distribution was assumed to be normal but this was not formally tested. For analysis of high-throughput qPCR data two-tailed t-tests or two-tailed Wilcoxon rank sum (i.e. Mann Whitney U tests) were performed, both at a significance threshold of $p < 0.05$, and depending on data normality as measured by the Shapiro Wilk's test.

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 or using this link: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE144136.

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.

Accession Numbers

GEO accession number for snRNA-seq data: GSE144136.

Method only 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. *2009* **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. *2015* **67**, 48 (2015).
58. Kuznetsova, A., Brockhoff, P.B. & Christensen, R.H.B. lmerTest Package: Tests in Linear Mixed Effects Models. *2017* **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.* Transcript Profiling Using Long-Read Sequencing Technologies. in *Gene Expression Analysis: Methods and Protocols* (ed. N. Raghavachari & N. Garcia-Reyero) 121-147 (Springer New York, New York, NY, 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. *Nature Methods* **9**, 676 (2012).

Supplementary Information

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

Corina Nagy[#,1], Malosree Maitra[#,1], Arnaud Tanti[1], Matthew Suderman[5], Jean-Francois Théroux[1], Maria Antonietta Davoli[1], Kelly Perlman[1], Volodymyr Yerko[1], Yu Chang Wang[3,4] Shreejoy J. Tripathy[6,7], Paul Pavlidis[6], Naguib Mechawar[1,2], Jiannis Ragoussis[3,4], Gustavo Turecki[1,2,3]*

[1]McGill Group for Suicide Studies, Douglas Mental Health University Institute, Montreal, QC, Canada. [2]Department of Psychiatry, McGill University, Montreal, QC, Canada. [3]Department of Human Genetics, McGill University, Montreal, QC, Canada. [4]McGill University and Genome Quebec Innovation Centre, Montreal, QC, Canada. [5]MRC Integrative Epidemiology Unit, University of Bristol, UK. [6]Michael Smith Laboratories and Department of Psychiatry, University of British Columbia, Vancouver, BC, Canada. [7]Krembil Centre for Neuroinformatics, Centre for Addiction and Mental Health, Toronto, ON, Canada. [8]Department of Psychiatry, University of Toronto, Toronto ON, Canada.

[#] These authors contributed equally

***Correspondence to**: Dr. Gustavo Turecki (gustavo.turecki@mcgill.ca)

6875 LaSalle Blvd, Verdun, Québec, Canada H4H 1R3

Tel.: (514) 761-6131 (ext. 2370)

Supplementary Methods

Intermediate Steps of Cell Type Clustering

Characterization of the 30 preliminary clusters in terms of known marker genes identified 3 astrocytic, 3 oligodendrocytic, 2 oligodendrocyte precursor, 1 microglial/macrophagic, 1 endothelial, 8 inhibitory neuron, and 9 excitatory neuron clusters (Supplementary Fig. 2c-d). In addition, there were three "mixed" cluster, which either expressed both neuronal and glial markers or expressed both excitatory neuronal and inhibitory neuronal markers.

In past studies of large single-nucleus human brain dataset, broad clusters were identified and subsequently sub-clustered[1]. We reclustered some of our large excitatory clusters (7 in total), including one containing ~27,000 cells, as shown in Supplementary Fig. 13a-c. This resulted in 10 new sub-clusters, two of which appeared to be composed of cell doublets (excitatory/inhibitory and neuronal/glial cells) (Supplementary Fig. 13b).

Two small clusters (Astros_1 and Inhib_4_SST) were omitted because they were each composed of a small number of cells from a single subject (subjects 12 and 10, respectively) (Supplementary Table 50-51). In Astros_1, there was increased GFAP (highest among three astrocytic clusters) and JUNB; in Inhib_4_SST there was increases JUN, FOS, and NPAS4 (Supplementary Fig. 14c). Thus, the presence of a population of "activated" cells in these subjects probably explains the two subject-specific clusters. All other clusters comprised of cells from the majority of subjects.

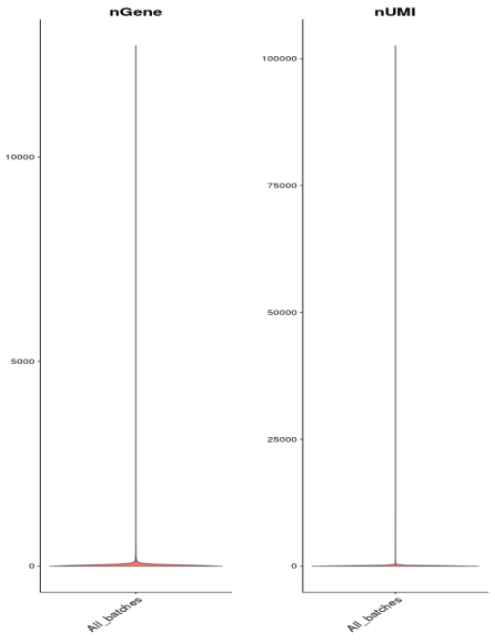
After removing all doublet clusters and clusters composed of cells from only a single subject, we retained 26 clusters (2 astrocytic, 3 oligodendrocytic, 2 oligodendrocyte precursor, 1 endothelial, 1 microglial/macrophagic, 7 inhibitory neuronal, and 10 excitatory neuronal) with relatively even

distribution from subjects, batches, and biological conditions (Supplementary Table 7a, 49-51, Supplementary Fig. 14a-b).

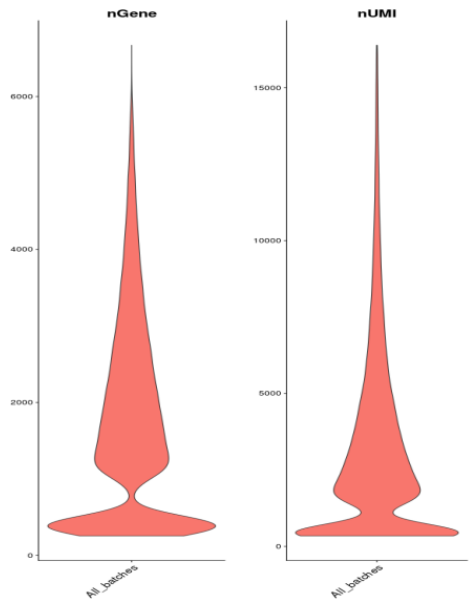
Closer Examination of Selected Differentially Expressed Genes

The per cell (nucleus) expression of selected differentially expressed genes in Ex7 and OPC2 is represented in the density in Supplementary Fig. 8a-f. While the heatmaps in Supplementary Fig. 9a-p depict the average expression per subject in the specified cluster for thee genes, the density plots give a clearer picture of the range of expression of these genes in individual nuclei within the cluster for controls and depressed cases. The overall difference in expression level between nuclei derived from control subjects and MDD cases are seen in these plots.

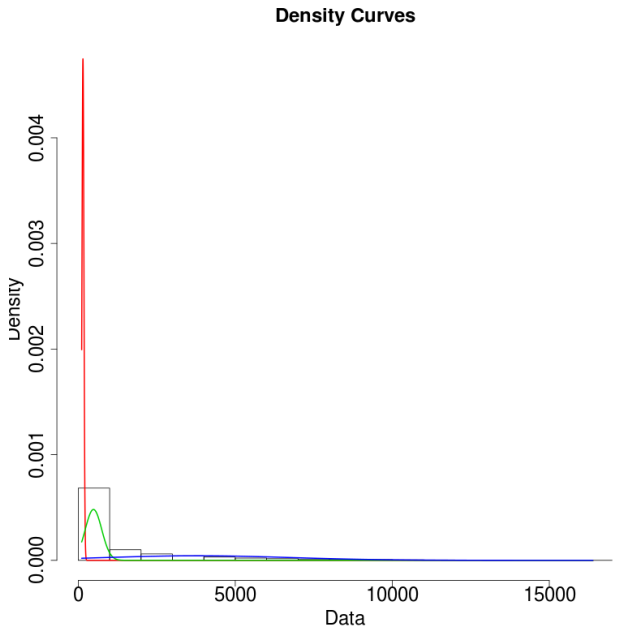
a


nGene
nUMI
10000
5000
0
100000
75000
50000
25000
0
All_batches
All_batches

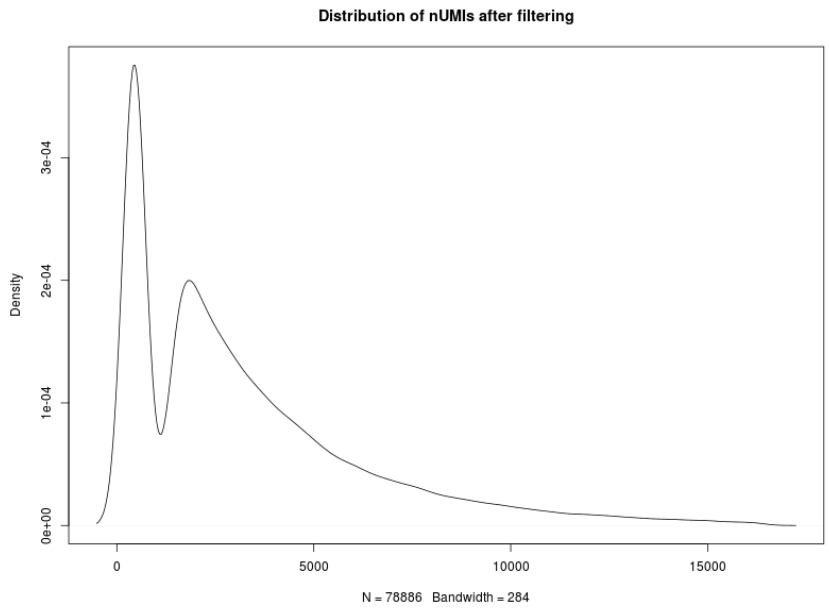
b


nGene
nUMI
6000
4000
2000
0
15000
10000
5000
0
All_batches
All_batches

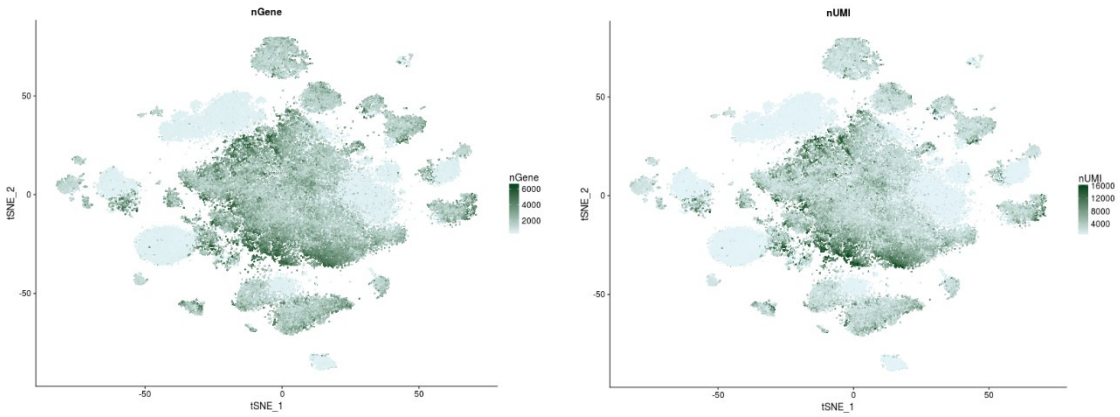
c


Density Curves
Density
0.000
0.001
0.002
0.003
0.004
0
5000
10000
15000
Data

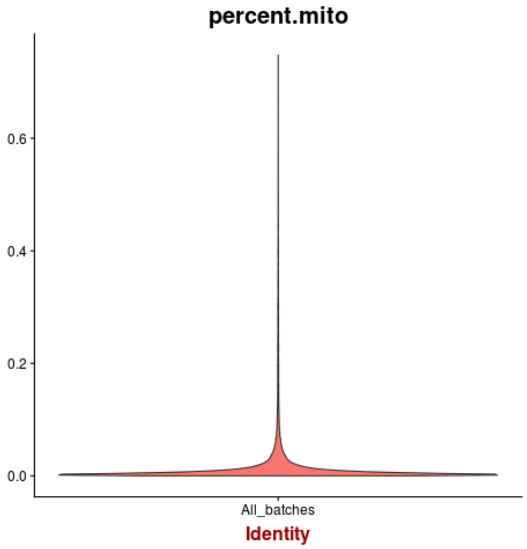
d


Distribution of nUMIs after filtering
Density
0e+00
1e-04
2e-04
3e-04
0
5000
10000
15000
N = 78886 Bandwidth = 284

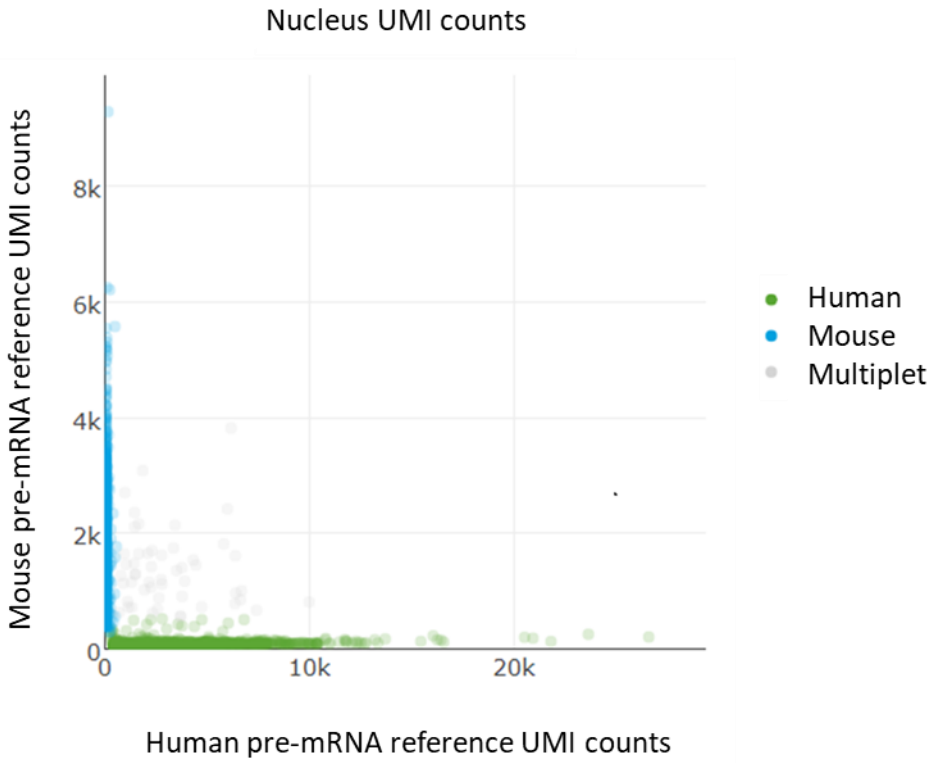
e


nGene
tSNE_2
tSNE_1
nGene
6000
4000
2000
nUMI
nUMI
16000
12000
8000
4000

f



g



Supplementary Figure 1

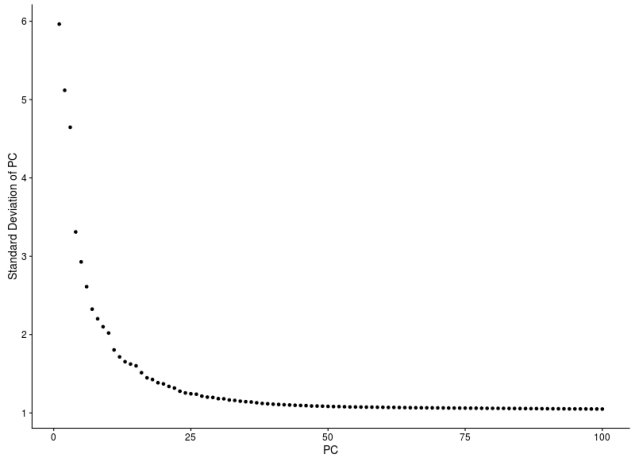
General data quality metrics

Violin plots showing distribution of numbers of UMIs and numbers of genes detected in each cell

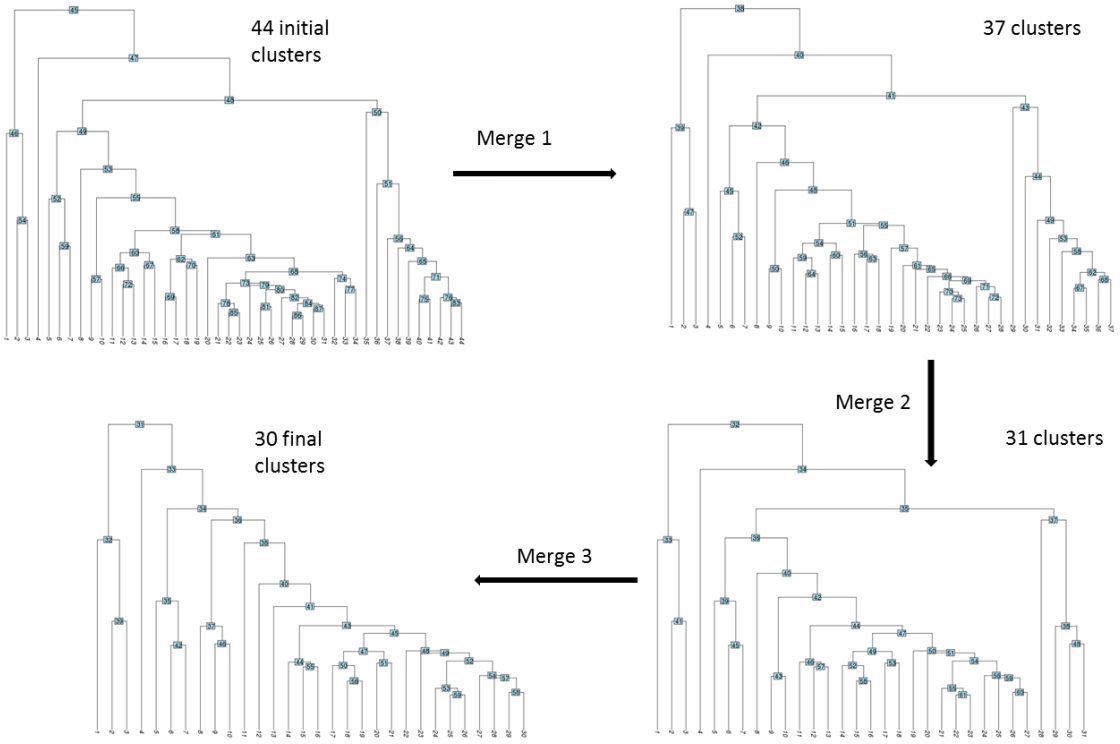
(a) before (n ~ 7 million barcodes detected pre-filtering from 34 subjects) and **(b)** after (n=78,886

individual nuclei from 34 subjects) custom filtering. **(c)** Normal distributions from mxmdl used for filtering (see Methods). Cells with high probability of belonging to the green or blue distribution were retained. **(d)** Distribution of number of UMIs per cell after all filtering steps showing two peaks likely roughly corresponding to non-neuronal and neuronal cells. **(e)** TSNE plots shaded by the number of genes (left), and the number of UMIs (right) detected in each cell, n=78,886 individual nuclei from 34 subjects. **(f)** The percentage of UMIs corresponding to mitochondrially encoded genes was below 5% for 90% of the cells in all batches (n= 78,886 individual nuclei from 34 subjects). **(g)** Barnyard plot produced by CellRanger for data generated by us using a mouse-human mixed sample depicting mouse-human doublets (total rate of 5.2%) using the pre-mRNA reference and default filtering. All violin plots extend from the minimum to the maximum value.

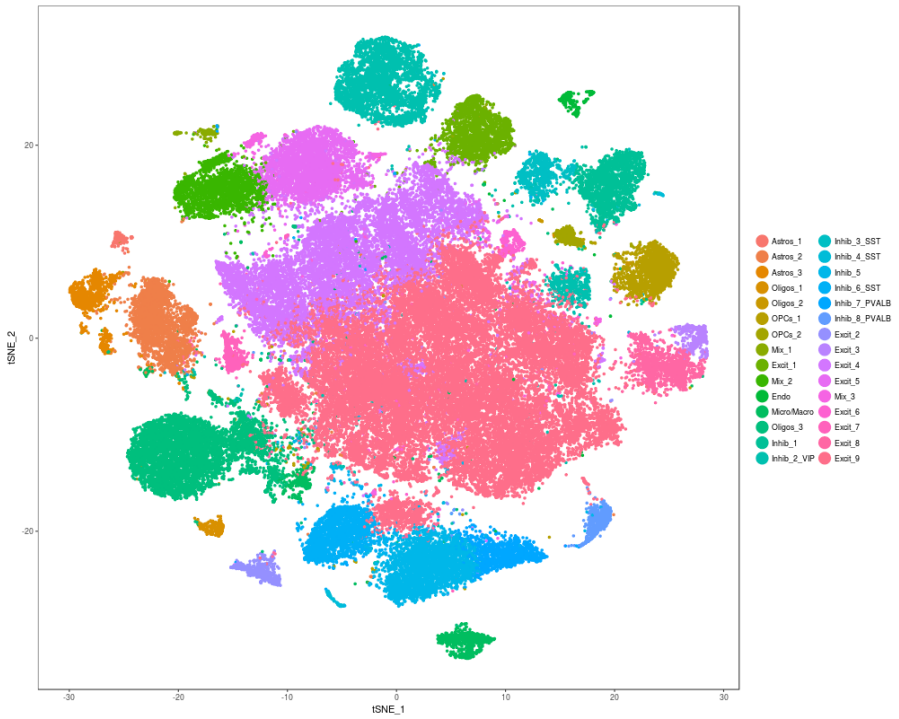
a


Standard Deviation of PC
PC
6
5
4
3
2
1
0
25
50
75
100

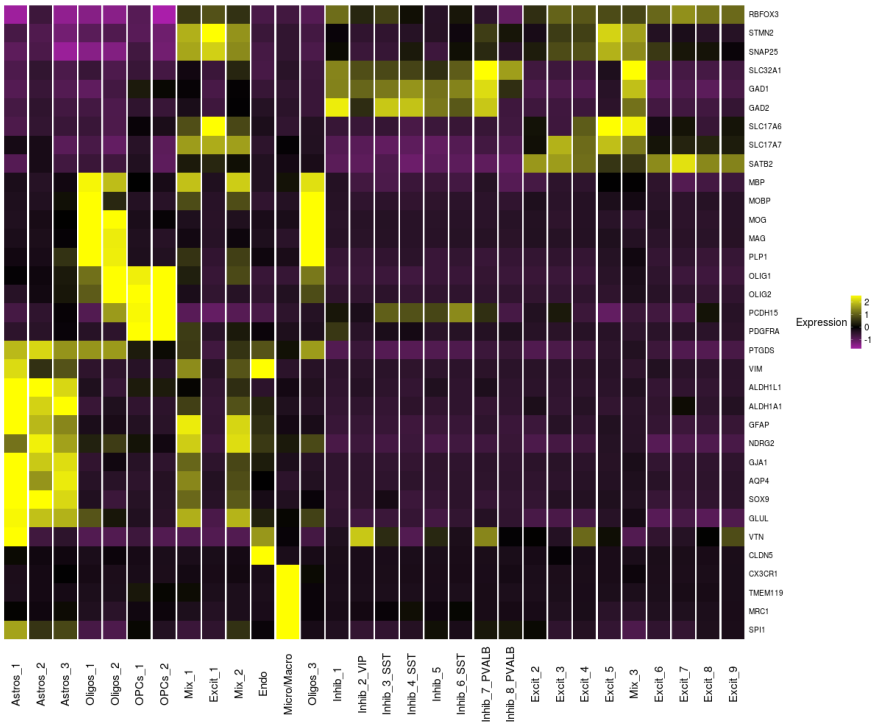
b


44 initial clusters
Merge 1
37 clusters
Merge 2
31 clusters
Merge 3
30 final clusters

c



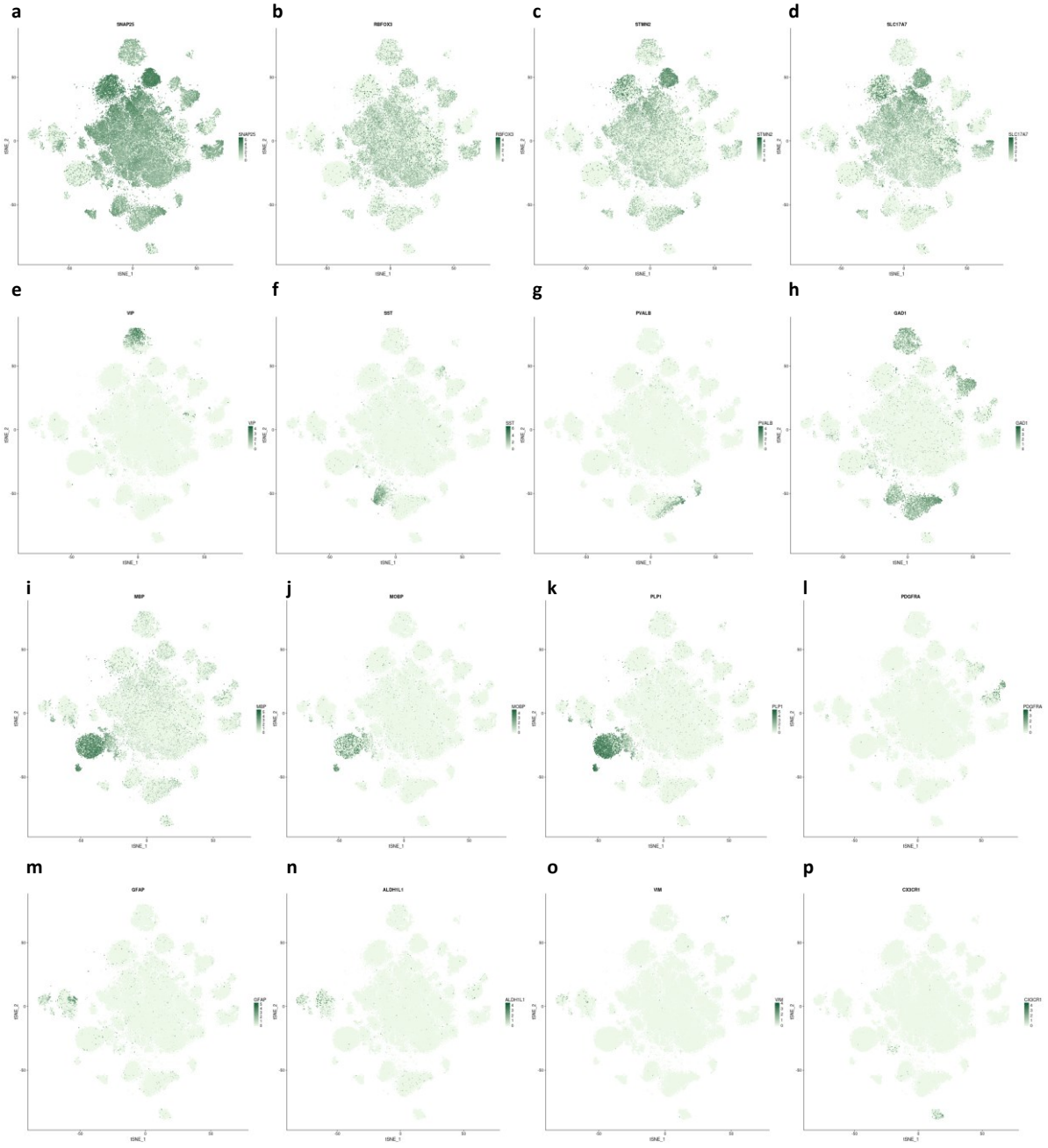
d



Supplementary Figure 2

Results from initial clustering analysis

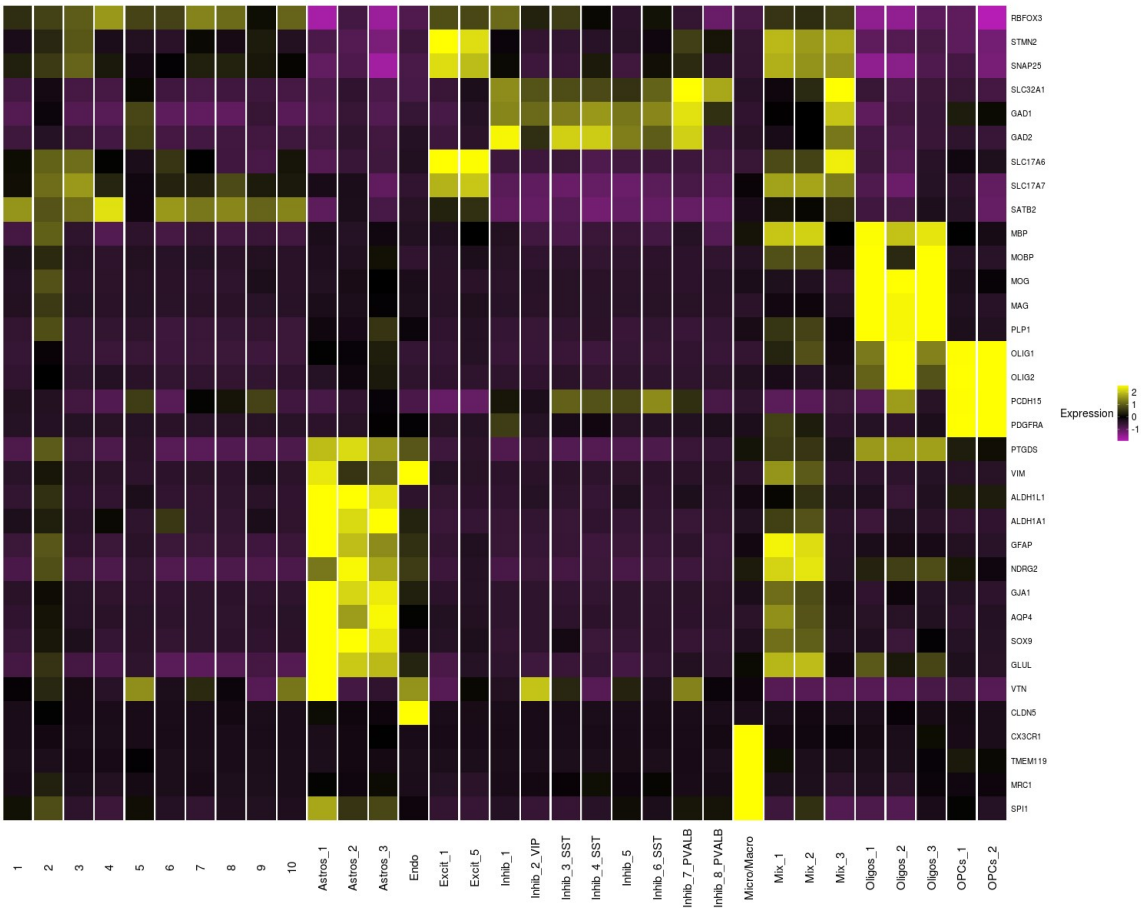
(a) Elbow plot showing the standard deviation of each of the 100 principal components calculated based on the top 2135 variable genes. First 50 principal components were used for downstream analysis. **(b)** Dendrograms depicting the relationships between clusters, created based on hierarchical clustering of the top 50 principle components, after every step of successive merging of the 43 initial clusters from unsupervised clustering (see Methods). **(c)** Unsupervised clustering followed by successive merging of similar clusters resulted in 30 clusters, n=78,886 individual nuclei from 34 subjects. **(d)** Average expression heatmap of cell-type marker genes (Macrophage/ Microglia: *SPI1, MRC1, TMEM119, CX3CR1*; Endothelial: *CLDN5, VTN, VIM*; Astrocytes: *GLUL, SOX9, AQP4, GJA1, NDRG2, GFAP, ALDH1A1, ALDH1L1, VIM*; OPCs: *PTGDS, PDGFRA, PCDH15, OLIG2, OLIG1*; Oligodendrocytes: *OLIG1, OLIG2, PLP1, MAG, MOG, MOBP, MBP*; Excitatory neurons: *SATB2, SLC17A7, SLC17A6*; Inhibitory neurons: *GAD1, GAD2, SLC32A1*; Neurons: *SNAP25, STMN2, RBFOX3*). Average marker gene expression in the 30 clusters was used for preliminary annotation.



Supplementary Figure 3

Individual TSNE plots representing the expression of various neuronal **(a-h)** and non-neuronal **(i-p)** cell type marker in a given cluster (n=78,886 individual nuclei from 34 subjects).

a



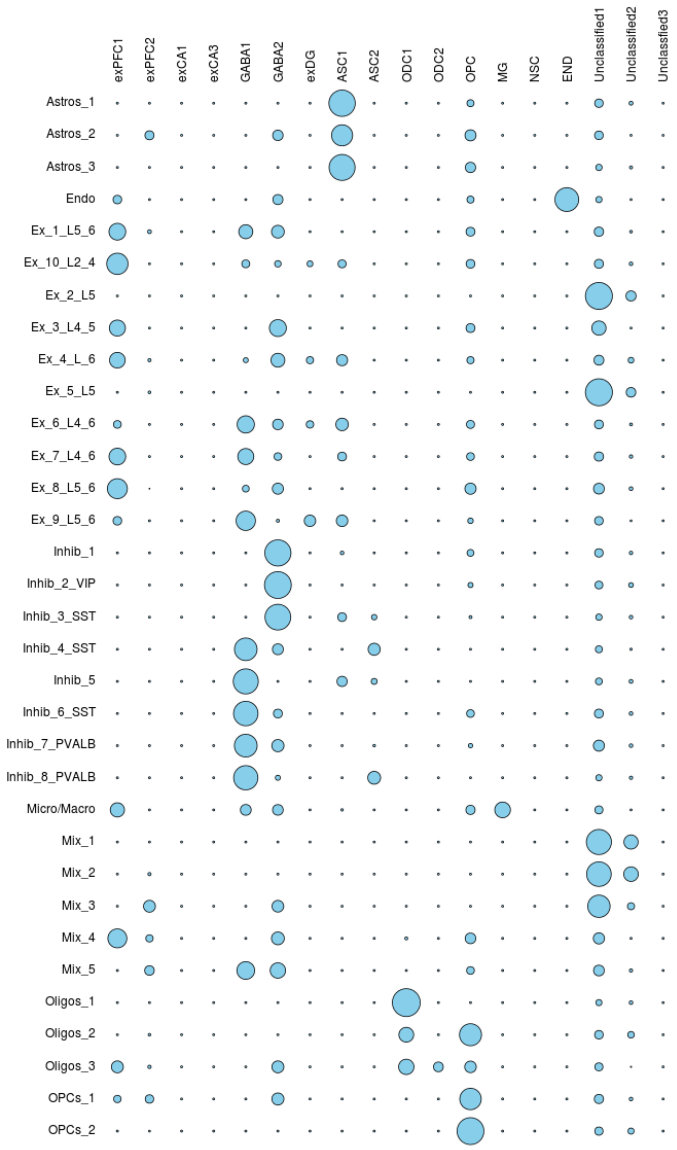
b

	Astros_1	Astros_2	Astros_3	Oligos_1	Oligos_2	OPCs_1	OPCs_2	Mix_1	Ex_2_L5	Mix_2	Mix_3	Ex_3_L4_5	Ex_5_L5	Endo	Micro/Macro	Oligos_3	Inhib_1	Inhib_2_VIP	Inhib_3_SST	Ex_1_L5_6	Ex_4_L_6	Ex_8_L5_6	Ex_6_L4_6	Ex_7_L4_6	Mix_4	Ex_9_L5_6	Mix_5	Ex_10_L2_4	Inhib_4_SST	Inhib_5	Inhib_6_SST	Inhib_7_PVALB	Inhib_8_PVALB
SNAP25	1.76	0	0	0	0	0	1.57	3.93	4.25	3.41	3.4	3.31	3.95	0	0	0	2.54	1.95	2.01	2.69	2.66	2.74	2.48	2.72	2.85	2.64	2.1	2.49	2.64	2	2.59	2.85	2.11
GLUL	3.3	1.63	2.51	1.8	0	0	0	2.11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SOX9	1.43	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AQP4	2.64	0	1.54	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GJA1	2.8	0	1.58	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NDRG2	1.85	2.14	2.08	0	0	0	0	2.49	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CLDN5	0	0	0	0	0	0	0	0	0	0	0	0	0	2.04	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
VTN	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SATB2	0	0	0	0	0	0	0	0	0	0	0	1.32	0	0	0	0	0	0	0	1.55	1.94	1.44	1.55	1.35	0	1.21	0	1.39	0	0	0	0	0	0
SLC17A7	0	0	0	0	0	0	0	2.23	2.44	0	0	2	0	0	0	0	0	0	0	0.92	1.09	1.42	1.22	1.19	0.93	1	0	0.75	0	0	0	0	0	0
SLC17A6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GAD1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.7	1.37	1.62	0	0	0	0	0	0	0	0	0	1.81	1.64	1.69	2.39	0
GAD2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.89	0	1.62	0	0	0	0	0	0	0	0	0	1.5	1.2	0	1.78	0
SLC32A1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SP1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MRC1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TMEM119	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CX3CR1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PLP1	0	0	0	4.12	2.93	0	0	0	0	0	0	0	0	0	0	3.25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MAG	0	0	0	1.67	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MOG	0	0	0	1.82	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MOBP	0	0	0	2.74	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MBP	0	0	0	3.75	3.11	0	0	3.01	0	3.06	0	0	0	0	0	3.28	0	0	0	0	0	0	0	0	1.26	0	0	0	0	0	0	0	0	0
PTGDS	3.34	3.26	2.89	2.92	2.74	0	0	1.39	0	0	0	0	0	0	0	2.94	0	0	0	0	0	0	0	0	1.38	0	0	0	0	0	0	0	0	0
PDGFRA	0	0	0	0	0	0	1.45	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PCDH15	0	0	0	0	2.27	3.3	3.78	0	0	0	0	0	0	0	0	0	0.79	0	1.9	0	0	0.62	0	0	0	0.81	0	0	1.59	1.38	2.24	1.23	0
OLIG1	0	0	0	0	0	0	1.83	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
OLIG2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Supplementary Figure 4

Cluster annotation based on marker genes

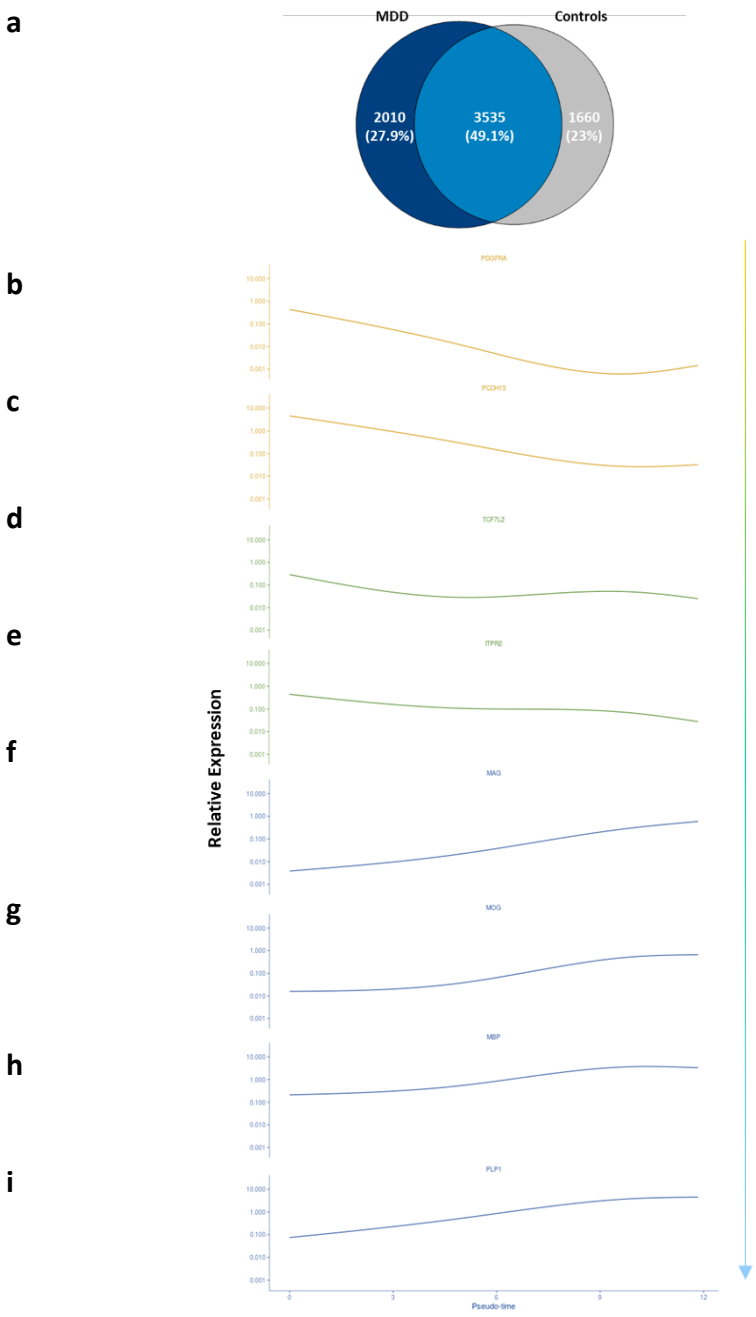
(a) Heatmap showing average expression of cell-type marker genes (as in Supplementary Fig. 2d) in the final 33 clusters. Numbers 1-10 correspond to the clusters obtained after reclustering the excitatory clusters. **(b)** Table showing the median expression of cell-type marker genes (darker red box corresponds to higher median expression) in the 33 final and fully annotated clusters.



Supplementary Figure 5

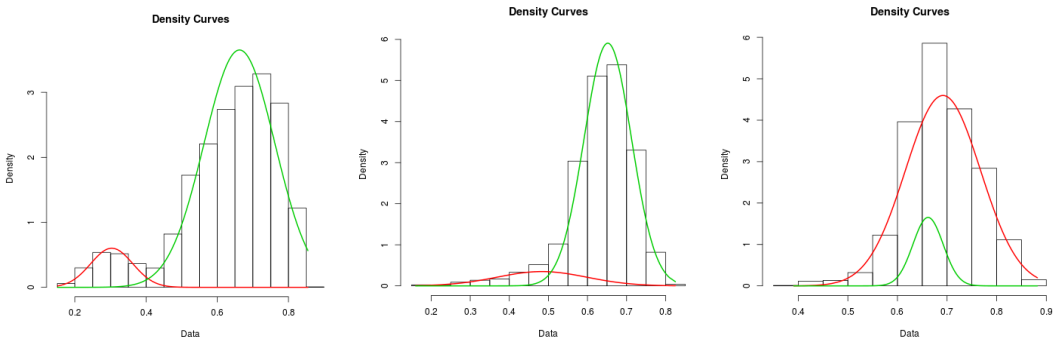
Cellular deconvolution analysis

Cell-type deconvolution was performed using the average gene expression in each of our clusters based on a previously published data from a similar sample set (Habib, N.*, et al. Nature Method* **14**, 955 (2017)). We found that cluster composition followed expected patterns across experiments. Given that we had a larger sample set, we were able to resolve more cell types than were identified in the human PFC in this dataset.



Supplementary Figure 6

(a) Approximately half of the genes that varied across pseudotime were shared between cases and controls. **(b-i)** Expression across pseudotime of **(b-c)** genes known to be highly expressed in OPCs or immature oligodendrocytes **(d-e)** transitionary, or **(f-i)** highly expressed in mature oligodendrocytes.



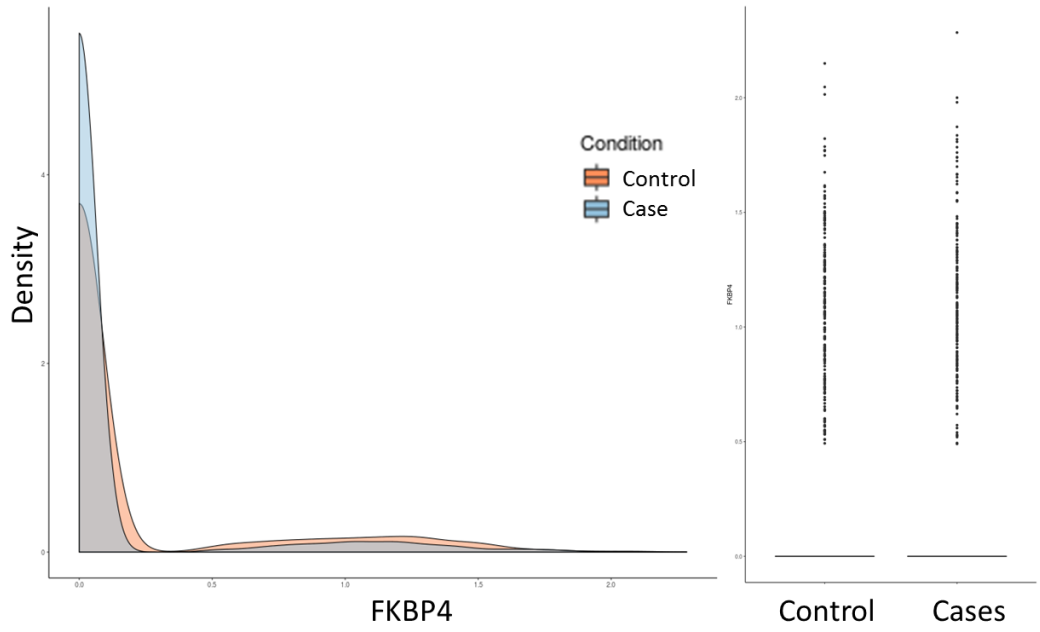
Supplementary Figure 7

Cell-filtering for cluster purification

Sample histograms of correlation between the gene expression profile of a cell and the median expression profile of its cluster (using highly variable genes) were plotted and normal distributions were fitted using mxmdl (see Methods). Cells with lower correlation and thus a higher probability of belonging to the red distribution were excluded, while cells with a higher probability of belonging to the green distribution were included. For clusters such as Ex_3_L4_5 (left) this removed lowly correlated cells. For clusters such as Ex_6_L4_6 (middle), where most cells were highly correlated to the median, not many cells needed to be removed. For clusters such as Ex_2_L5 (right), the two distributions overlapped and none of the cells were removed.

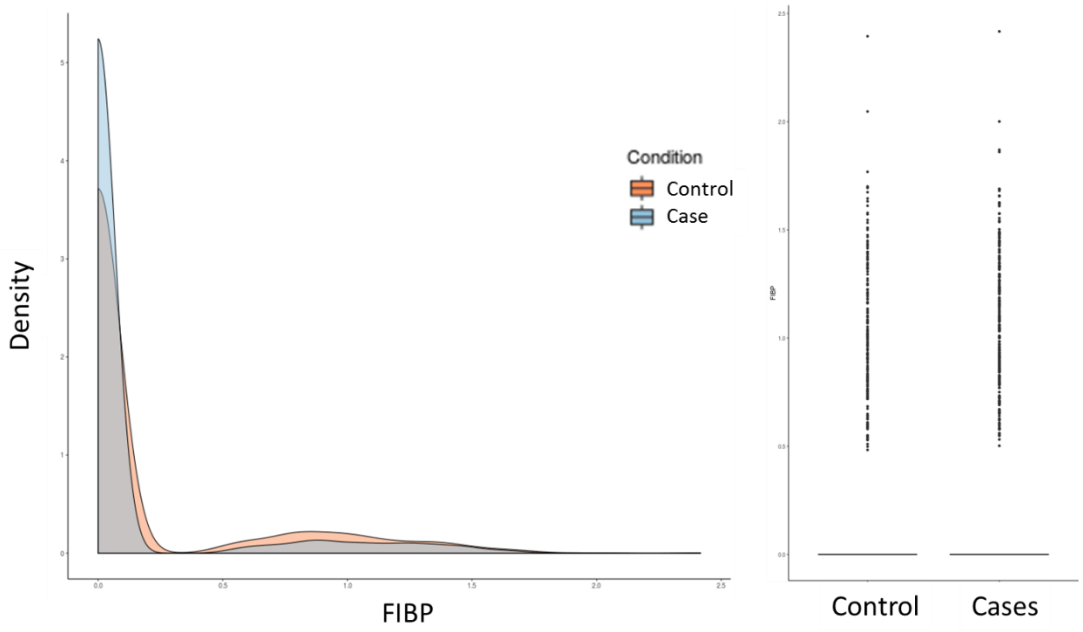
a

Ex_7

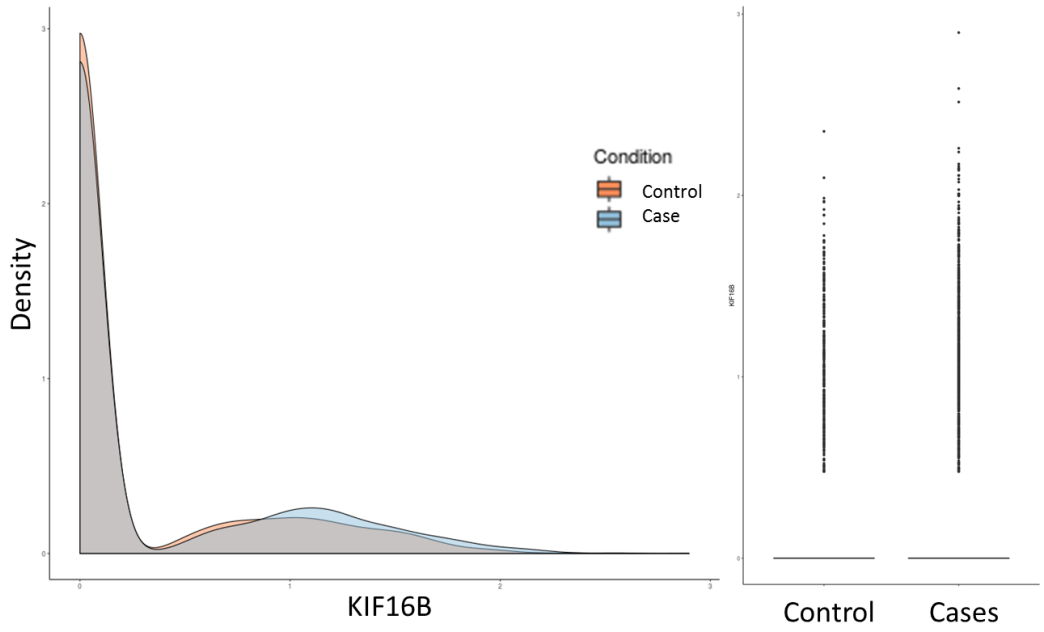

Condition
Control
Case
Density
FKBP4
Control
Cases

b

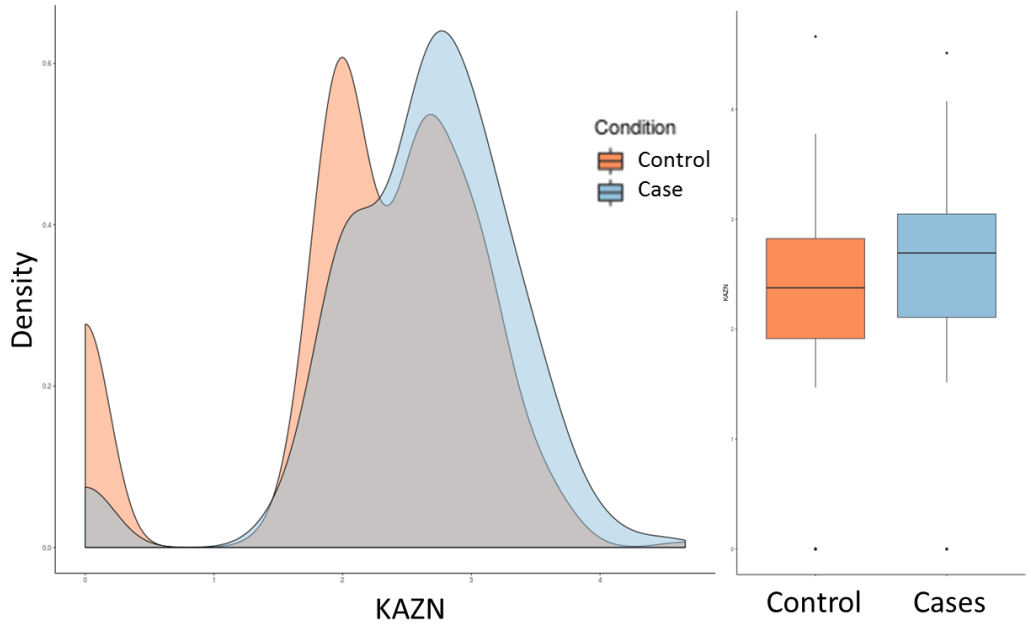
Ex_7


Condition
Control
Case
Density
FIBP
Control
Cases

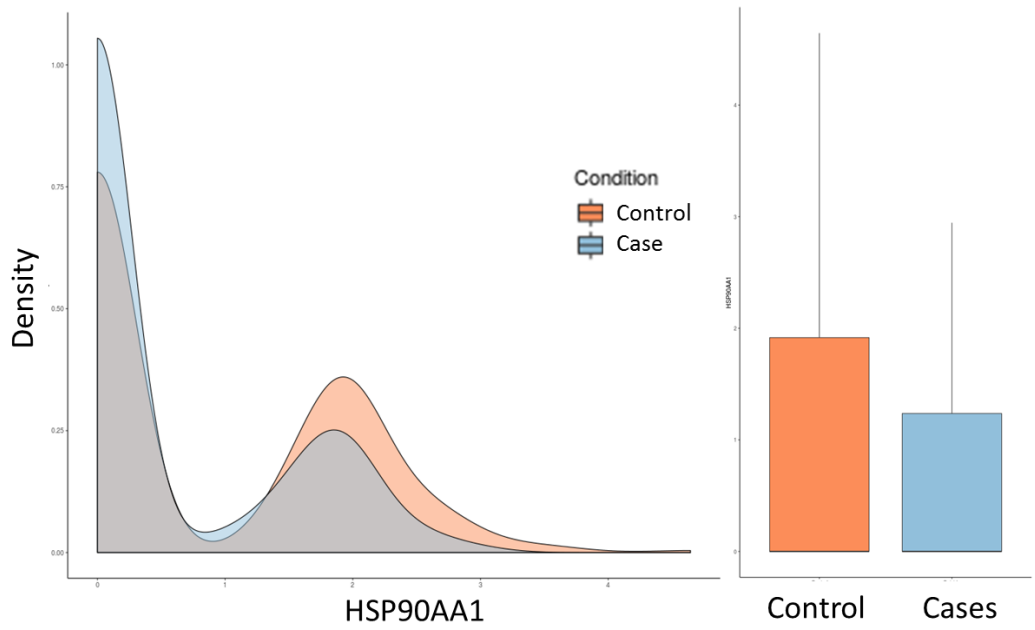
c Ex_7


Condition
Control
Case
Density
KIF16B
Control
Cases

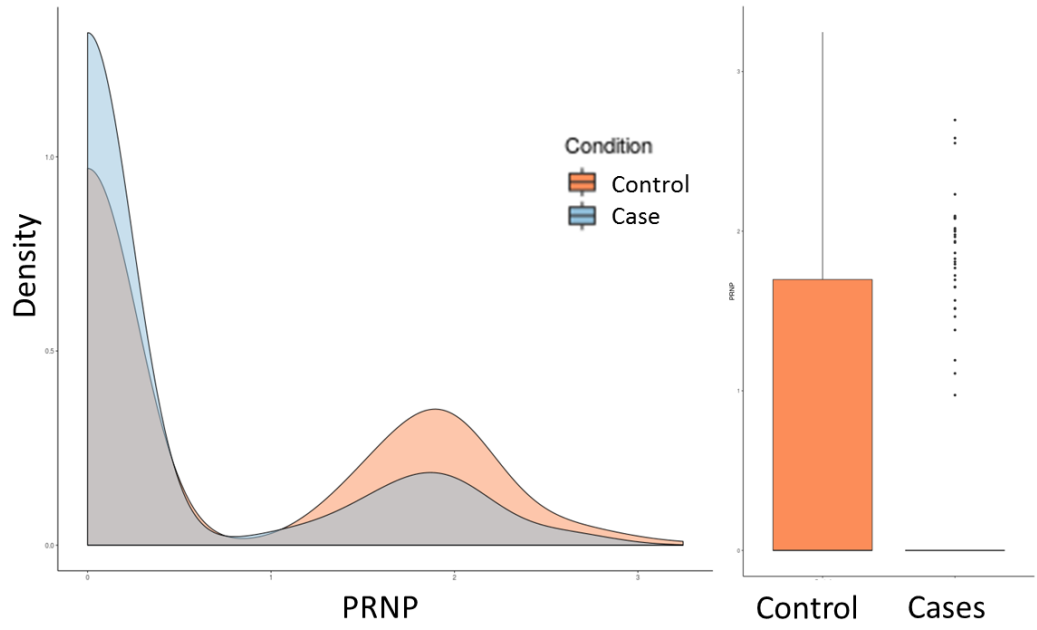
d OPC2


Condition
Control
Case
Density
KAZN
Control
Cases

e



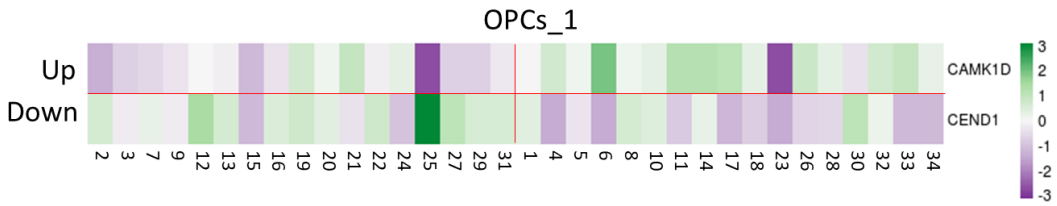
f



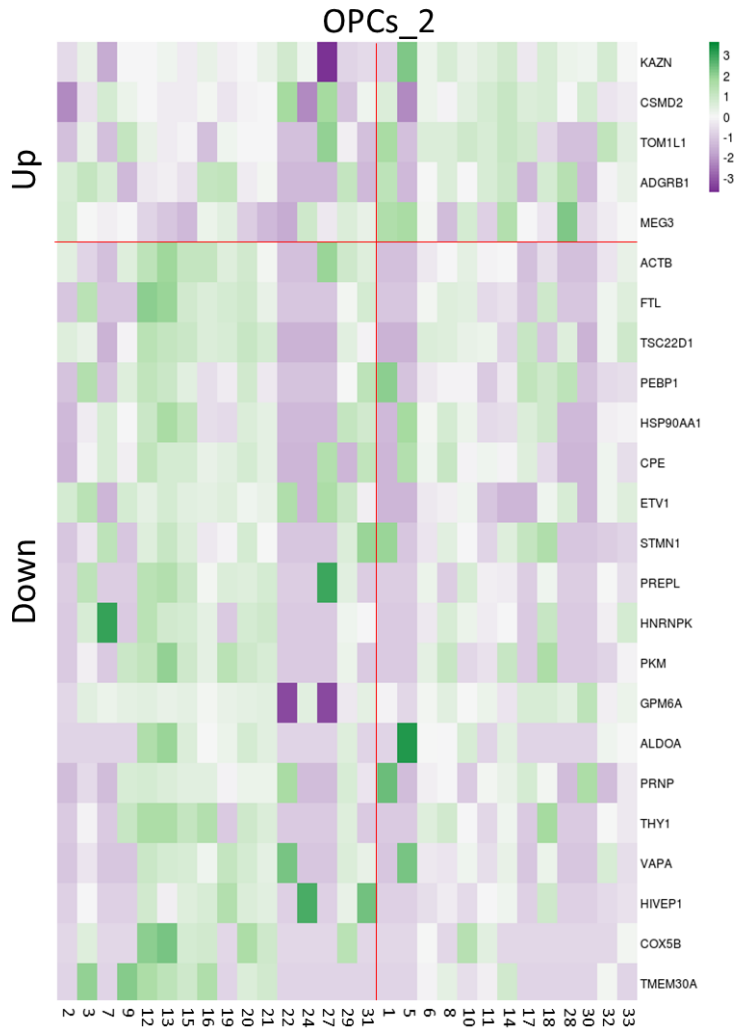
Supplementary Figure 8

Density plots (left) and boxplots (right) showing distribution of normalized expression values in single-nuclei for **(a)** *FKBP4*, **(b)** *FIBP*, and **(c)** *KIF16B* in Ex7 (n=1201 individual nuclei from 15 controls and individual 2170 nuclei from 17 cases) and for **(d)** *KAZN*, **(e)** *HSP90AA1*, and **(f)** *PRNP* in OPC2 (n=312 individual nuclei from 16 controls, and 164 individual nuclei from 13 cases). Only nuclei retained in purified cluster are represented. Controls in orange, cases in blue. For the boxplots, the lower and upper hinges correspond to the 25th and 75th percentiles respectively. The upper and lower whiskers extend from the upper hinge to the largest value and from the lower hinge to the smallest value respectively, and no further from the hinge than 1.5 times the interquartile range. Data points beyond this range are plotted individually. The central line is the median.

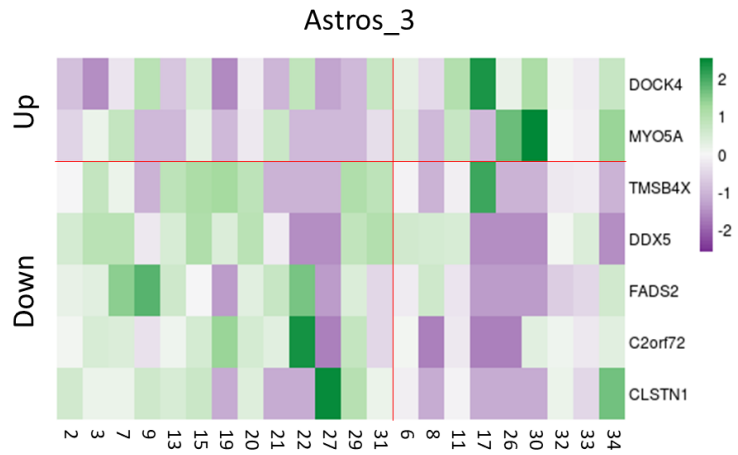
a


OPCs_1
Up
Down
CAMK1D
CEND1
3
2
1
0
-1
-2
-3
2 3 7 9 12 13 15 16 19 20 21 22 24 25 27 29 31 1 4 5 6 8 10 11 14 17 18 23 26 28 30 32 33 34

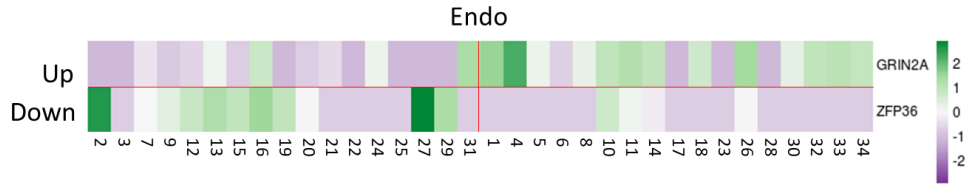
b


OPCs_2
Up
Down
KAZN
CSMD2
TOM1L1
ADGRB1
MEG3
ACTB
FTL
TSC22D1
PEBP1
HSP90AA1
CPE
ETV1
STMN1
PREPL
HNRNPK
PKM
GPM6A
ALDOA
PRNP
THY1
VAPA
HIVEP1
COX5B
TMEM30A
3
2
1
0
-1
-2
-3
2 3 7 9 12 13 15 16 19 20 21 22 24 27 29 31 1 5 6 8 10 11 14 17 18 28 30 32 33

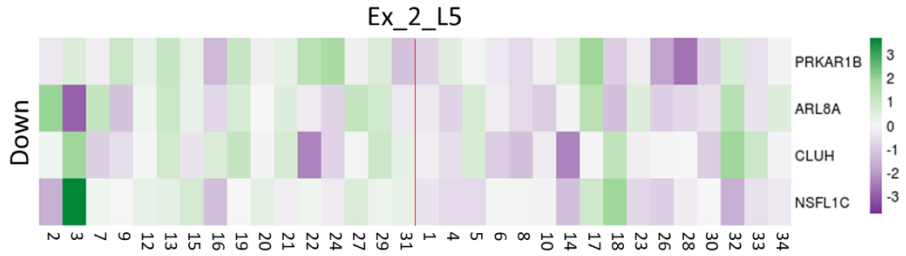
c


Astros_3
Up
Down
DOCK4
MYO5A
TMSB4X
DDX5
FADS2
C2orf72
CLSTN1
2
1
0
-1
-2
2 3 7 9 13 15 19 20 21 22 27 29 31 6 8 11 17 26 30 32 33 34

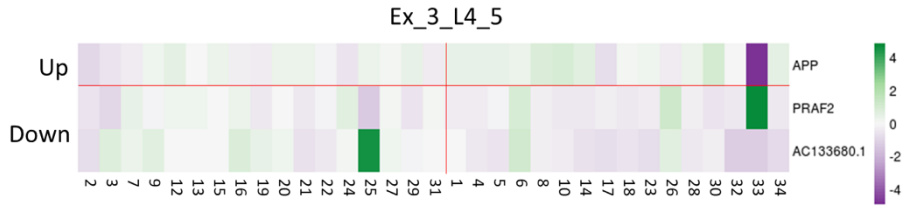
d


Endo
Up
Down
GRIN2A
ZFP36
2
1
0
-1
-2
2 3 7 9 12 13 15 16 19 20 21 22 24 25 27 29 31 1 4 5 6 8 10 11 14 17 18 23 26 28 30 32 33 34

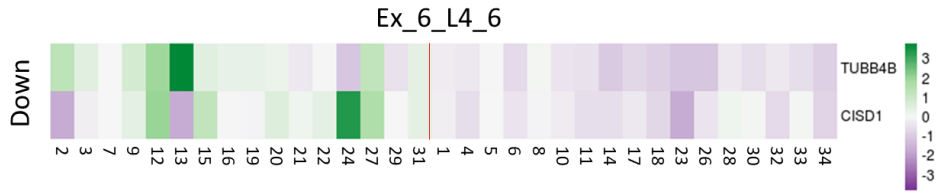
e


Ex_2_L5
Down
PRKAR1B
ARL8A
CLUH
NSFL1C
3
2
1
0
-1
-2
-3
2 3 7 9 12 13 15 16 19 20 21 22 24 27 29 31 1 4 5 6 8 10 14 17 18 23 26 28 30 32 33 34

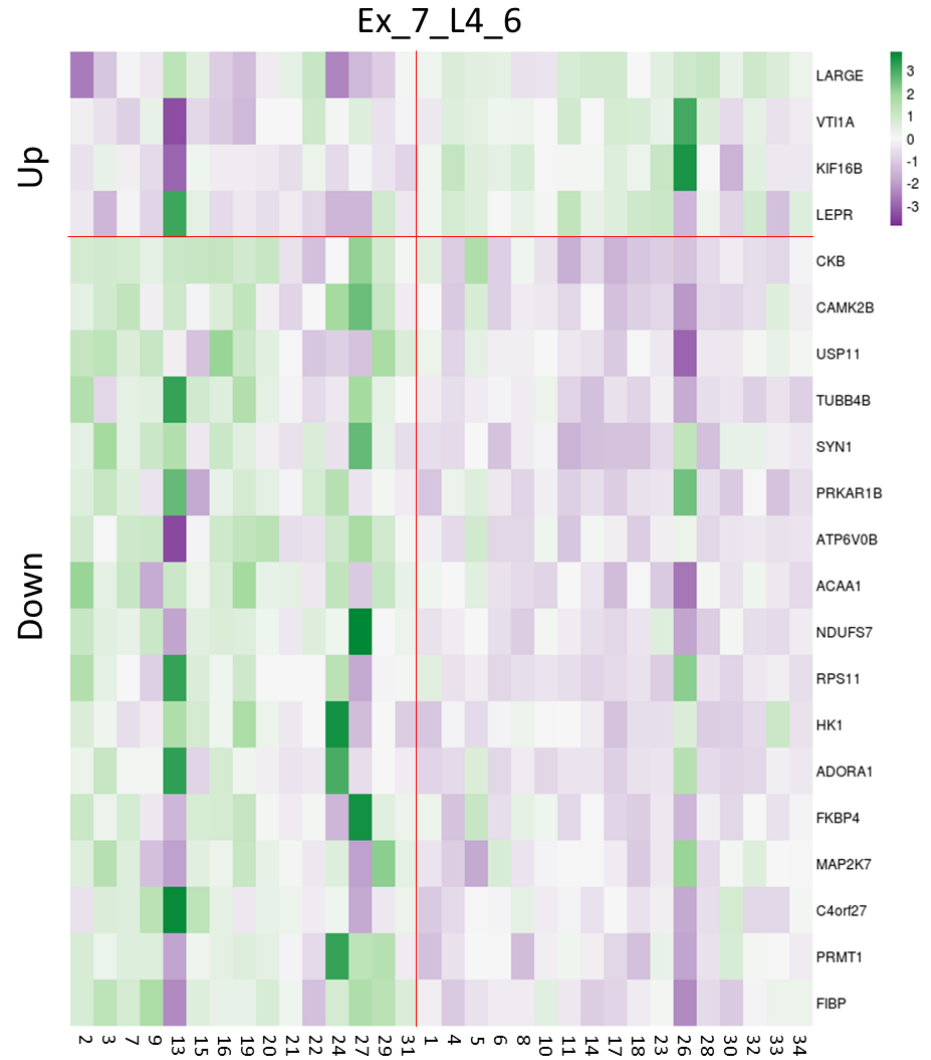
f


Ex_3_L4_5
Up
Down
APP
PRAF2
AC133680.1
4
2
0
-2
-4
2 3 7 9 12 13 15 16 19 20 21 22 24 25 27 29 31 1 4 5 6 8 10 14 17 18 23 26 28 30 32 33 34

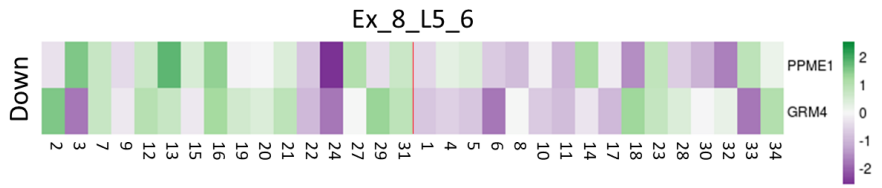
g


Ex_6_L4_6
Down
TUBB4B
CISD1
3
2
1
0
-1
-2
-3
2 3 7 9 12 13 15 16 19 20 21 22 24 27 29 31 1 4 5 6 8 10 11 14 17 18 23 26 28 30 32 33 34

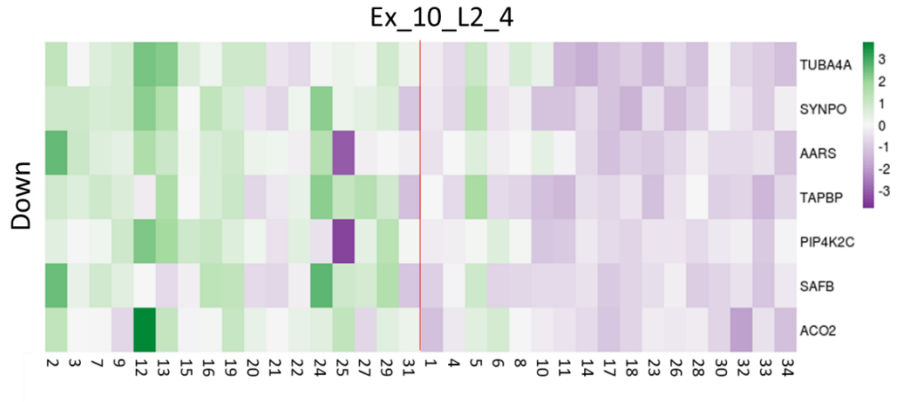
h


Ex_7_L4_6
Up
Down
LARGE
VTI1A
KIF16B
LEPR
CKB
CAMK2B
USP11
TUBB4B
SYN1
PRKAR1B
ATP6V0B
ACAA1
NDUFS7
RPS11
HK1
ADORA1
FKBP4
MAP2K7
C4orf27
PRMT1
FIBP
3
2
1
0
-1
-2
-3
2 3 7 9 13 15 16 19 20 21 22 24 27 29 31 1 4 5 6 8 10 11 14 17 18 23 26 28 30 32 33 34

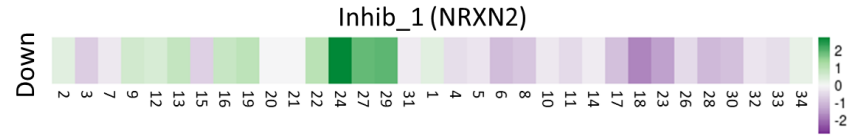
i


Ex_8_L5_6
Down
PPME1
GRM4
2
1
0
-1
-2
2 3 7 9 12 13 15 16 19 20 21 22 24 27 29 31 1 4 5 6 8 10 11 14 17 18 23 28 30 32 33 34

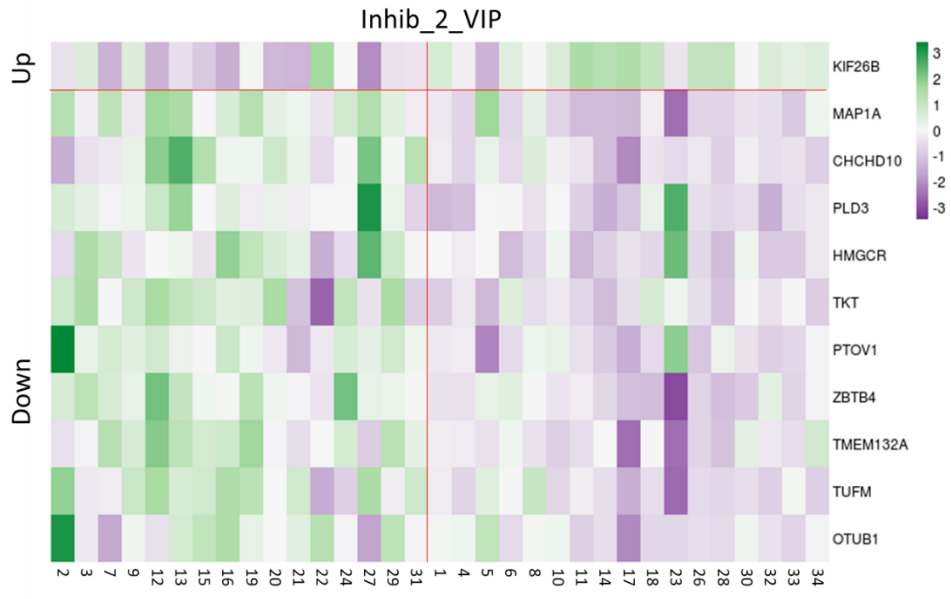
j


Ex_10_L2_4
Down
TUBA4A
SYNPO
AARS
TAPBP
PIP4K2C
SAFB
ACO2
3
2
1
0
-1
-2
-3
2 3 7 9 12 13 15 16 19 20 21 22 24 25 27 29 31 1 4 5 6 8 10 11 14 17 18 23 26 28 30 32 33 34

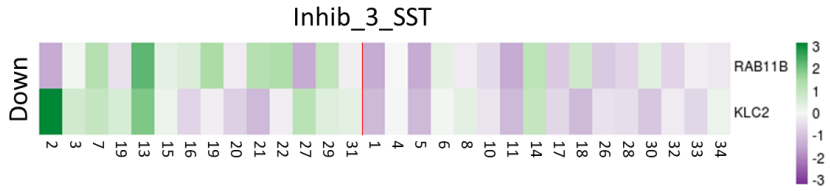
k


Inhib_1 (NRXN2)
Down
2
1
0
-1
-2
2 3 7 9 12 13 15 16 19 20 21 22 24 27 29 31 1 4 5 6 8 10 11 14 17 18 23 26 28 30 32 33 34

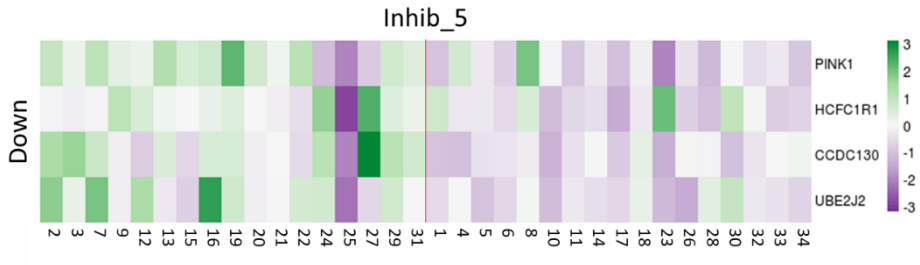
l


Inhib_2_VIP
Up
Down
KIF26B
MAP1A
CHCHD10
PLD3
HMGCR
TKT
PTOV1
ZBTB4
TMEM132A
TUFM
OTUB1
3
2
1
0
-1
-2
-3
2 3 7 9 12 13 15 16 19 20 21 22 24 27 29 31 1 4 5 6 8 10 11 14 17 18 23 26 28 30 32 33 34

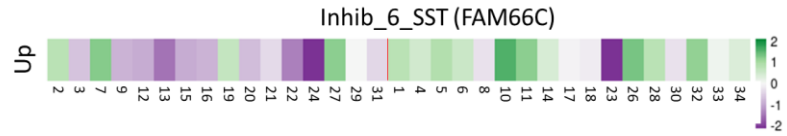
m



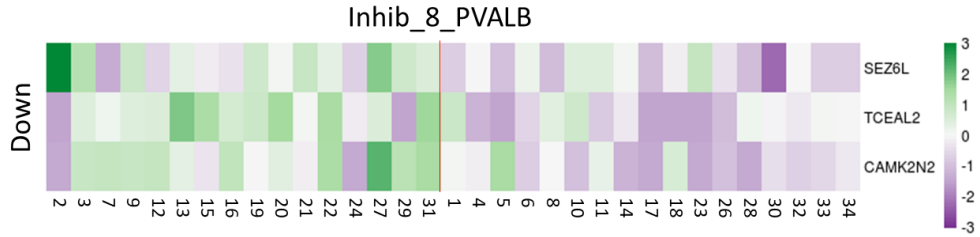
n



o



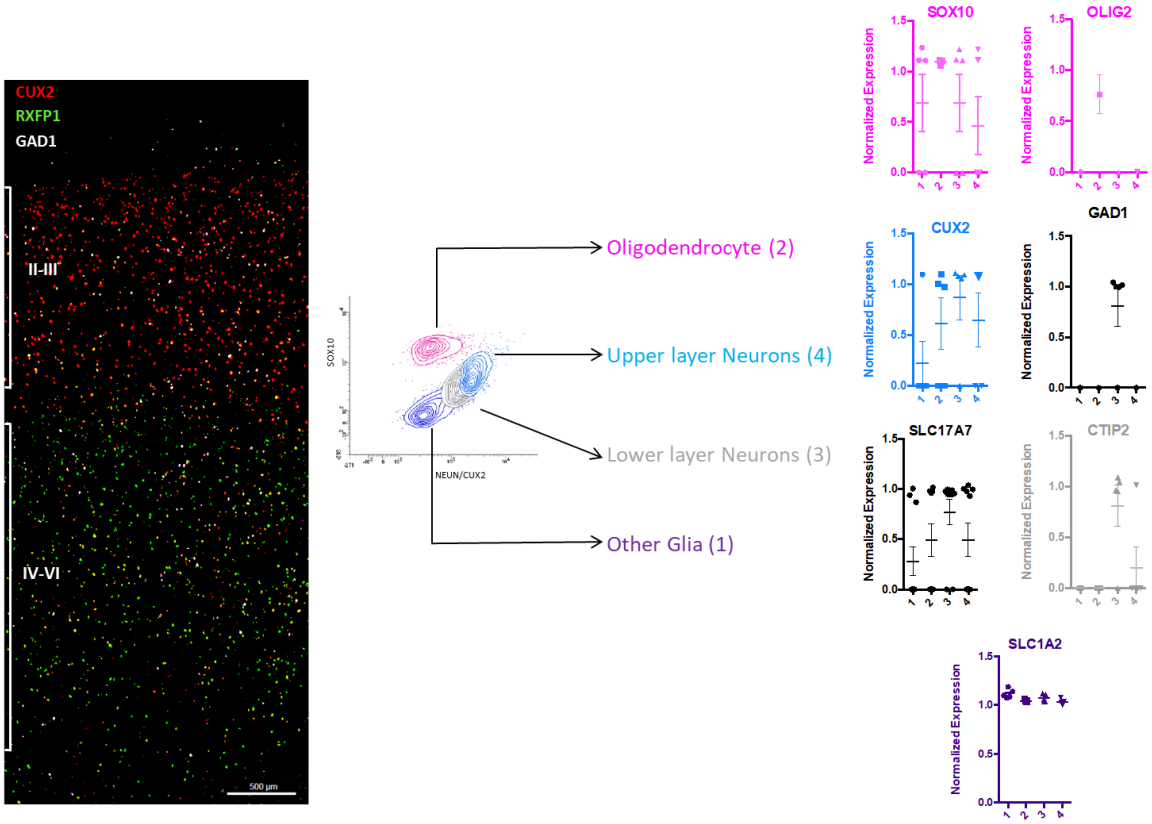
p



Supplemental Figure 9

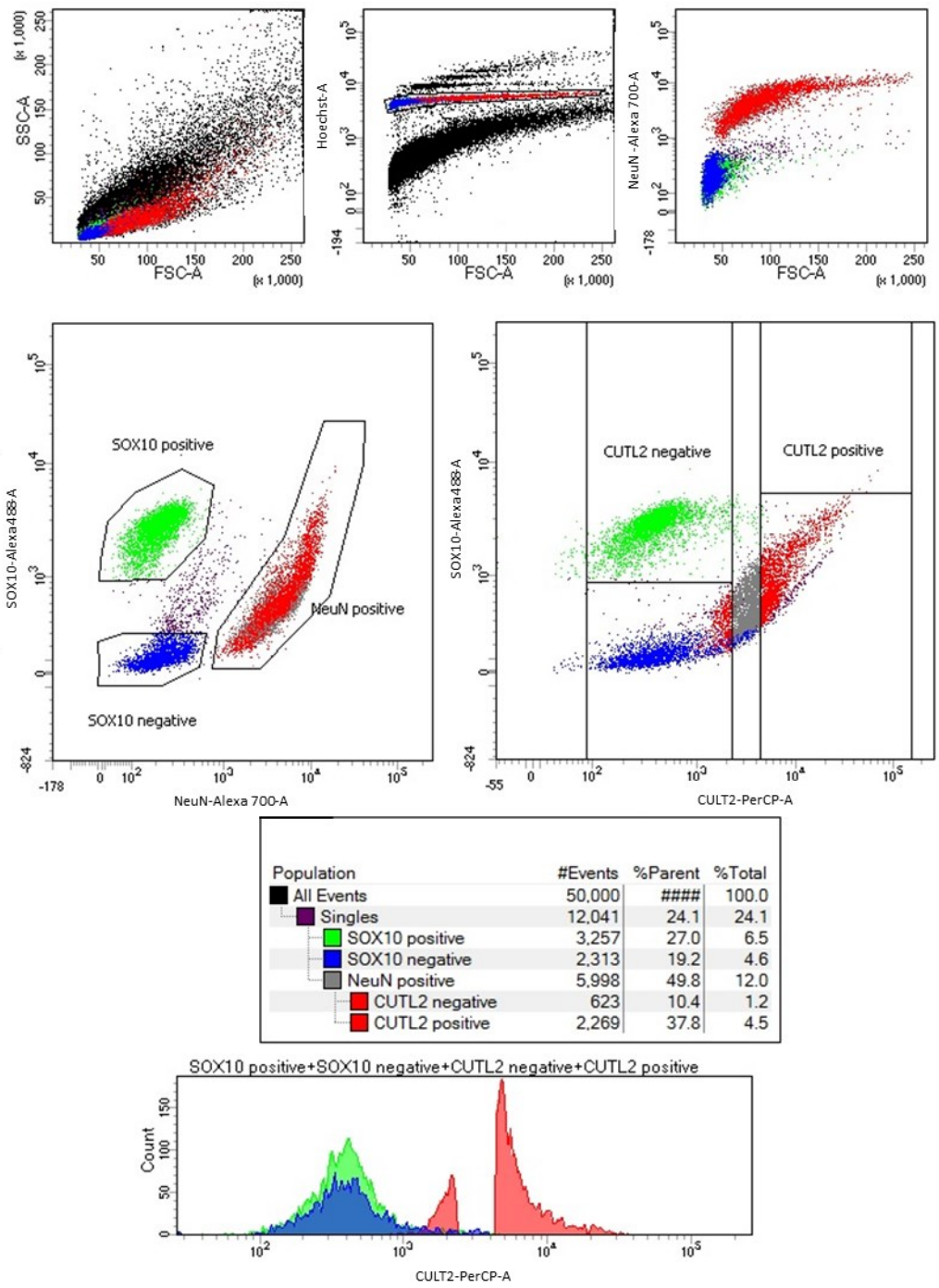
Differential expression heatmaps

(a-p) Heatmaps of average expression for differentially expressed genes per subject in each cluster (controls on the left, cases on the right) showing significant differential expression.



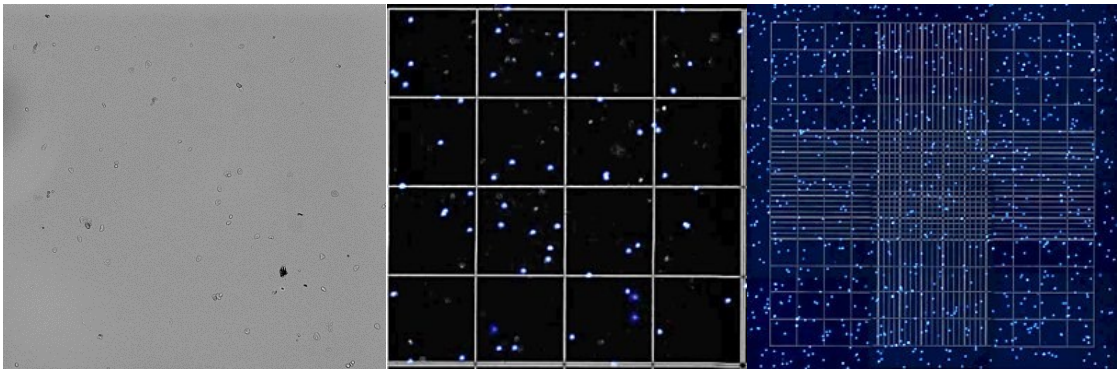
Supplementary Figure 10

Summary of validations results from high-throughput qPCR. (Left) RNAScope in situ hybridization image showing expression of *CUX2, RXFP1,* and *GAD1* in the dlPFC of one subject from our cohort. The experiment was performed once to confirm the pattern of *CUX2* expression. (Top middle) Representative FANS plot from 34 sorts performed for 34 subjects depicting separation of nuclei (refer to Supplementary Fig. 12). (Top right) qPCR validation of expected marker gene expression to verify FANS separation (error bars represent ± S.E.M). *SOX10* (n = 5 technical replicates), *OLIG2* (n = 5 technical replicates), *CUX2* (n = 5 technical replicates), and *SLC1A2* (n = 5 technical replicates) all show highest expression in the expected population. *GAD1* (n = 5 technical replicates), *SLC17A7* (n = 10 technical replicates), and *CTIP2* (n = 5 technical replicates) are expressed more highly in the neuronal population.



Supplementary Figure 11

Representative plots from one subject illustrating the gating strategy for FANS as described in detail in the Methods.

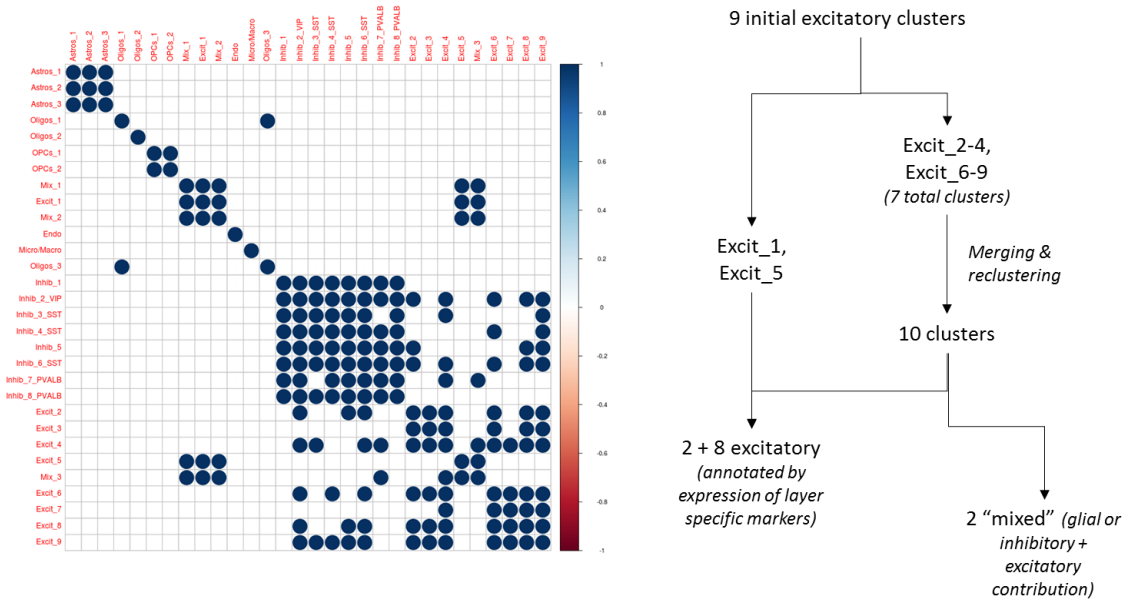


Supplementary Figure 12

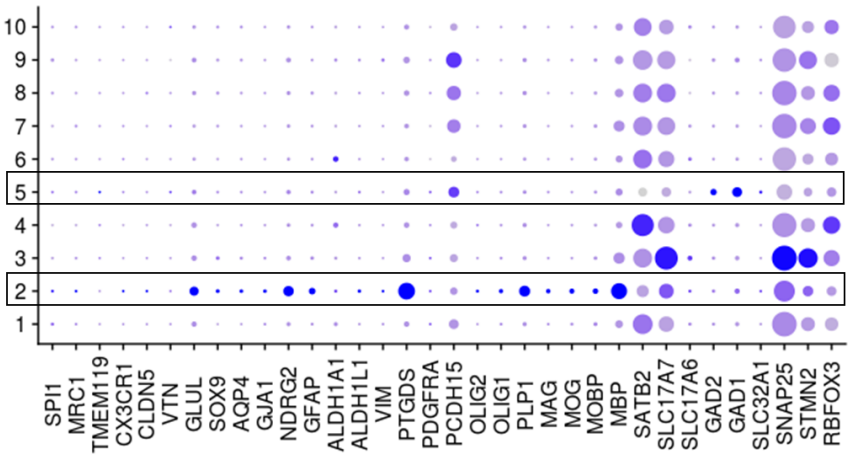
Microscope images of nuclei after extraction with brightfield and with Hoechst staining for DNA.

Representative image from 36 samples prepared for 34 subjects.

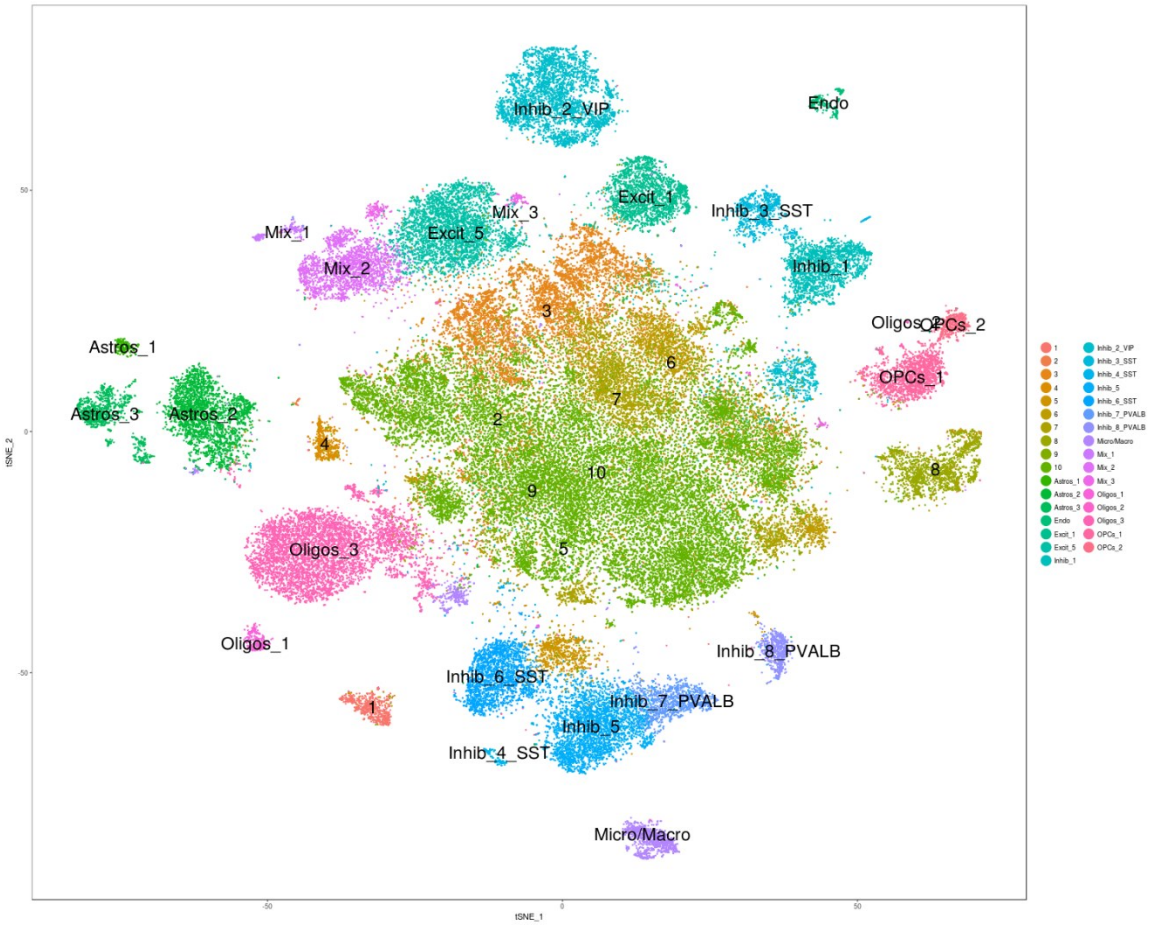
a



b



c

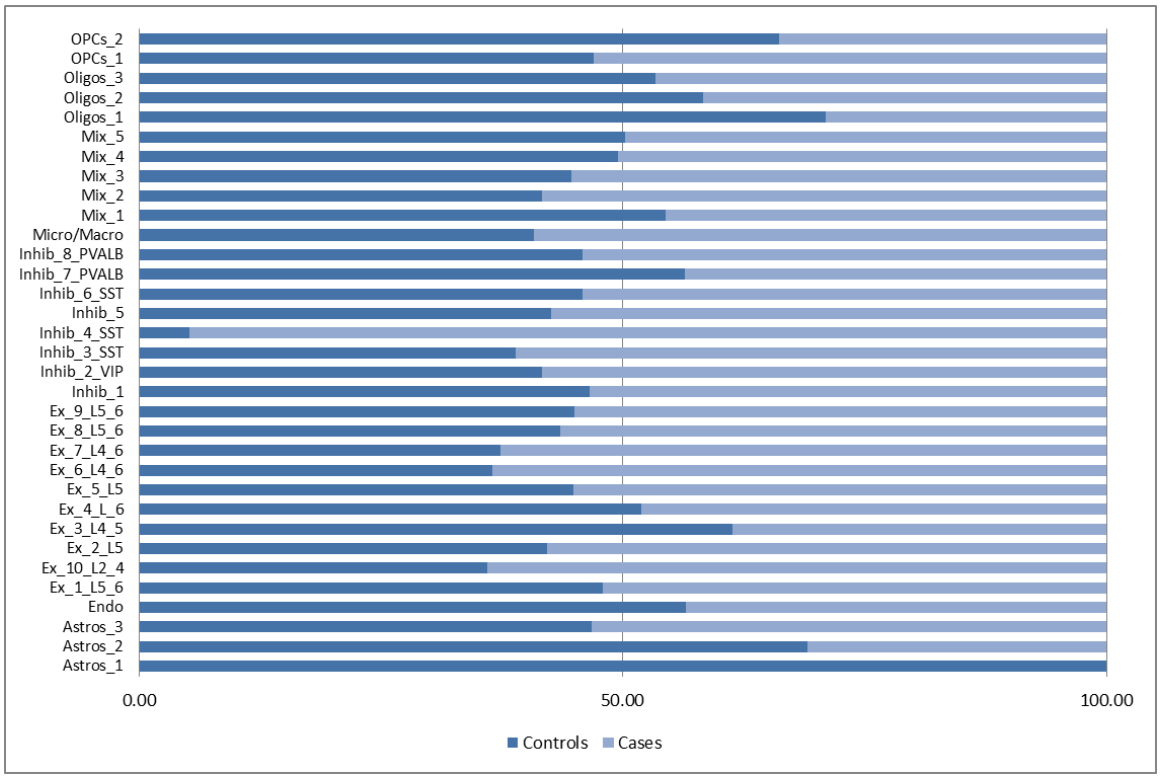


Supplementary Figure 13

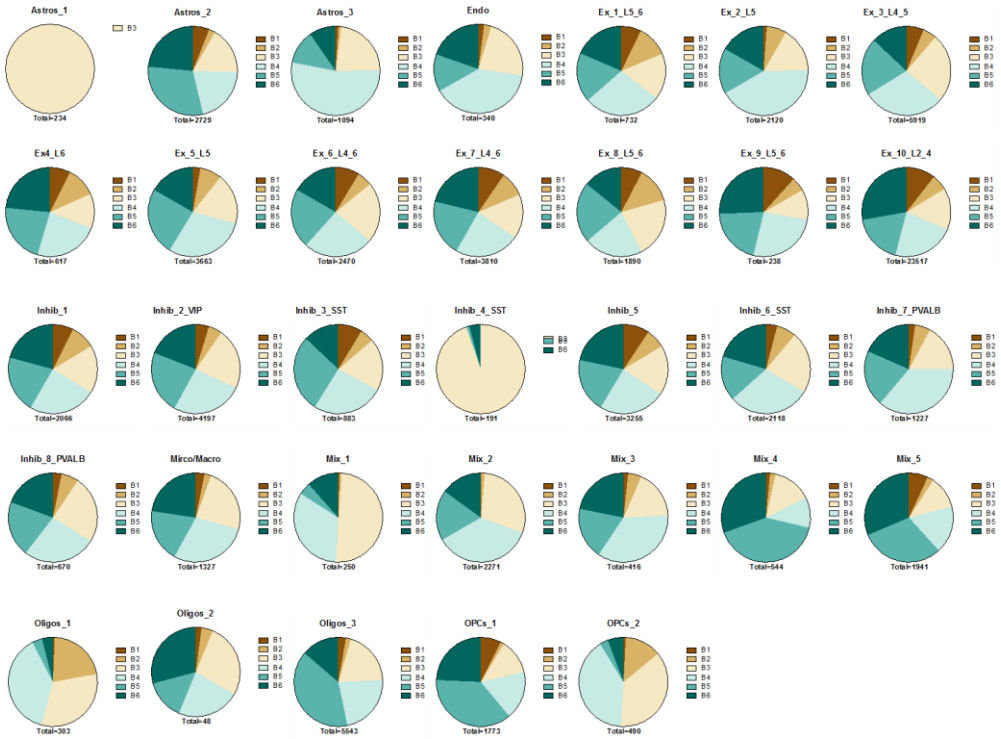
Reclustering of excitatory clusters

(a) (Left) Pearson's correlation between gene expression (n= 30062 genes) in pairs of clusters. Only the correlation coefficients > 0.95 are shown. Similar cell-types show this high-level of correlation. (Right) Flowchart clarifying the relationship between the 9 initial and 10 final excitatory clusters. Based on gene expression correlation patterns cells in clusters Excit_2-4 and Excit_6-9 were re-clustered, while cells in cluster Excit_1 and Excit_5 were not. **(b)** DotPlot showing expression of cell-type marker genes (as in Supplementary Figure 2d) in the 10 clusters obtained after reclustering. The highlighted clusters were marked as mixed, because they expressed marker genes of multiple cell-types. **(c)** TSNE with 33 clusters resulting after re-clustering of selected excitatory clusters shown in Supplementary Fig. 2c into more refined excitatory subtypes (n= 78,886 individual nuclei from 34 subjects).

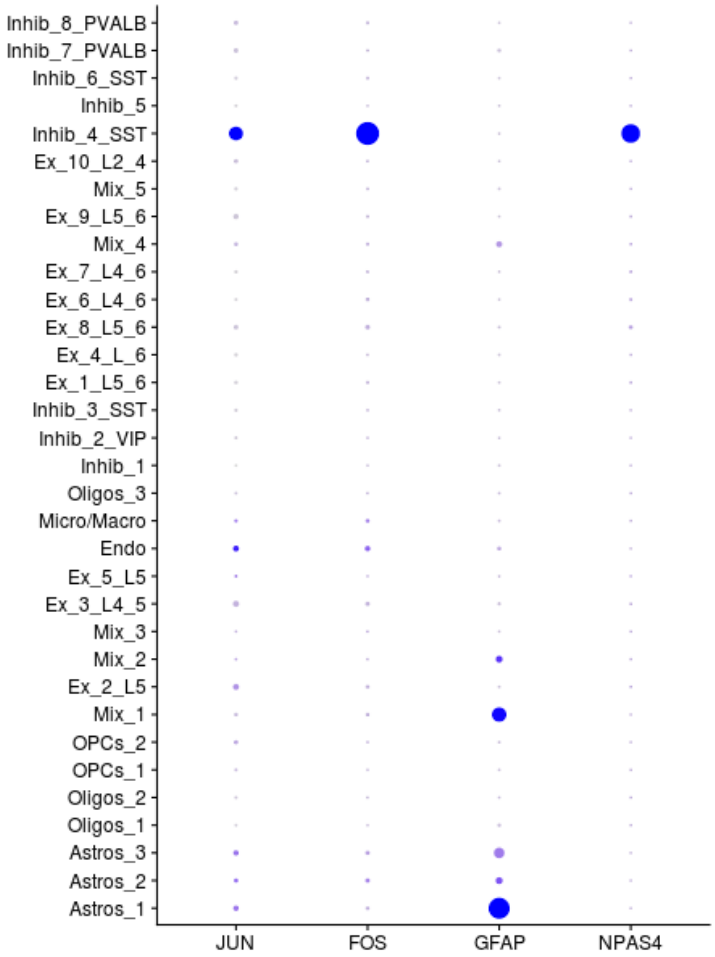
a



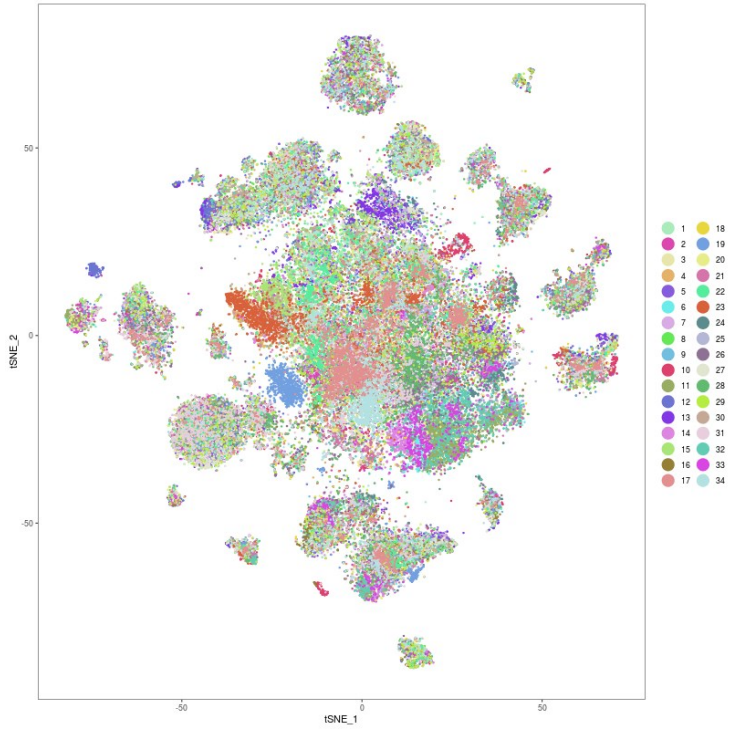
b


Astros_1
Total=234
Astros_2
Total=2729
Astros_3
Total=1094
Endo
Total=340
Ex_1_L5_6
Total=732
Ex_2_L5
Total=2120
Ex_3_L4_5
Total=5919
Ex4_L6
Total=617
Ex_5_L5
Total=3663
Ex_6_L4_6
Total=2470
Ex_7_L4_6
Total=3810
Ex_8_L5_6
Total=1890
Ex_9_L5_6
Total=238
Ex_10_L2_4
Total=23517
Inhib_1
Total=2066
Inhib_2_VIP
Total=4197
Inhib_3_SST
Total=883
Inhib_4_SST
Total=191
Inhib_5
Total=3255
Inhib_6_SST
Total=2118
Inhib_7_PVALB
Total=1227
Inhib_8_PVALB
Total=670
Mirco/Macro
Total=1327
Mix_1
Total=250
Mix_2
Total=2271
Mix_3
Total=416
Mix_4
Total=544
Mix_5
Total=1941
Oligos_1
Total=303
Oligos_2
Total=48
Oligos_3
Total=5543
OPCs_1
Total=1773
OPCs_2
Total=490
B1
B2
B3
B4
B5
B6

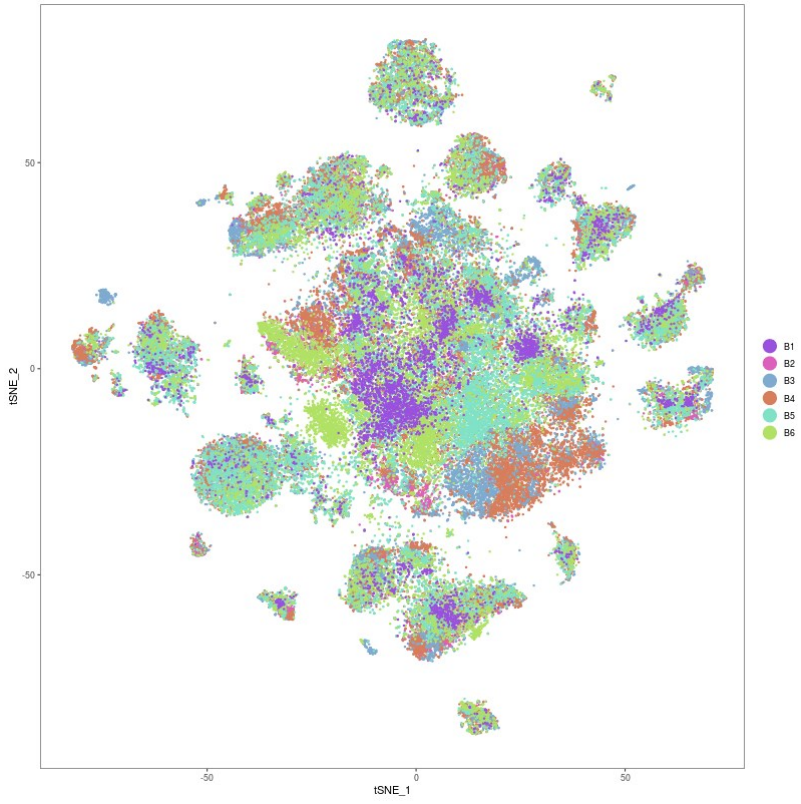
c


Inhib_8_PVALB
Inhib_7_PVALB
Inhib_6_SST
Inhib_5
Inhib_4_SST
Ex_10_L2_4
Mix_5
Ex_9_L5_6
Mix_4
Ex_7_L4_6
Ex_6_L4_6
Ex_8_L5_6
Ex_4_L_6
Ex_1_L5_6
Inhib_3_SST
Inhib_2_VIP
Inhib_1
Oligos_3
Micro/Macro
Endo
Ex_5_L5
Ex_3_L4_5
Mix_3
Mix_2
Ex_2_L5
Mix_1
OPCs_2
OPCs_1
Oligos_2
Oligos_1
Astros_3
Astros_2
Astros_1
JUN
FOS
GFAP
NPAS4

d


tSNE_2
tSNE_1
50
0
-50
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34

e


tSNE_2
tSNE_1
50
0
-50
B1
B2
B3
B4
B5
B6

Supplementary Figure 14

Quality control metrics for final clusters

(a) Contribution of each group (depressed cases versus controls) to a given cluster. In most cases there is a relatively even contribution. In highly skewed instances like cluster Astros_1 or Inhib_4_SST, these clusters were discarded. **(b)** The samples were processed in multiple batches. Batches 1 and 2 were relatively small batches with 2 samples each while batches 4-6 were full 8-sample captures (as much as can be processed by one 10x chip). In most cases all batches contributed to each cluster. In some cases, e.g. Astros_1 only one batch contributed and as a result, this cluster is removed from downstream analysis. **(c)** Clusters Inhib_4_SST and Astros_1 were excluded because they were primarily contributed by a single subject each. Inhib_4_SST specifically showed increased expression of activity dependent genes *NPAS4, JUN,* and *FOS*. Astros_1 showed highly increased expression of *GFAP* compared to other astrocytic clusters. **(d & e)** tSNE plots colored by subject and batch (n=78,886 individual nuclei from 34 subjects). Linear models examining the effect of batch and subject on the residuals of the first tSNE component (residualized on the second tSNE component) showed an R-squared value of <0.002 for batch and < 0.04 for subject, emphasizing the absence of prominent batch effects on clustering.

Chapter 3

Preface to Chapter 3

Encouraged by our findings of cell type specific gene expression differences associated with depression in our male cohort we extended our dataset with a cohort of female individuals. Our initial study focused on males to simplify experimental design in the application of a novel technology for profiling the human brain in depression and to facilitate comparison with previous literature which was biased towards males. Subsequently, given the considerable evidence for sex differences in depression-associated transcriptomic patterns we generated crucial complementary data in females, identified cell types with prominent depression-associated gene expression differences in females, re-evaluated our male dataset in parallel with improved contemporary analysis methodology, and finally performed statistical comparison of the findings between the sexes highlighting the points of similarity and difference, as described in the following chapter.

Cell type specific transcriptomic differences in depression show similar patterns between males and females but implicate distinct cell types and genes

Author list

Malosree Maitra[1,2,3], Haruka Mitsuhashi[1,2,3], Reza Rahimian[1,3], Anjali Chawla[1,2,3], Jennie Yang[1,3], Laura M. Fiori[1,3], Maria Antonietta Davoli[1,3], Kelly Perlman[1,2,3], Zahia Aouabed[1,3], Deborah C. Mash[4], Matthew Suderman[5,6], Naguib Mechawar[1,2,3], Gustavo Turecki[1,2,3*], Corina Nagy[1,2,3*]

* These authors jointly supervised this work

Affiliations

1- McGill Group for Suicide Studies, Douglas Institute, Verdun, QC, Canada

2- Integrated Program in Neuroscience, McGill University, Montreal, QC, Canada

3- Douglas Institute, Department of Psychiatry, McGill University, Montreal, QC, Canada

4- Dr. Kiran C. Patel College of Allopathic Medicine, Nova Southeastern University, Ft. Lauderdale, FL, USA

5- Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK

6- MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK

Corresponding authors

Dr. Gustavo Turecki (gustavo.turecki@mcgill.ca)

6875 LaSalle Blvd, Verdun, Québec, Canada H4H 1R3

Tel.: (514) 761-6131 (ext. 2370)

Dr. Corina Nagy (corina.nagy@mcgill.ca)

6875 LaSalle Blvd, Verdun, Québec, Canada H4H 1R3

Tel.: (514) 761-6131 (ext. 6160)

Abstract

Major depressive disorder (MDD) is a common, heterogenous, and potentially serious psychiatric illness. Diverse brain cell types have been implicated in MDD etiology. Significant sexual differences exist in MDD clinical presentation and outcome, and recent evidence suggests different molecular bases for male and female MDD. We evaluated over 160,000 nuclei from 71 female and male donors, leveraging new and pre-existing single-nucleus RNA-sequencing data from the dorsolateral prefrontal cortex. Cell type specific transcriptome-wide threshold-free MDD-associated gene expression patterns were similar between the sexes, but significant differentially expressed genes (DEGs) diverged. Among 7 broad cell types and 41 clusters evaluated, microglia and parvalbumin interneurons contributed the most DEGs in females, while deep layer excitatory neurons, astrocytes, and oligodendrocyte precursors were the major contributors in males. Further, the Mic1 cluster with 38% of female DEGs and the ExN10_L46 cluster with 53% of male DEGs, stood out in the meta-analysis of both sexes.

Keywords

major depressive disorder; sex differences; single-nucleus RNA-sequencing; dorsolateral prefrontal cortex; microglia; PV interneurons

Introduction

Major depressive disorder (MDD) is a serious and potentially debilitating mental illness affecting 200-300 million people worldwide[1]. MDD is a leading cause of disability globally[1] and some prominent symptoms in patients with MDD include persistent low mood, decreased interest and/or pleasure, sleep and appetite disturbances, feelings of worthlessness, and suicidal thoughts[2]. A number of genetic variants have been identified which contribute to the heritability of MDD[3] and brain transcriptomic differences[4] are detected in this disease, but the molecular etiology of MDD is still only partially understood.

There are known dissimilarities in the epidemiology and pathophysiology of MDD between the sexes. Notably, it is twice as prevalent in women than men[5]. Symptomatology differs in that, women are more likely to have comorbid anxiety, so-called atypical depression, and recurrent episodes, while men are more likely to have comorbid substance use disorders and to die by suicide[6-8]. Sex specific molecular profiles in MDD and corresponding animal models are often attributed to hormonal differences either during development or in adulthood, to the contributions of sex-chromosomes, or to inherent sex differences in the monoaminergic system or the hypothalamic-pituitary-adrenal axis (HPA), among other factors[6,9].

Recent studies in humans have attempted to address the gap in our knowledge of molecular sex differences in depression by examining MDD-associated sex-specific brain transcriptomic differences in human patients[7,10]. Using bioinformatic and meta-analysis approaches, combined with validation in animal models, these studies found that, overall, MDD-associated differences in brain transcriptomics are primarily sex-specific across brain regions, with very little overlap of

differentially expressed genes (DEGs) and discordance in overall patterns of difference between the sexes.

Single-nucleus RNA-sequencing (snRNA-seq) can disentangle cell type specific transcriptomic contributions to complex neuropsychiatric conditions[11-15], and our recent snRNA-seq results[16] revealed disruptions in deep layer excitatory neurons and immature oligodendrocyte precursor cells (OPCs) in the prefrontal cortex (PFC) of males with MDD. Given the higher prevalence of MDD among women, the known sex-specific differences in MDD, and growing evidence that male and female MDD may be mediated by distinct brain molecular mechanisms, we conducted a study in a cohort of female individuals and applied an updated unified analysis pipeline to both the female and previously generated male cohorts. With a total of 71 individuals, 37 cases and 34 controls and over 160,000 single-nuclei profiled, our dataset represents the largest snRNA-seq study of the human brain in MDD to date. We found that the DEGs detected and the cell types with prominent differences were distinct in males and females. However, the overall patterns of MDD-associated gene expression difference within each cell type were consistent between the sexes. Whereas in males our analysis indicated a strong involvement of deep layer excitatory neurons, astrocytes, and OPCs – consistent with our previous report, in females we found a striking contribution of microglia and parvalbumin (PV) interneurons to MDD pathology.

Results

Profiling cells of the human dorsolateral prefrontal cortex (dlPFC)

snRNA-seq data was generated from the dlPFC for 20 female subjects with MDD and 18 neurotypical female controls (Figure 1a, schematic; Table 1, demographic and sample

characteristics; Supplementary Data 1, sequencing metrics) and combined with previously generated data from males[16]. After pre-processing with a unified pipeline (methods: Sequencing, alignment, and generation of count matrices), we retained 160,711 high-quality nuclei with comparable contributions of sex (51% from females) and disease status (58% MDD). We used Harmony[17] to correct for covariates, including batch effects (Supplementary Figure 1a-d), and applied the scclusteval[18] workflow to optimize the Seurat clustering parameters (Supplementary Figure 2a-b) resulting in the identification of 41 nuclei clusters. Clusters mostly did not appear to be driven by batch, sex, brain bank, or subject (Supplementary Figure 3a-e, g).

Of the 41 clusters, 40 could each be confidently annotated to one of 7 major brain cell types (methods: Cluster annotation, Figure 1b-d, Supplementary Figure 4, Supplementary Data 2) – excitatory neurons (48% of nuclei), inhibitory neurons (18% of nuclei), oligodendrocytes (14% of nuclei), astrocytes (8% of nuclei), OPCs (5% of nuclei), endothelial cells (2.5% of nuclei), and microglia (2% of nuclei). The one unassigned cluster displayed a mixed expression profile of neuronal and glial marker genes (2% of nuclei).

We annotated 30 neuronal clusters, both excitatory (20 clusters) and inhibitory (10 clusters), using known subtype markers (Supplementary Figure 5a-b). Excitatory neuronal clusters were annotated according to their layer of origin and inhibitory neuronal clusters according to their developmental origin, where applicable. For non-neuronal cells, we identified one microglial cluster, two clusters of astrocytes, and three clusters each of oligodendrocytes and OPCs. Clusters annotated to the oligodendrocyte lineage (OL) were further characterized using pseudotime trajectory analysis (methods: Pseudotime trajectory analysis; Supplementary Figure 5c-d).

Using the gene expression patterns of our clusters and matching them to published clusters in several human brain datasets[19,20], we found close correspondence between observed cell types (Figure 1e, Supplementary Figure 6), further emphasizing that the quality of our data, clustering, and annotation are at par with other recent snRNA-seq datasets for the human brain. Besides the single cluster with mixed expression profile, two other clusters showed evidence of possible technical effects, ExN17 and ExN5 (methods: Assessment of clustering quality).

Cell types with altered proportions in MDD

We next examined whether proportions of nuclei in broad cell types and clusters differed between cases and controls. We observed that the proportions of nuclei per subject contributing to the broad astrocytic and OPC cell types were significantly decreased in cases compared to controls (two-sided Wilcoxon-test, FDR =3.46×10^{-4}, Ast; FDR= 5.32×10^{-4}, OPC; Figure 1f, Supplementary Data 3) and there were concomitant increases in excitatory neurons (FDR 0.0477). Similarly, there were reduced proportions of nuclei in both astrocytic clusters (Ast1, FDR 0.00188; Ast2, FDR 0.00291) and in two of three OPC clusters (OPC1, FDR 0.009799; OPC2, FDR 0.0168; Figure 1f, Supplementary Data 3). The robustness of these differences was supported by sub-sampling analysis (methods: Cell type proportions comparison). Splitting the male and female datasets revealed similar patterns as observed for the combined data (Supplementary Figure 7). These results are similar to those found in analyses of other brain disorders[11,21], and indicate that there may be decreased proportion of astrocytes and OPCs may be reduced in MDD. Here the FDR refers to Benjamini and Hochberg correction.

Global cell type specific transcriptomic changes are largely concordant between the sexes

We next asked whether there are sex-specific differences in the gene expression patterns of individual cell types. To answer this question, we performed differential gene expression analysis comparing cases and controls in broad cell types and clusters, in males and females separately. In both males and females, we observed a high proportion of common DEGs between broad and cluster level analyses. However, consistent with previous studies showing distinct brain transcriptomic changes in males and females with MDD[10,22], few DEGs were common to both sexes (Figure 2a). To compare overall patterns of depression-associated gene expression in males and females beyond those genes passing significance thresholds, we performed rank-rank hypergeometric overlap (RRHO) analysis[23] (methods: Comparison of male and female results). Specifically, we used RRHO2 to compare the orderings of the genes induced by MDD association statistics in males compared to females. These orderings were generally moderately to strongly concordant between the sexes (Figure 2b). Some evidence of discordance was visible only for Oli and OPC. There was a significant overlap between males and females in genes less expressed in MDD in Ast, ExN, and InN (warm colors in top right quadrant of RRHO plots) and an overlap in genes more expressed in MDD in Mic (warm colors in bottom left quadrant of RRHO plot).

At the cluster level there was some evidence of discordance between the sexes, with 8 out of 34 clusters compared showing discordant patterns. This encompassed certain neuronal clusters, primarily excitatory neuronal, including ExN4_L35, ExN7, ExN12_L56, ExN13_L56, InN10_ADARB2 (Supplementary Figure 8). Within the oligodendrocyte lineage, discordance is apparent for the Oli2, Oli3, and OPC1 clusters (Figure 2c). Supplementary Data 4 summarizes the maximum -log10 p-values from RRHO2 analyses and the classification of the results into weak, moderate, strong categories or concordance and discordance – with strongly concordant or

discordant results providing the most convincing evidence for similarity or difference between the sexes. As can be seen from Supplementary Data 4, we can see moderate or strong evidence for concordance between males and females in not only neuronal cell types, but also glia (e.g.- microglia and astrocytes).

Taken together we find that, although cell type specific statistically significant MDD-associated DEGs differ between the sexes, a threshold-free ranking approach to comparison shows considerable concordance between males and females for the majority of broad cell types and clusters.

We further assessed whether the similarities in cell type specific MDD-associated gene expression differences between males and females was likely to arise by chance using permutation analysis, which supported our conclusion that the similarities are not driven by chance (methods: Permutation analysis, Supplementary Figure 9e-j, Supplementary Data 4). For broad cell types, excluding the cluster annotated as having a mixed contribution of cell types, on average 91% of the time the real data yielded a higher correlation between male and female results than the permuted data. For clusters with concordant patterns, on average 90% of the time the real data yielded a higher correlation between males and females than the permuted data. However, for clusters with evidence of discordance, the real correlation was higher than permuted correlation only 42% of the time.

Cell types with strongest MDD associations differ by sex

Next, we identified the cell types with the strongest evidence of dysregulation due to MDD in each sex. In males, our reanalysis indicated results consistent with those we reported previously,

i.e., for broad cell types we identified the highest number of DEGs in astrocytes (90/151, 60%) and OPCs (54/151, 36%) (Figure 3a, Supplementary Figure 10a-d, Supplementary Data 5), whereas at the cluster level (Figure 3b, Supplementary Figure 10e-h & 11, Supplementary Data 5), the highest number of DEGs were found in a cluster of deep layer excitatory neurons – ExN10_L46 (238/447, 53%) and a cluster of astrocytes – Ast1 (98/447, 22%). A summary of the proportions of upregulated versus downregulated genes and unique DEGs versus DEGs shared across clusters is provided in Figure 3e. Correlations between gene expression fold differences calculated in our reanalysis and our previous analysis are provided in Supplementary Figure 10i-l (methods: Differential expression analysis - Comparison of male differential expression results to previous results).

In females, for broad cell types, we detected a high number of DEGs in microglia only (74/85, 87%) (Figure 3c, Supplementary Figure 12a-d, Supplementary Data 6). The same analysis at the cluster level (Figure 3d, Supplementary Figure 12e-h, Supplementary Data 6) consistently showed the highest number of DEGs in the Mic1 (Figure 3f; 68/180 DEGs, 38%) cluster with a large proportion (53/68, 78%) overlapping with the microglial DEGs at the broad level. We focused on cluster level results for follow up analyses (methods: Differential expression analysis, for justification) and assessed the robustness of our microglial findings against misclassified or contaminating cells (Supplementary Figure 12i).

The majority of microglial DEGs (47/68, 69%) were confirmed to be both transcribed and translated in microglia using a TRAP gene expression dataset in a lipopolysaccharide challenge mouse model[24].

In addition to microglia, several inhibitory neuronal clusters (Figure 3g), including two *PVALB* expressing clusters – InN1_PV and InN9_PV as well as an *SST* expressing cluster – InN2_SST and an *ADARB2* expressing cluster – InN8_ADARB2 contained the majority of remaining DEGs. Our results thus pointed to dysregulation of microglia and inhibitory neurons, especially PV interneurons in females with MDD which further prompted us to explore the biological pathways within and possible interactions between these cell types which could be altered in MDD, as detailed below.

Further, our permutation analyses revealed that at the broad level the number of unique DEGs identified with the real data was higher than 93% of permutations for females and 97% of permutations for males (Supplementary Figure 9a-b). At the cluster level, for males, the real number of unique DEGs was higher than the number of permuted DEGs 94% of the time (Supplementary Figure 9c). The evidence from permutations was weaker at the cluster level for females with 60% of permutations revealing fewer unique DEGs than the real data (Supplementary Figure 9d).

Meta-analysis reveals additive effects of depression-associated transcriptomic changes in males and females

To maximize statistical power to observe gene expression differences common to both males and females, we performed meta-analyses of the male and female data within each broad cell type and cluster. For broad cell types, the meta-analysis revealed upregulated genes in microglia and downregulated genes in astrocytes, with the majority of DEGs from the separate male and female analyses retained (Figure 4, Supplementary Data 7). There were more DEGs in microglia (172

DEGs) than observed in the female dataset alone (74 DEGs), whereas there were fewer DEGs in astrocytes (53 DEGs) than identified in males alone (90 DEGs). 49/90 (54%) DEGs in the broad astrocytic cluster in males and 56/74 (76%) DEGs in the broad microglial cluster in females were recapitulated in the meta-analysis. There were 22 DEGs in OPCs in the meta-analysis, but the number was less than half compared to the independent analysis of the male dataset (54 DEGs) whereas for oligodendrocytes the number of DEGs was higher when the data were meta-analyzed (21 versus 7 DEGs in the male dataset alone). The decrease in number of MDD-associated DEGs in OPCs when combining the male and female cohorts indicates that gene expression differences in OPCs in MDD are dissimilar between the sexes. This agrees with the discordance of depression-related transcriptomic changes between sexes in OPCs in our RRHO2 analysis.

At the cluster level, we found that upregulated DEGs in Mic1 and downregulated DEGs in ExN10_L46 stood out as the top findings in the meta-analysis (Figure 4, Supplementary Data 7). Once again, we found more microglial DEGs (128 DEGs) via the meta-analysis compared to the female data alone (68 DEGs) and more DEGs in ExN10_L46 (254 DEGs) than with the male data alone (238 DEGs).

Given the overall between-sex concordance in MDD-associated gene expression changes detected in RRHO2, it is not surprising that clusters with prominent differential expression from the individual cohorts also stood out in the meta-analysis. Taken together these results further support that the global patterns of change in gene expression within cell type are generally

consistent between males and females, especially for excitatory neurons and microglia, with a few notable exceptions such as OPCs.

Female cell type specific DEGs are enriched for previous MDD-linked genes

The relevance of the DEGs we have identified to psychiatric disorders was evaluated by referring to the PsyGeNET[25] text-mining database. Compared to other disorders, depressive disorders had the most gene-disease associations with the female cell type specific DEGs (> 60; Figure 5a). The next largest number of gene-disease associations was for schizophrenia (< 40). Statistically, the overlap of all DEGs at the cluster level with disease-associated genes in PsyGeNET was significant only for two disease categories, Depressive disorders (hypergeometric test, $p = 0.0378$) and Alcohol use disorders (hypergeometric test, $p = 0.0141$). Further, for the top 5 clusters with highest numbers of DEGs in the female cluster-level analysis, gene-disease associations for depression and related disorders in PsyGeNET were identified for several DEGs (Figure 5b-c). Therefore, our cell type specific DEG findings in females recapitulated previously reported gene-disease associations.

Disease-relevant biological pathways revealed by cell type specific transcriptomic changes in females with MDD

To explore the underlying pathways associated with the cell type specific transcriptomic changes in females with MDD, we performed pre-ranked gene set enrichment analysis (GSEA; methods: Pre-ranked gene set enrichment analysis). Female microglia from cases showed significant negative enrichment scores for inflammation-related Reactome pathway gene sets including “Interferon Gamma signaling”, “Interleukin 4 and Interleukin 13 signaling”, “Interleukin 10 signaling”, and “TNFR2 non-canonical NF-KB pathway” (Figure 5d, Supplementary Data 8).

"Neuronal system" gene sets were positively enriched with contributions from "Voltage-gated potassium channels", "Class C/3 metabotropic glutamate/pheromone receptors", and "Neurexins and neuroligins" among others (Figure 5d). Interestingly both pro- and anti-inflammatory immune signaling pathway gene sets were downregulated which may indicate that MDD-associated dysregulation of gene expression in microglia involves more than just a microglial inflammatory response.

Further, both PV interneuron clusters showed a negative enrichment of heat shock factor 1 (HSF1) related terms – "HSF1 activation" in InN9_PV and "HSF1 dependent transactivation" in InN1_PV. Moreover, both clusters showed an enrichment of the gene sets "Cellular response to external stimuli" and "Metabolism of RNA". The InN1_PV cluster showed further enrichment of immune gene sets such as "Innate immune system", "Adaptive immune system", and "Cytokine signaling in immune system" and interestingly in the context of sex differences in depression, "ESR mediated signaling", pertaining to the estrogen receptor.

Thus, our GSEA of the female microglia and PV interneuron differential expression results revealed dysregulated Reactome pathway gene sets which are functionally relevant in these cell types and plausibly associated with sex differences.

Assessing the relationship between microglia and PV interneuron dysregulation in females with MDD using protein-protein interaction assessment

To further assess the functional relevance of striking gene expression differences in microglia and PV interneurons in females with MDD, we examined whether the protein products of DEGs in these clusters belonged to interacting networks. STRING[26] protein-protein interaction (PPI)

analysis (methods: STRING analysis) revealed links between the protein products of several DEGs in the microglia and the PV interneurons. We focused on the top two interactions, based on the STRING confidence score. These interactions were between protein products of DEGs coming from microglia and PV interneurons and with the same direction of change (Figure 5e). The *ROBO2* gene, which encodes a canonical cell migration guidance receptor[27], was increased in microglia whereas one of its corresponding ligands, *SLIT3*[27] was increased in expression in the InN9_PV cluster. Additionally, *ADAMSTL1* and *THSD4* (also known as *ADAMTSL6*), two members of the ADAMTS-like family of proteins, which have extracellular matrix (ECM) binding properties[28], were upregulated in microglia and in the InN1_PV cluster respectively. The PPI network analysis results point to the intriguing possibility that changes in communication between microglia and PV interneurons through the ECM and cell surface molecules contribute to depression-associated brain pathology in females.

Assessing the relationship between microglia and PV interneuron dysregulation in females with MDD using ligand-receptor interaction assessment

Building upon the indications from PPI assessment we explored the possible changes in ligand-receptor expression in microglia and PV interneurons between female cases and controls with CellChat[29] (methods: CellChat analysis). CellChat identified more interacting ligand-receptor pairs and estimated increased communication strength overall within and between and within microglia and PV interneurons in cases compared to controls (Figure 5f). CellChat further identified several signaling pathways (groups of related ligand-receptor pairs) with decreased (top pathway: GAS) and increased (top pathway: SPP1) communication in cases compared to

controls (Figure 5f). Within these top signaling pathways, we specifically identified a probable increase in SPP1 to integrin communication and decrease in GAS6-MERTK communication from microglia to PV interneurons and vice versa, respectively (Supplementary Figure 13).

WGCNA confirms MDD dysregulated pathways in female microglia and PV interneurons

Next, we performed weighted-gene co-expression network analysis (WGCNA) using the pseudobulk gene expression profiles to identify correlated modules of genes associated with MDD in microglia and PV interneurons in females.

In microglia, 8 modules out of 44 had a significant correlation with case-control status (p-value < 0.05; Figure 6a). Further, the MEturquoise module which is positively correlated with MDD-status (correlation 0.627, $p = 7.26 \times 10^{-5}$) showed a significant overlap ($p = 5 \times 10^{-56}$; methods: Weighted gene co-expression network analysis) with upregulated DEGs in microglia in female cases (Figure 6b). MEturquoise also showed an enrichment of Reactome pathway gene sets related to ion channels, neurotransmitter receptors, and the neuronal system (Figure 6c) similar to gene sets found upregulated in microglia in female cases by GSEA.

In PV interneurons (including nuclei in the InN1_PV and InN9_PV clusters), 16 of 55 modules were significantly associated with case-control status ($p < 0.05$, Figure 6d). Additionally, downregulated DEGs from InN1_PV and InN9_PV significantly overlapped ($p = 3.24 \times 10^{-7}$) with the genes from the MEturquoise module (Figure 6e). The MEturquoise module which is negatively associated with MDD (correlation -0.582, p = 0.00016), had over-representation of 489 Reactome pathway gene sets. Of these, 30 pathways overlapped with the main downregulated pathways previously identified with GSEA in InN1_PV or InN9_PV (Figure 6f). The overlapping

pathways included “HSF1 activation”, “HSF1 dependent transactivation”, and “ESR mediated signaling”. Further, upregulated DEGs in InN1_PV and InN9_PV significantly overlapped with the genes of two modules which had a positive association with MDD-status: MEred (correlation 0.568, p = 0.0002) and MEgreenyellow (correlation 0.426, p = 0.0085).

Overall, the female microglia and PV interneuron WGCNA results further support our MDD-associated DEG and Reactome Pathway findings in these clusters.

Discussion

Cell type specificity of depression associated transcriptomic changes

There is a sizable body of postmortem literature describing differences from cellular morphology to proteomic and transcriptomic profiles in individuals with depression. Classic cytological experiments from the turn of the century identified abnormalities in morphology and distribution of cell types, but also put into question cell number, size, and neuropil density, particularly for neurons and astrocytes[30-34]. Transcriptomic studies have to some extent implicated all broad cell types[35-37]. Results from our present and previous study confirm this implication of multiple cell types including excitatory and inhibitory neurons, astrocytes, OPCs, and microglia.

Our study highlights potential cell type specific transcriptomic targets for treatment and intervention in MDD. Given that different cell types appear to be implicated in MDD in males and females, approaches to treatment may need to be different as well. At the very least, our findings strengthen the evidence in support of including female subjects in pre-clinical and clinical research, which had been historically neglected and continues to be neglected in biomedical research.

Sex-specificity of depression-associated transcriptomic differences

Only recently have postmortem transcriptomic studies of MDD begun to incorporate sex as a biological factor [7,10]. These studies, which analyzed bulk tissue samples, reported distinct gene expression differences in males and females with very limited overlap of DEGs. Our findings are consistent with these studies in that the cell types with most prominent differences in gene expression – and the DEGs within these cell types – were quite separate for males and females (Figure 2a). However, in contrast, we found that within each cluster and broad cell type the threshold-free patterns of MDD-associated difference in gene expression were highly concordant between the sexes in most cases, except for most oligodendrocyte lineage clusters (Figure 2b-c). This overall agreement between the sexes was confirmed by a meta-analysis of the male and female data (Figure 4a-c). Possibly, by using single-nucleus methodology, our results provided better resolution for threshold free analyses.

Notably, recent reviews on the sex specificity of transcriptomic differences in MDD[22,38] suggest that in females with MDD there is reduced microglial activation and increased synaptic connectivity while the opposite is true for males. This theory is supported by the downregulation we observed in MDD females in microglial inflammatory pathways such as interferon and NF-KB signaling.

Immune response is innately different across sexes leading to inflammatory responses that vary with age and sex resulting in a bias in susceptibility to the development of diseases from autoimmune to infections to cancer[39]. Microglia, the resident immune cell of the brain, showed the most significant difference in gene expression compared to control subjects specifically in females and not males (Figure 3a-d). It has been hypothesized that these differences result from

distinct starting points between sexes[6]. Notably, many microglial immune functions are mediated by gonadal hormones including transcriptional regulators such as suppressor of cytokine signaling 3, *SOCS3*[40], which is downregulated in our data and influences the expression of other cytokines.

Microglial contributions to MDD in females

The salient implication of microglia specifically in females is consistent with differences in the distribution, structure, transcriptome, and proteome of microglia between the sexes, in both health and disease[41-43]. Furthermore, the number and phenotype of microglia differ by sex in the rodent brain[44-46], and several recent rodent studies demonstrated sex-specificity of microglial response to stress in various brain regions[47-49]. These studies describe changes in genes involved in cellular stress and immune function with brain-region and sex-specific variation. This is roughly analogous to the female-specific pathway dysregulation we observed in microglia and PV interneurons in MDD (Figure 5d, Supplementary Data 8).

Most studies examining peripheral markers report increased inflammation in MDD[50], but studies in brain tissue have reported increases[50], decreases[51], or changes in both directions[52] in the expression of pro-inflammatory molecules. Moreover, several depression-linked genetic variants in pro-inflammatory genes, including in *IL1B*, *TNFA*, and *CRP*, are associated with decreased expression[53]. Recently the concept of a pro-inflammatory versus anti-inflammatory state of microglia has been challenged[54]. In the brain, amidst close interactions with multiple cell types, microglia adopt more diverse states with varying levels of pro- and anti-inflammatory markers, and this is being underscored by single-cell data[54]. Our results reflect altered microglial

transcription in MDD females versus controls, with pro-inflammatory (interferon and NF-KB signaling) and anti-inflammatory (IL4, IL13, and IL10 signaling) pathways simultaneously downregulated (Supplementary Data 8).

We observed evidence that further "neuronal" pathways – including neurotransmitter signaling and ion channels – were upregulated in female MDD microglia, in both differential expression and WGCNA results (Supplementary Data 8, Figure 6c). Microglia have long been known to express neurotransmitter receptors and ion channels. Mounting evidence suggests these canonically "neuronal" gene products regulate microglial activity[55-58], and our results suggest that changes in their expression may contribute to MDD pathophysiology, at least in females.

PV interneuron and microglia crosstalk in females with MDD

Together with striking changes in microglial gene expression, we observed dysregulation in PV interneurons. PV interneurons, among other interneuron subtypes, are implicated in stress and depression with evidence for sex-specific changes[59,60]. Most PV interneurons are encapsulated by ECM structures called perineuronal nets (PNNs) which help protect them from cellular stress, and microglia are known to regulate PNNs[61]. Oxidative and cellular stress relate to PV neuron and PNN deficits in animal models[62] and cellular stress may be part of the molecular pathology in MDD[63].

We found evidence of dysregulated cellular stress pathways, such as heat-shock factor activation, in PV interneurons in MDD females via differential expression analysis and WGCNA (Supplementary Data 8, Figure 6f). Moreover, both analyses pointed to dysregulation of estrogen receptor mediated signaling. The expression of many genes is regulated by the ligand-bound

estrogen receptor and difference in estrogen levels are known to contribute to differences in brain physiology between the sexes[41].

Beyond effects in individual cell types, our results imply potentially impaired communication between PV interneurons and microglia in females with MDD (Figure 5e-f). Microglial synaptic regulation involves migration of microglia towards specific neurons and in glioblastomas this migration can be regulated by SLIT-ROBO signaling[27]. The *SLIT3* gene and its corresponding receptor gene *ROBO2* were upregulated in PV interneurons and microglia respectively in females with MDD. Of note, genetic variation in *SLIT3* has been associated with depression[64].

We observed that the ECM-binding protein genes *ADAMTSL1* and *THSD4* were upregulated in microglia and PV interneurons, respectively, in MDD females. These recently characterized ADAMTS-like proteins lack the enzymatic domains through which ADAMTSs break PNN components, but they have been proposed to protect these components from degradation by mimicking ADAMTS binding[28,65]. We therefore conjecture that microglial migration cued by PV interneurons, followed by concerted alterations of the ECM by these two cell types stabilize PNNs in females with MDD. A recent study – including males and females – reported increased PNN number in the PFC of MDD subjects who experienced early life adversity[66], and our molecular findings might underlie one sex-specific mechanism for PNN alterations in MDD.

Our preliminary assessment also points to downregulation of PV interneuron to microglia signaling via GAS6-MERTK and upregulation of SPP1 to integrin signaling in the opposite direction in females with MDD. Together, MERTK and GAS6 promote homeostasis and neuronal survival and they are disrupted in several nervous system disorders[67]. On the other hand, microglial osteopontin (SPP1), promotes remyelination in multiple sclerosis and is neuroprotective near

infarcts in stroke but in Alzheimer's disease it is part of the "disease associated microglia" signature[68]. The role of these signaling molecules in depression, if any, are yet to be determined.

Limitations

This study has limitations that should be considered. We could not directly compare male and female cell type specific transcriptomes or assess the interaction of sex and disease status given that we are using data from two sex-specific datasets. Thus, the implication of different cell types in MDD between males and females could be partly attributable to differences in methodology (such as library preparation chemistry, tissue collection approach, or nuclei isolation protocol, among other factors) for generating the two datasets. However, we attempted to mitigate this by applying a unified pre-processing pipeline and joint definition of cell types. Our findings are consistent with previous evidence for sex-specific mechanisms for depression etiology in animal models and human studies[6,9,69,70].

Our permutation analysis indicated our DEGs at the cluster level for females may not be as robust as for the male cluster level analysis and the broad analysis for both sexes. However, our main findings in females at the cluster level are in microglia, and 78% of microglial DEGs in the female cluster analysis are also present in the female broad analysis, results that were robust, according to the permutation approach. Further our DEGs from the female cluster level analysis were supported by our WGCNA results, partially mitigating the concern that the DEGs can be an artefact of the differential expression analysis strategy.

Although our study included data from over 160,000 nuclei, the number of subjects was small relative to the large number of genes tested for associations with MDD. The relatively small number of subjects included in this study limits our statistical power to detect cell type specific

disease relevant genes and pathways. Further, our results may not be generalizable to all populations and this work will need to be extended with larger sample sizes from diverse populations. However, the number of subjects included in our study compares favourably to most published snRNA-seq studies of neuropsychiatric conditions to date, which have included anywhere between 11 to 48 subjects[11,12,15,71,72].

We did not identify a separate sub-population of disease-associated microglia as observed in some neurological disorders[73]. This may partly be due to the lack of cytoplasmic transcripts in snRNA-seq limiting the information about microglial states[74]. Nevertheless, a recent study highlighted similarities between cellular and nuclear microglia RNA-seq data from mouse and human – fresh and frozen – CNS samples[75]. Nuclear microglia transcriptomes are a reliable proxy for cellular transcriptomes and are less affected by cell isolation-based transcriptional artifacts[75]. We were able to detect inflammatory pathway dysregulation in female microglia despite the limitations.

Our CellChat and STRING results are speculative. We cannot draw conclusions about the proximity of microglia to PV interneurons or the presence of PNNs, as snRNA-seq involves dissociation of the tissue with loss of spatial and structural information. Neither can we conclude that protein expression is changed for our DEGs. Future studies using spatial transcriptomic techniques coupled with immunohistochemistry may better answer these questions.

Lastly, a few clusters may be of lower quality (biased by batch or according to quality parameters, and inconsistency with other datasets or with cluster enriched genes; ExN17, ExN5 and Mix). However, given that these clusters did not contribute substantially to our differential expression results, their impact on our main conclusions is likely to be limited.

Outcomes

We provide a cell type and sex-specific assessment of transcriptomic changes in the dlPFC in MDD using snRNA-seq. Our dataset represents a rich resource which will stimulate further fruitful investigations of sex- and cell type specific molecular pathways in depression. While most transcriptomic changes in males with MDD are observed in deep layer excitatory neurons, astrocytes, and OPCs, in females the changes are concentrated in microglia and PV interneurons. Although major dysregulated cell types and genes are distinct for each sex, within broad cell types and clusters the patterns of transcriptomic differences in MDD are primarily concordant between males and females. Finally, preliminary evidence hints that in females with MDD, impaired communication between microglia and PV interneurons may be an important feature of MDD molecular pathology.

Methods

Male snRNA-seq dataset

We used published snRNA-seq data from a cohort of male subjects with or without MDD[16]. We started with the raw FASTQ files available through GEO (GSE144136, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE144136) and reprocessed the data, dropping two runs from one subject (number 25) with low quality results based on the previous analysis. All male samples for the study had been obtained from the Douglas Bell-Canada Brain Bank.

Post-mortem brain samples in the female cohort

This study was approved by the Douglas Institute IRB. Human post-mortem dlPFC tissue was obtained from the Douglas- Bell-Canada Brain Bank (www.douglasbrainbank.ca, all female case samples and eight female control samples) and from the University of Miami Miller School of Medicine Brain Endowment Bank (https://med.miami.edu/programs/brain-endowment-bank, ten female control samples). Informed consent from next of kin was obtained for each individual included in this study. Frozen histological grade samples of gray and white matter were dissected from the dlPFC (Brodmann Area 9) by expert neuroanatomists and stored at –80 °C. Psychological autopsies were performed using proxy-based interviews complemented by medical charts, as previously described[76]. A summary of sample demographic characteristics is provided in Table 1. All cases included in this study died while affected by MDD or unspecified depressive disorder, whereas controls were neurotypical individuals who died suddenly without prolonged agonal periods and did not have evidence of axis I disorders. The post-mortem interval (PMI) represents the delay between an individual's death and collection and processing of the brain. One female case subject and three female control subjects were Hispanic, two female control subjects were African American, and race information was missing for one female case. All other female subjects were Caucasian, as were all subjects in the male cohort.

Nuclei extraction, single-nuclei capture, and library preparation for female cohort

Nuclei were extracted from coronal cryosections or tissue shavings across the cortical layers and white matter, weighing between 40-65 mg, obtained using a cryostat at -20 °C with thickness set to 100 microns. Nuclei were extracted as previously described[77]. Two versions of the iodixanol gradient were used – a weaker gradient using 17.5% and 15% (w/v) concentrations of iodixanol (batches 3F, 7F, 2F) and a stronger gradient using the 29% and 25% (w/v) concentrations of

iodixanol (batches 6F, 8F, 12F), as previously published[78], and we found the stronger gradient to perform better. Nuclei were resuspended in wash buffer and stained using Hoescht 33342 (1:2000). 10 uL of nuclei were loaded onto EVE cell counting slides (MBI) and imaged using an Olympus VS120 Slide Scanner (10X magnification) and counted using the QuPath[79] software (version 0.2.0) with the "Watershed cell detection" functionality.

We used the 10x Genomics Chromium controller for single-cell gene expression to isolate single nuclei for downstream RNA library preparation with 10x Genomics Chromium Single Cell 3' reagents. For samples processed with version 2 of the Chromium chemistry (Supplementary Data 1), we followed the protocols as outlined by the user guide (CG00052_SingleCell3_ReagentKitv2UserGuide_RevB; latest version at https://assets.ctfassets.net/an68im79xiti/RT8DYoZzhDJRBMrJCmVxl/6a0ed8015d89bf9602128a4c9f8962c8/CG00052_SingleCell3_ReagentKitv2UserGuide_RevF.pdf), whereas for sample processed with version 3 of the Chromium chemistry (Supplementary Data 1) we followed the protocols as outlined by the user guide (CG000204_ChromiumNextGEMSingleCell3_v3.1_Rev_D, https://assets.ctfassets.net/an68im79xiti/1eX2FPdpeCgnCJtw4fj9Hx/7cb84edaa9eca04b607f9193162994de/CG000204_ChromiumNextGEMSingleCell3_v3.1_Rev_D.pdf). The catalog numbers for the 10X Genomics single-cell RNA-seq kits for the v2 chemistry and v3 chemistry were 120237 and 1000121, respectively. The only modification was for loading concentration, which we increased by 30% as we assessed the capture of nuclei to be slightly less efficient than cell encapsulation. Nuclei were loaded to capture 3000 per sample, but because of a systematic error in counting the actual number of nuclei captured per sample was variable (Supplementary Data 1).

Sequencing, alignment, and generation of count matrices

The majority of samples in the female cohort (36) were sequenced using the Illumina NovaSeq 6000 but two samples were sequenced using BGI DNB-seq technology. Sequencing metrics are provided in Supplementary Data 1. All samples from the male cohort were realigned. Alignment was performed and count matrices were generated with Cell Ranger version 5.0.1 against the GRCh38 reference available on the 10X Genomics website (refdata-gex-GRCh38-2020-A, https://support.10xgenomics.com/single-cell-gene-expression/software/release-notes/build). We ran the "cellranger count" command using the "--include-introns" option and all other options set to default.

An initial 174,178 nuclei were obtained with Cell Ranger default cell filtering. The median value of mean reads per cell was 71,279, the average mapping rate to the transcriptome was 68.8%, the average fraction of reads in cells was 71 %, and the average sequencing saturation was 78.5% (Supplementary Data 1). There was higher intronic mapping rate (Kruskal-Wallis test p-value 0.0029) and a lower exonic mapping rate (Kruskal-Wallis test p-value 0.0048) for cases compared to controls, but no significant differences in any other sequencing quality control metrics (Supplementary Data 1).

The filtered gene barcode matrices were individually loaded into R[80] (versions 4.0.2 and 4.1.2) for downstream analysis and processed with Seurat[81] (4.0.3.9000 and 4.0.5). Percentage of reads from mitochondrially encoded genes were calculated before filtering, added as metadata, and used as a quality control parameter for nuclei filtering, after which the mitochondrial genes were removed for downstream analysis. The parameters for filtering were as follows:

Male cohort: nCount_RNA < 35000, nFeature_RNA > 350, percent.mt < 10

Female cohort v2 chemistry: nCount_RNA < 25000, nFeature_RNA > 250, percent.mt < 10

Female cohort v3 chemistry: nCount_RNA < 120000, nFeature_RNA > 350, percent.mt < 10

After filtering, we obtained 79,058 nuclei in the male cohort (43,347 from cases, 35,711 from controls) and 81,653 nuclei in the female cohort (49,926 from cases, 31,727 from controls). In the female cohort, after filtering, the median across samples of the median number of UMIs per cell and the median the number of genes per cell were 2758.5 and 1711.5 respectively (Supplementary Data 1). In the males, the corresponding numbers were 2530.5 and 1638.25 respectively (Supplementary Data 1).

Dimensionality reduction and data integration

We performed SCTransform on each Seurat object individually and used the SelectIntegrationFeatures function to set the variable genes for downstream analysis. We scaled each cell to 10000 counts and ran log normalization. We regressed out nCount_RNA and percent.mt from the counts to get scaled gene expression values for variable genes, which was used as input for calculating 100 PCA components. We corrected PCA components with Harmony[17] to account for batch, chemistry, and sample specific effects. This helped align the datasets as seen in the UMAP projections produced before and after correction (Supplementary Figure 1a-d). All UMAPs in figures were created using Seurat.

Clustering

We tested of a range of combinations of clustering parameters for the Seurat package (FindClusters function) using the scclusteval[18] sub-sampling (80% of all cells, 100 times) and

stability comparison workflow using Jaccard indices, with some customization. With each sub-sampling, PCA and Harmony were recalculated. The parameters tested were: k-param: 20, 30; resolution: 0.1, 0.3, 0.5, 0.7, 0.9, 1.1, 1.3, 1.5; number of Harmony corrected PCs to use: 70, 80.

We then set a threshold for the minimum stability with a chooseR-like[82] approach based on the bootstrapped medians of the median Jaccard index across all the clusters and all the parameter sets tested. We selected parameters that maximized the number of clusters while passing the threshold of cluster stability: 70 Harmony corrected PCA components, a k-nearest neighbors' parameter of 30, and a resolution of 0.7 (Supplementary Figure 2a-b). Repeating the Harmony correction with a seed set followed by clustering with the optimal parameters produced 41 clusters. Final UMAPs were produced using all 100 Harmony corrected PCA components and all calculation parameters set to default.

Cluster annotation

Genes enriched in clusters were calculated using the wilcoxauc function from presto[83] with default parameters, and filtered with the following criteria: padj < 0.05, logFC > log(1.5), pct_in-pct_out > 10. For annotation, the following known cell type marker genes were assessed in the cluster enriched genes:

Macrophage/microglia: *SPI1, MRC1, TMEM119, CX3CR1;* Endothelial: *CLDN5, VTN, VIM;* Astrocytes: *GLUL, SOX9, AQP4, GJA1, NDRG2, GFAP, ALDH1A1, ALDH1L1;* OPCs: *PDGFRA, PCDH15, OLIG2, OLIG1;* Oligodendrocytes: *PLP1, MAG, MOG, MOBP, MBP;* Neurons: *SNAP25, RBFOX3;* Excitatory neurons: *SATB2, SLC17A7, SLC17A6;* Inhibitory neurons: *GAD1, GAD2, SLC32A1,* Inhibitory neuronal subtypes: *VIP, PVALB, SST, ADARB2, LHX6, LAMP5, PAX6*

Additionally, expression of cell type specific genes from BRETIGEA[84] were assessed using the Seurat AddModuleScore function (Supplementary Figure 4).

Twenty clusters of excitatory cells were identified (Supplementary Figure 5a) including four superficial cortical layer neuronal clusters (ExN1_L24, ExN2_L23, ExN8_L24, ExN9_L23), ten deep cortical layer neuronal clusters (ExN3_L46, ExN4_L35, ExN10_L46, ExN11_L56, ExN12_L56, ExN13_L56, ExN15_L56, ExN16_L56, ExN19_L56, ExN20_L56) and six excitatory neuronal clusters without an obvious pattern of cortical layer specific marker expression (ExN5, ExN6, ExN7, ExN14, ExN17, ExN18). The layer annotations of excitatory neuronal clusters were supported by assessment of enrichment for genes known to be specific to the different layers of the cortex using spatial transcriptomics results from Maynard et al.[85] (data in supplementary table 4 of the cited publication).

We identified 10 inhibitory clusters (Supplementary Data 2, Supplementary Figure 5b), that can broadly be divided into cells likely derived from the medial ganglionic eminence (MGE; InN1_PV, InN9_PV, InN2_SST, InN5_SST) based on *LHX6, SST,* or *PVALB* enrichment, or the caudal ganglionic eminence (CGE; InN3_VIP, InN4_VIP, InN6_LAMP5, InN8_ADARB2, InN10_ADARB2) based on *ADARB2* enrichment. The InN2_SST cluster was enriched for *SST* and *GAD1* expression but had no *LHX6* enrichment. The InN8_ADARB2 (also referred to interchangeably as InN8_Mix) cluster also showed enrichment for *SST*. One inhibitory neuron cluster with enrichment for both *ADARB2* and *LHX6* (InN7_Mix), which has been previously reported[86].

Assessment of clustering quality

Contribution of batches, groups, brain banks, and subjects was relatively uniform across clusters (Supplementary Figure 3a-d, g). Endothelial, microglial, and oligodendrocyte lineage cells showed a higher percentage of contribution from the females compared to the males, possibly due a different dissection strategy used for the two cohorts such that for the female cohort more white matter tissue was included in the nuclei extractions. All but one cluster (number 34, later annotated as ExN17) had contributions from both the male and female cohorts and one cluster was primarily composed of cells from the female cohort (number 11, ExN5). ExN17 also showed exceptionally high numbers of UMIs detected per nucleus (Supplementary Figure 3f). Moreover, one cluster (number 17, which was later annotated as showing a mixed expression profile – Mix) had relatively high percentage of mitochondrial reads (Supplementary Figure 3f). These clusters are likely driven by technical effects rather than representing biologically driven cell subtypes or cell states, but they only represented < 6% of our data.

Comparison to other datasets

MetaNeighbor

We used MetaNeighbor[87] to compare the clusters in our dataset to several published datasets[16,19,20]. For the Song et al., 2020 data we used the h5_a[88], h7[89], h10[90], and h14[86] datasets which contain adult human cortical cells or nuclei and were reprocessed by the authors. We used our own dataset as a reference to train the model, for consistency of comparisons across the datasets and limited the analysis to the same variable genes we used for PCA and clustering. MetaNeighbor best hits plots are shown in Figure 1e and Supplementary Figure 6.

Spatial label transfer

We used Seurat to transfer the labels for layer annotation from a spatial transcriptomics dataset[85] to our dataset (Supplementary Figure 5a). Each tissue section of the spatial transcriptomics data was treated separately and one section each from two different subjects were assessed (data shown for one subject and section - 151673). Both the spatial and snRNA-seq data were preprocessed with SCTransform and transfer anchors were identified using the "canonical correlation analysis" option before transferring the labels.

Pseudotime trajectory analysis

We used "slingshot"[91] to build a pseudotime trajectory with our OL nuclei (Supplementary Figure 5c). We built the pseudotime trajectory with the male and female datasets combined. OL nuclei were subset and UMAP was rerun using the following parameters: dims= 1:10, min.dist=0.1, spread = 5, n.neighbors = 100, chosen to capture the global patterns in the data. Slingshot was run with the resulting UMAP as input and using the following parameters: extend = "n", start.clus = "OPC2", end.clus = "Oli3", stretch = 0.1, thresh = 0.3, once again chosen to capture the broad patterns in the data. The start and end clusters were chosen based on their position in the UMAP, and cluster labels were provided. The oligodendrocyte lineage (OL) clusters were arranged from OPC2 at one end of the pseudotime trajectory, followed by OPC1 and OPC3, a small cluster possibly corresponding to committed oligodendrocyte precursors (COPs). At the other end of the pseudotime trajectory Oli2, Oli1, and Oli3 were placed sequentially and could represent the order of oligodendrocyte clusters from myelinating to mature states.

We fit the expression of genes along pseudotime by splitting the data for males and females before using tradeSeq[92] (Supplementary Figure 5d). We ran fitGAM on the UMI counts for each gene, with age, PMI, pH, and batch as covariates, with conditions set to case and control status,

and nknots of 5, based on evaluation of a range of nknot values. The fitted expression of OL marker genes was visualized using the plotSmoothers function. The pseudotime trajectory analysis was performed following the vignette available here: https://kstreet13.github.io/bioc2020trajectories/articles/workshopTrajectories.html

Cell type proportions comparison

The percentage of nuclei in each cluster and each broad cell type for each sample was calculated and compared between cases and controls with Wilcoxon tests using rstatix[93]. To further mitigate the effect of outliers we obtained p-values for the Wilcoxon test using bootstrapping with 10000 replicates (R package boot[94]; Supplementary Data 3) which supported the initial results. Lastly, we also examined the distribution of p-values (Supplementary Data 3) from the Wilcoxon test rerun after randomly sub-sampling 70% of the nuclei 100 times similar to a previous study[11] and confirmed the pattern of changes in proportion preserved after sub-sampling (Supplementary Data 3). Wilcoxon tests were also repeated for the male and female datasets separately (Supplementary Figure 7) as described above. All boxplots in Figure 1 and Supplementary Figure 7 are made using the geom_boxplot function from ggplot2, and the detailed description of the boxplot elements can be found, in the documentation for the function which is linked here: https://ggplot2.tidyverse.org/reference/geom_boxplot.html.

Differential expression analysis

We performed pseudobulk differential gene expression analysis using muscat[95] and edgeR[96] at the broad cell type and cluster levels in males and females separately. Pseudobulk expression profiles were obtained by summing the raw UMI counts for each gene for each sample within the

broad cell type or cluster. Only one run of the male sample 24 was included in these analyses (M24_2 excluded). Additionally, subjects were only included if they had a minimum of 10 cells in the broad cell type or 5 cells minimum in the cluster. The covariates included age, pH, PMI, and batch and muscat's default gene and sample filtering were disabled. Further, internal checks within muscat excluded clusters where the number of samples with sufficient cells was not enough, given the model being used. DEGs were selected using an FDR (Benjamini & Hochberg) adjusted local (within cluster or broad cell type) p-value <0.05 and logFC > log2(1.1) and non-zero expression value in at least 3 samples. The isOutlier function (nmads 5, log "FALSE") from scater[97] was used to flag potential outliers on the CPMs from edgeR, as an additional assessment for genes that were called as differentially expressed. Flagged outliers were not removed from analysis. For one female subject with missing pH, F35, the average pH across all female subjects was substituted.

Since the only difference between the broad and cluster level microglial results lies in the exclusion criteria for subjects based on number of cells contributed the input data and outcomes were similar between these analyses, and we focused on the cluster level results in females for follow-up analyses.

All Venn diagrams to show overlap of differentially expressed genes were made with ggvenn (version 0.1.9). Heatmaps for differentially expressed genes were made with ComplexHeatmap (version 2.10.0).

Comparison of male differential expression results to previous results

For the male differential expression results, using linear regression we compared the log fold changes per gene for top clusters with highest numbers of DEGs from the current analysis with the per gene estimates for similar clusters with high numbers of DEGs in our previous analysis[16]. Considering only the top 1000 genes in common ranked by the p-values in the current analysis, we found moderate positive relationships with R-squared values in the 0.13-0.32 range (Supplementary Figure 10i-l). Considering that the analysis approaches were quite distinct at every upstream and downstream step, these results support a similar pattern of changes in gene expression in the male data as we had previously reported.

Sub-clustering of microglia for differential expression analysis in females

A subset of microglia clustered next to oligodendrocytes in the UMAP, which could reflect misclassified cells, doublets, or even immune oligodendroglia[98,99] or white matter microglia[100]. To determine the robustness of our microglial results to the presence of this subset of cells, we sub-clustered the microglial cluster. We found variable features within the microglial population, reran PCA and Harmony, and optimized clustering parameters (resolution 0.01, other parameters default) using silhouette scores. We excluded any subclusters which expressed oligodendrocyte lineage markers (*PLP1* and *ZFPM2*). Then we reran differential expression analysis on female microglia using the same parameters as initially used and found that new per gene logFCs showed a strong positive association (linear regression) with the initial results (Supplementary Figure 12i). Given that sub-setting microglia to most confident nuclei did not substantially alter the results and for downstream analysis, we proceeded with the DEGs obtained using the full microglial cluster.

Comparison of male and female results

Rank-rank hypergeometric overlap

We performed a threshold free, rank-rank hypergeometric overlap (RRHO) analysis with RRHO2[23]. Within each cluster or broad cell type, genes were scored using the product of the logFC and the negative log base 10 uncorrected p-value from differential expression analysis in the male and female datasets separately. The scored gene lists were provided to RRHO2_initialize function (method "hyper" and log10.ind "TRUE") and the results were plotted using the RRHO2_heatmap function.

Meta-analysis by Fisher combination of p-values

To meta-analyze the male and female differential expression results per broad cell type and per cluster, we used Fisher combination of p-values as implemented in the metaRNASeq[101] R package on the uncorrected p-values after filtering out genes detected in less than 3 samples. We also used an FDR (Benjamini & Hochberg) adjusted p-value threshold of 0.05 for genes to be considered significantly changed in the meta-analysis and removed any genes with opposite direction of change between the two datasets.

Permutation analysis

We permuted the cases versus control labels 100 times, within each batch, within the male and female datasets separately, and re-ran our differential expression analysis to obtain a distribution of the number of cell type specific unique DEGs (counting once any DEGs repeated across multiple clusters or cell types), at the broad and cluster level, for males and females, with randomly permuted groups (Supplementary Figure 9a-d). We also calculated the Spearman

correlation between the differential expression gene scores (log fold change multiplied by the negative log base 10 uncorrected p-value, as used in RRHO2 analysis) between male and female datasets per cell type and cluster. We plotted the distribution of correlation coefficients obtained between the male and female datasets using permuted case versus control labels for broad cell types (Supplementary Figure 9e-j). Further, we assessed for each cluster what percentage of Spearman correlation coefficients calculated using permuted results were less than the Spearman correlation coefficient observed with the real labels (Supplementary Data 4), with a higher percentage representing a correlation that is less likely to appear by chance with random case versus control labels.

Functional interpretation of female differential expression results

Pre-ranked gene set enrichment analysis (GSEA)

For the microglial (Mic1) and PV interneuron (InN1_PV, InN9_PV) differential expression results we individually performed pre-ranked Gene Set Enrichment Analysis[102] with FGSEA[103] using the same ranking metric as used for RRHO2 (product of log fold change and the negative log base 10 of the uncorrected p-value). We evaluated the Reactome pathway[104] gene sets obtained from msigdbr[105]. The following parameters were used for the fgsea function: eps = 0.0, minSize = 15, maxSize = 1000 and any pathways with Benjamini-Hochberg adjusted p-value < 0.1 were considered to be significant. Finally, we ran collapsedPathways with pval.threshold= 0.01 to get the main pathways for each cluster.

PsyGeNET analysis

With the list of DEGs from the female dataset across all clusters, we ran enrichedPD from psygenet2r[106] with database = "ALL" and other parameters set to default to find the psychiatric disorders for which our DEGs showed an enrichment. Next, we ran psygenetGene with database= "ALL" and other parameters set to default, and created a geneAttrPlot for the evidence index for all DEGs from all clusters in females to summarize the links between our DEGs and psychiatric disorders reported in PsyGeNET[25]. Additionally, we similarly ran psygenetGene, individually on the DEGs from Mic1, InN1_PV, InN2_SST, InN9_PV, InN8_ADARB2, and plotted the corresponding gene-disease association heatmaps with plot type = "GDA heatmap".

STRING analysis

We used STRING DB[26] (version 11.5) to assess the relationships between the protein products of our DEGs in female microglia and PV interneurons. The entire list of DEGs from these clusters (Mic1, InN1_PV, and InN9_PV) were provided as input and the confidence level was set to high (interaction score > 0.7). We then exported the network to Cytoscape (3.9.1), colored genes by direction of change in expression, shaped DEG nodes based on their cluster of origin, and labelled the edges with the confidence scores for the interactions.

CellChat analysis

We subset the relevant nuclei from females in Mic1, InN1_PV, and InN9_PV and performed CellChat[29] analysis. We relabelled all PV interneuron nuclei as InN_PV. For cases and controls independently, we sequentially ran identifyOverExpressedGenes and identifyOverExpressedInteractions with lenient default parameters to find the ligand-receptor gene combinations overexpressed in these cell types. Next, we ran computeCommunProb (with

nboot = 1000) followed by computeCommunProbPathway, netAnalysis_computeCentrality, and aggregateNet with default parameters to find the ligand-receptor pathways present. Lastly, we merged the case and control objects and ran computeNetSimilarityPairwise with type "functional". Finally, we used the compareInteractions, rankNet, and netVisual_bubble to visualize the results. We used the following vignette for CellChat analysis: https://github.com/sqjin/CellChat/blob/master/tutorial/Comparison_analysis_of_multiple_datasets.html

Weighted gene co-expression network analysis (WGCNA)

Weighted gene co-expression network analysis (WGCNA) was performed to identify co-expression modules using the snRNA-seq expression data[107]. First, the aggregated expression for each female sample in microglia and PV interneuron clusters (InN1_PV and InN9_PV combined) was calculated by summing the counts per gene across all nuclei. We excluded subjects that did not have at least 5 microglial nuclei or 5 PV interneuron nuclei (InN1_PV and InN9_PV combined). To account for known external sample traits, the counts were corrected for age, pH, PMI, and batch (same as covariates used for differential gene expression analysis) using limma[108]. In addition, lowly expressed genes with total counts of below 5 were removed. A soft thresholding power of 10 and 12, respectively, with a minimum module size of 30 genes, were used for network construction and module detection for microglia and PV interneurons. Each module was correlated with the phenotype (healthy control vs MDD), and significance was determined using a p-value < 0.05.

To further characterize modules correlated with MDD, Fisher tests for overlap were performed to calculate the over-representation of DEGs as described previously[109]. In addition, the functional annotation of modules was determined using Reactome Pathway gene set over-representation analysis provided by clusterprofiler[110].

Data availability

Raw sequencing data (FASTQ files) for the female cohort generated in this study is available on GEO (accession number: GSE213982, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE213982) along with the processed gene-barcode matrix and metadata including both male and female cohorts. The raw sequencing data for the male cohort are also available on GEO (accession number: GSE144136, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE144136). Additionally, the processed data from this study are available on the UCSC Cell Browser for easy visualization: https://dlpfc-mdd.cells.ucsc.edu. Source data for all figures in this paper are provided on Zenodo: https://doi.org/10.5281/zenodo.7884086. The reference genome version used is available on the 10X Genomics website (refdata-gex-GRCh38-2020-A, https://support.10xgenomics.com/single-cell-gene-expression/software/release-notes/build).

Allen Brain Institute motor cortex data used for MetaNeighbor comparison are available for download here: https://portal.brain-map.org/atlases-and-data/rnaseq/human-m1-10x. STAB reprocessed data from published snRNA-seq and scRNA-seq datasets used for MetaNeighbor comparison is available here: https://mai.fudan.edu.cn/stab/help/. The spatial transcriptomics

data used here for label transfer is available using the spatialLIBD (version 1.6.0) R package and through the AWS download links provided here: https://github.com/LieberInstitute/spatialLIBD.

Code availability

All scripts used to analyze data are provided in a Github repository (https://github.com/MgssGroup/snRNASeq_public/). The DOI for the version of the scripts used is: https://zenodo.org/badge/latestdoi/634913347.

References

1. James, S.L*., et al.* Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet* **392**, 1789-1858 (2018).
2. American Psychiatric, A. Diagnostic and statistical manual of mental disorders : DSM-5-TR. (American Psychiatric Association Publishing, Washington, DC, 2022).
3. Howard, D.M*., et al.* Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nature Neuroscience* **22**, 343-352 (2019).
4. Gandal, M.J*., et al.* Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science* **359**, 693-697 (2018).
5. Kessler, R.C. & Bromet, E.J. The epidemiology of depression across cultures. *Annu Rev Public Health* **34**, 119-138 (2013).
6. Seney, M.L. & Sibille, E. Sex differences in mood disorders: perspectives from humans and rodent models. *Biol Sex Differ* **5**, 17 (2014).
7. Seney, M.L*., et al.* Opposite Molecular Signatures of Depression in Men and Women. *Biol Psychiatry* **84**, 18-27 (2018).
8. Bachmann, S. Epidemiology of Suicide and the Psychiatric Perspective. *International Journal of Environmental Research and Public Health* **15**, 1425 (2018).
9. Labaka, A., Goñi-Balentziaga, O., Lebeña, A. & Pérez-Tejada, J. Biological Sex Differences in Depression: A Systematic Review. *Biol Res Nurs* **20**, 383-392 (2018).
10. Labonté, B*., et al.* Sex-specific transcriptional signatures in human depression. *Nat Med* **23**, 1102-1111 (2017).
11. Morabito, S*., et al.* Single-nucleus chromatin accessibility and transcriptomic characterization of Alzheimer's disease. *Nature Genetics* **53**, 1143-1155 (2021).
12. Smajić, S*., et al.* Single-cell sequencing of human midbrain reveals glial activation and a Parkinson-specific neuronal state. *Brain* **145**, 964-978 (2022).
13. Reiner, B.C*., et al.* Single-nuclei transcriptomics of schizophrenia prefrontal cortex primarily implicates neuronal subtypes. *bioRxiv*, 2020.2007.2029.227355 (2020).
14. Ruzicka, W.B*., et al.* Single-cell dissection of schizophrenia reveals neurodevelopmental-synaptic axis and transcriptional resilience. *medRxiv*, 2020.2011.2006.20225342 (2020).
15. Velmeshev, D*., et al.* Single-cell genomics identifies cell type–specific molecular changes in autism. *Science* **364**, 685 (2019).
16. Nagy, C*., et al.* Single-nucleus transcriptomics of the prefrontal cortex in major depressive disorder implicates oligodendrocyte precursor cells and excitatory neurons. *Nature Neuroscience* **23**, 771-781 (2020).
17. Korsunsky, I*., et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nature Methods* **16**, 1289-1296 (2019).
18. Tang, M*., et al.* Evaluating single-cell cluster stability using the Jaccard similarity index. *Bioinformatics* **37**, 2212-2214 (2021).

19. Song, L.*, et al.* STAB: a spatio-temporal cell atlas of the human brain. *Nucleic Acids Res* **49**, D1029-d1037 (2021).
20. Bakken, T.E.*, et al.* Comparative cellular analysis of motor cortex in human, marmoset and mouse. *Nature* **598**, 111-119 (2021).
21. Batiuk, M.Y.*, et al.* Selective vulnerability of supragranular layer neurons in schizophrenia. *bioRxiv*, 2020.2011.2017.386458 (2021).
22. Seney, M.L., Glausier, J. & Sibille, E. Large-Scale Transcriptomics Studies Provide Insight Into Sex Differences in Depression. *Biol Psychiatry* **91**, 14-24 (2022).
23. Cahill, K.M., Huo, Z., Tseng, G.C., Logan, R.W. & Seney, M.L. Improved identification of concordant and discordant gene expression signatures using an updated rank-rank hypergeometric overlap approach. *Scientific Reports* **8**, 9588 (2018).
24. Boutej, H.*, et al.* Diverging mRNA and Protein Networks in Activated Microglia Reveal SRSF3 Suppresses Translation of Highly Upregulated Innate Immune Transcripts. *Cell Reports* **21**, 3220-3233 (2017).
25. Gutierrez-Sacristan, A.*, et al.* Text mining and expert curation to develop a database on psychiatric diseases and their genes. *Database : the journal of biological databases and curation* **2017**(2017).
26. Szklarczyk, D.*, et al.* The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res* **49**, D605-d612 (2021).
27. Geraldo, L.H.*, et al.* SLIT2/ROBO signaling in tumor-associated microglia and macrophages drives glioblastoma immunosuppression and vascular dysmorphia. *The Journal of Clinical Investigation* **131**(2021).
28. Porter, S., Clark, I.M., Kevorkian, L. & Edwards, D.R. The ADAMTS metalloproteinases. *Biochem J* **386**, 15-27 (2005).
29. Jin, S.*, et al.* Inference and analysis of cell-cell communication using CellChat. *Nature Communications* **12**, 1088 (2021).
30. Stockmeier, C.A. & Rajkowska, G. Cellular abnormalities in depression: evidence from postmortem brain tissue. *Dialogues Clin Neurosci* **6**, 185-197 (2004).
31. Rajkowska, G.*, et al.* Morphometric evidence for neuronal and glial prefrontal cell pathology in major depression. *Biol Psychiatry* **45**, 1085-1098 (1999).
32. Cotter, D., Mackay, D., Landau, S., Kerwin, R. & Everall, I. Reduced glial cell density and neuronal size in the anterior cingulate cortex in major depressive disorder. *Arch Gen Psychiatry* **58**, 545-553 (2001).
33. Cotter, D.*, et al.* Reduced neuronal size and glial cell density in area 9 of the dorsolateral prefrontal cortex in subjects with major depressive disorder. *Cereb Cortex* **12**, 386-394 (2002).
34. Ongür, D., Drevets, W.C. & Price, J.L. Glial reduction in the subgenual prefrontal cortex in mood disorders. *Proc Natl Acad Sci U S A* **95**, 13290-13295 (1998).
35. Sequeira, A.*, et al.* Global brain gene expression analysis links glutamatergic and GABAergic alterations to suicide and major depression. *PLoS One* **4**, e6585 (2009).
36. Tripp, A., Kota, R.S., Lewis, D.A. & Sibille, E. Reduced somatostatin in subgenual anterior cingulate cortex in major depression. *Neurobiol Dis* **42**, 116-124 (2011).

37. 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).
38. Bollinger, J.L. Uncovering microglial pathways driving sex-specific neurobiological effects in stress and depression. *Brain, Behavior, & Immunity - Health* **16**, 100320 (2021).
39. Klein, S.L. & Flanagan, K.L. Sex differences in immune responses. *Nat Rev Immunol* **16**, 626-638 (2016).
40. Shi, M.*, et al.* Estrogen receptor-regulated SOCS3 modulation via JAK2/STAT3 pathway is involved in BPF-induced M1 polarization of macrophages. *Toxicology* **433-434**, 152404 (2020).
41. Rahimian, R., Cordeau, P. & Kriz, J. Brain Response to Injuries: When Microglia Go Sexist. *Neuroscience* **405**, 14-23 (2019).
42. Rahimian, R., Belliveau, C., Chen, R. & Mechawar, N. Microglial Inflammatory-Metabolic Pathways and Their Potential Therapeutic Implication in Major Depressive Disorder. *Front Psychiatry* **13**, 871997 (2022).
43. Guneykaya, D.*, et al.* Transcriptional and Translational Differences of Microglia from Male and Female Brains. *Cell Reports* **24**, 2773-2783.e2776 (2018).
44. Caplan, H.W., Cox, C.S. & Bedi, S.S. Do microglia play a role in sex differences in TBI? *J Neurosci Res* **95**, 509-517 (2017).
45. Schwarz, J.M. & Bilbo, S.D. Sex, glia, and development: interactions in health and disease. *Horm Behav* **62**, 243-253 (2012).
46. Lenz, K.M., Nugent, B.M., Haliyur, R. & McCarthy, M.M. Microglia are essential to masculinization of brain and behavior. *J Neurosci* **33**, 2761-2772 (2013).
47. Bollinger, J.L., Bergeon Burns, C.M. & Wellman, C.L. Differential effects of stress on microglial cell activation in male and female medial prefrontal cortex. *Brain Behav Immun* **52**, 88-97 (2016).
48. Woodburn, S.C., Bollinger, J.L. & Wohleb, E.S. Synaptic and behavioral effects of chronic stress are linked to dynamic and sex-specific changes in microglia function and astrocyte dystrophy. *Neurobiology of Stress* **14**, 100312 (2021).
49. Bollinger, J.L., Collins, K.E., Patel, R. & Wellman, C.L. Behavioral stress alters corticolimbic microglia in a sex- and brain region-specific manner. *PLoS One* **12**, e0187631 (2017).
50. Wang, Q., Roy, B., Turecki, G., Shelton, R.C. & Dwivedi, Y. Role of Complex Epigenetic Switching in Tumor Necrosis Factor-α Upregulation in the Prefrontal Cortex of Suicide Subjects. *Am J Psychiatry* **175**, 262-274 (2018).
51. Morrison, F.G.*, et al.* Reduced interleukin 1A gene expression in the dorsolateral prefrontal cortex of individuals with PTSD and depression. *Neuroscience Letters* **692**, 204-209 (2019).
52. Mahajan, G.J.*, et al.* Altered neuro-inflammatory gene expression in hippocampus in major depressive disorder. *Progress in Neuro-Psychopharmacology and Biological Psychiatry* **82**, 177-186 (2018).
53. Barnes, J., Mondelli, V. & Pariante, C.M. Genetic Contributions of Inflammation to Depression. *Neuropsychopharmacology* **42**, 81-98 (2017).

54. Ransohoff, R.M. A polarizing question: do M1 and M2 microglia exist? *Nature Neuroscience* **19**, 987-991 (2016).
55. Liu, H., Leak, R.K. & Hu, X. Neurotransmitter receptors on microglia. *Stroke Vasc Neurol* **1**, 52-58 (2016).
56. Eder, C. Ion channels in microglia (brain macrophages). *American Journal of Physiology-Cell Physiology* **275**, C327-C342 (1998).
57. Stolero, N. & Frenkel, D. The dialog between neurons and microglia in Alzheimer's disease: The neurotransmitters view. *Journal of Neurochemistry* **158**, 1412-1424 (2021).
58. Bhandage, A.K. & Barragan, A. GABAergic signaling by cells of the immune system: more the rule than the exception. *Cellular and Molecular Life Sciences* **78**, 5667-5679 (2021).
59. Wellman, C.L*., et al.* Sex Differences in Risk and Resilience: Stress Effects on the Neural Substrates of Emotion and Motivation. *J Neurosci* **38**, 9423-9432 (2018).
60. Pesarico, A.P*., et al.* Chronic Stress Modulates Interneuronal Plasticity: Effects on PSA-NCAM and Perineuronal Nets in Cortical and Extracortical Regions. *Frontiers in Cellular Neuroscience* **13**(2019).
61. Gibel-Russo, R., Benacom, D. & Di Nardo, A.A. Non-Cell-Autonomous Factors Implicated in Parvalbumin Interneuron Maturation and Critical Periods. *Frontiers in Neural Circuits* **16**(2022).
62. Steullet, P*., et al.* Oxidative stress-driven parvalbumin interneuron impairment as a common mechanism in models of schizophrenia. *Molecular Psychiatry* **22**, 936-943 (2017).
63. Kowalczyk, M*., et al.* Cellular Response to Unfolded Proteins in Depression. *Life* **11**, 1376 (2021).
64. Glessner, J.T*., et al.* Duplication of the SLIT3 Locus on 5q35.1 Predisposes to Major Depressive Disorder. *PLOS ONE* **5**, e15463 (2010).
65. Hirohata, S*., et al.* Punctin, a Novel ADAMTS-like Molecule, ADAMTSL-1, in Extracellular Matrix*. *Journal of Biological Chemistry* **277**, 12182-12189 (2002).
66. Tanti, A*., et al.* Child abuse associates with increased recruitment of perineuronal nets in the ventromedial prefrontal cortex: a possible implication of oligodendrocyte progenitor cells. *Molecular Psychiatry* **27**, 1552-1561 (2022).
67. Burstyn-Cohen, T. & Hochberg, A. TAM Signaling in the Nervous System. *Brain Plasticity* **7**, 33-46 (2021).
68. Rosmus, D.-D., Lange, C., Ludwig, F., Ajami, B. & Wieghofer, P. The Role of Osteopontin in Microglia Biology: Current Concepts and Future Perspectives. *Biomedicines* **10**, 840 (2022).
69. Brivio, E., Lopez, J.P. & Chen, A. Sex differences: Transcriptional signatures of stress exposure in male and female brains. *Genes Brain Behav* **19**, e12643 (2020).
70. Blokland, G.A.M*., et al.* Sex-Dependent Shared and Nonshared Genetic Architecture Across Mood and Psychotic Disorders. *Biol Psychiatry* **91**, 102-117 (2022).
71. Lau, S.-F., Cao, H., Fu, A.K.Y. & Ip, N.Y. Single-nucleus transcriptome analysis reveals dysregulation of angiogenic endothelial cells and neuroprotective glia in Alzheimer’s disease. *Proceedings of the National Academy of Sciences* **117**, 25800-25809 (2020).

72. Mathys, H.*, et al.* Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332-337 (2019).
73. Chen, Y. & Colonna, M. Microglia in Alzheimer's disease at single-cell level. Are there common patterns in humans and mice? *J Exp Med* **218**(2021).
74. Thrupp, N.*, et al.* Single-Nucleus RNA-Seq Is Not Suitable for Detection of Microglial Activation Genes in Humans. *Cell Rep* **32**, 108189 (2020).
75. Gerrits, E., Heng, Y., Boddeke, E. & Eggen, B.J.L. Transcriptional profiling of microglia; current state of the art and future perspectives. *Glia* **68**, 740-755 (2020).
76. 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).
77. Maitra, M.*, et al.* Extraction of nuclei from archived postmortem tissues for single-nucleus sequencing applications. *Nature Protocols* **16**, 2788-2801 (2021).
78. Kriaucionis, S. & Heintz, N. The Nuclear DNA Base 5-Hydroxymethylcytosine Is Present in Purkinje Neurons and the Brain. *Science* **324**, 929-930 (2009).
79. Bankhead, P.*, et al.* QuPath: Open source software for digital pathology image analysis. *Scientific Reports* **7**, 16878 (2017).
80. R Core Team. R: A Language and Environment for Statistical Computing. (2021).
81. Hao, Y.*, et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e3529 (2021).
82. Patterson-Cross, R.B., Levine, A.J. & Menon, V. Selecting single cell clustering parameter values using subsampling-based robustness metrics. *BMC Bioinformatics* **22**, 39 (2021).
83. Korsunsky, I., Nathan, A., Millard, N. & Raychaudhuri, S. Presto scales Wilcoxon and auROC analyses to millions of observations. *bioRxiv*, 653253 (2019).
84. McKenzie, A.T.*, et al.* Brain Cell Type Specific Gene Expression and Co-expression Network Architectures. *Scientific Reports* **8**, 8868 (2018).
85. Maynard, K.R.*, et al.* Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nature Neuroscience* **24**, 425-436 (2021).
86. Hodge, R.D.*, et al.* Conserved cell types with divergent features in human versus mouse cortex. *Nature* **573**, 61-68 (2019).
87. Fischer, S., Crow, M., Harris, B.D. & Gillis, J. Scaling up reproducible research for single-cell transcriptomics using MetaNeighbor. *Nature Protocols* **16**, 4031-4067 (2021).
88. Li, M.*, et al.* Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science* **362**(2018).
89. 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).
90. Habib, N.*, et al.* Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nature Methods* **14**, 955 (2017).
91. Street, K.*, et al.* Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477 (2018).
92. Van den Berge, K.*, et al.* Trajectory-based differential expression analysis for single-cell sequencing data. *Nature Communications* **11**, 1201 (2020).
93. Kassambara, A. rstatix: Pipe-Friendly Framework for Basic Statistical Tests. (2021).

94. Canty, A. & Ripley, B.D. boot: Bootstrap R (S-Plus) Functions. (2021).
95. Crowell, H.L*., et al.* muscat detects subpopulation-specific state transitions from multi-sample multi-condition single-cell transcriptomics data. *Nature Communications* **11**, 6077 (2020).
96. Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140 (2010).
97. McCarthy, D.J., Campbell, K.R., Lun, A.T.L. & Wills, Q.F. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* **33**, 1179-1186 (2017).
98. Jäkel, S*., et al.* Altered human oligodendrocyte heterogeneity in multiple sclerosis. *Nature* **566**, 543-547 (2019).
99. Kokkosis, A.G., Madeira, M.M., Mullahy, M.R. & Tsirka, S.E. Chronic stress disrupts the homeostasis and progeny progression of oligodendroglial lineage cells, associating immune oligodendrocytes with prefrontal cortex hypomyelination. *Molecular Psychiatry* **27**, 2833-2848 (2022).
100. Rahimian, R., Perlman, K., Canonne, C. & Mechawar, N. Targeting microglia-oligodendrocyte crosstalk in neurodegenerative and psychiatric disorders. *Drug Discov Today* (2022).
101. Rau, A., Marot, G. & Jaffrézic, F. Differential meta-analysis of RNA-seq data from multiple studies. *BMC Bioinformatics* **15**, 91 (2014).
102. Subramanian, A*., et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* **102**, 15545-15550 (2005).
103. Korotkevich, G*., et al.* Fast gene set enrichment analysis. *bioRxiv*, 060012 (2021).
104. Gillespie, M*., et al.* The reactome pathway knowledgebase 2022. *Nucleic Acids Research* **50**, D687-D692 (2021).
105. Dolgalev, I. msigdbr: MSigDB Gene Sets for Multiple Organisms in a Tidy Data Format. (2021).
106. Gutierrez-Sacristan, A., Hernandez-Ferrer, C., Gonzalez, J.R. & Furlong, L.I. psygenet2r: psygenet2r - An R package for querying PsyGeNET and to perform comorbidity studies in psychiatric disorders. (2021).
107. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
108. Ritchie, M.E*., et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* **43**, e47-e47 (2015).
109. Perez, J.M*., et al.* Hippocampal subfield transcriptome analysis in schizophrenia psychosis. *Molecular Psychiatry* **26**, 2577-2589 (2021).
110. Wu, T*., et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation* **2**, 100141 (2021).

Acknowledgments

GT holds a Canada Research Chair (Tier 1) and is supported by grants from the Canadian Institute of Health Research (CIHR) (FDN148374 and ENP161427 (ERA-NET ERA PerMed)). MS works in a unit funded by the UK Medical Research Council (MC_UU_00011/5). CN is supported by grants from NARSAD (BBRF #31204) and CIHR (PJT183904).

We acknowledge the expert help of the Douglas–Bell Canada Brain Bank staff. This Project has been made possible with the financial support of Health Canada, through the Canada Brain Research Fund, an innovative partnership between the Government of Canada (through Health Canada) and Brain Canada, and in part by funding from the Canada First Research Excellence Fund, awarded to McGill University for the Healthy Brains for Healthy Lives initiative, 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.

This research was enabled in part by support provided by Calcul Québec (https://www.calculquebec.ca/en/) and the Digital Research Alliance of Canada (alliancecan.ca). We thank Senthilkumar Kailasam and the Canadian Centre for Computational Genomics team at McGill University for their services and assistance with differential expression analysis. We thank Claudia Belliveau for her suggestions on data interpretation. We thank Ryan Denniston for help with data handling and management.

Author Contributions Statement

M.M. processed samples, performed data analysis, and wrote the manuscript. H.M., A.C., and Z.A. supported data analysis. R.R. and K.Y. interpreted data and reviewed the manuscript. J.Y., L.M.F., and M.A.D. processed samples. D.C.M provided samples. M.S. provided bioinformatic expertise and reviewed the manuscript. N.M. contributed to sample procurement and reviewed the manuscript. G.T. obtained funding. G.T. and C.N. jointly supervised the study design, implementation, and manuscript preparation. C.N. additionally performed snRNA-seq experiments.

Competing Interests Statement

The authors declare no competing interests.

Tables

Table 1: Demographic and sample characteristics of cohorts

Group	**Case (n = 37)**		**Control (n = 34)**	
Sex	**Female (n = 20)**	**Male (n = 17)**	**Female (n = 18)**	**Male (n = 16)**
Age	45.10±3.19 (0.92)	41.06±4.66 (0.67)	47.89±4.45 (0.92)	38.38±4.58 (0.67)
PMI*	41.49±3.07 (0.02)	41.69±4.76 (0.10)	30.27±4.73 (0.02)	32.02±4.81 (0.10)
pH	6.58±0.08 (0.06)	6.60±0.07 (0.30)	6.34±0.08 (0.06)	6.50±0.06 (0.30)

(*) significantly different between female cases and controls (p-value < 0.05)
Numbers in brackets are uncorrected p-values from Kruskal-Wallis test between the two conditions for the same sex. Numeric values in each cell represent the mean±SEM. pH was unavailable for one female subject, F35.

Figures and Figure Legends

Figure 1

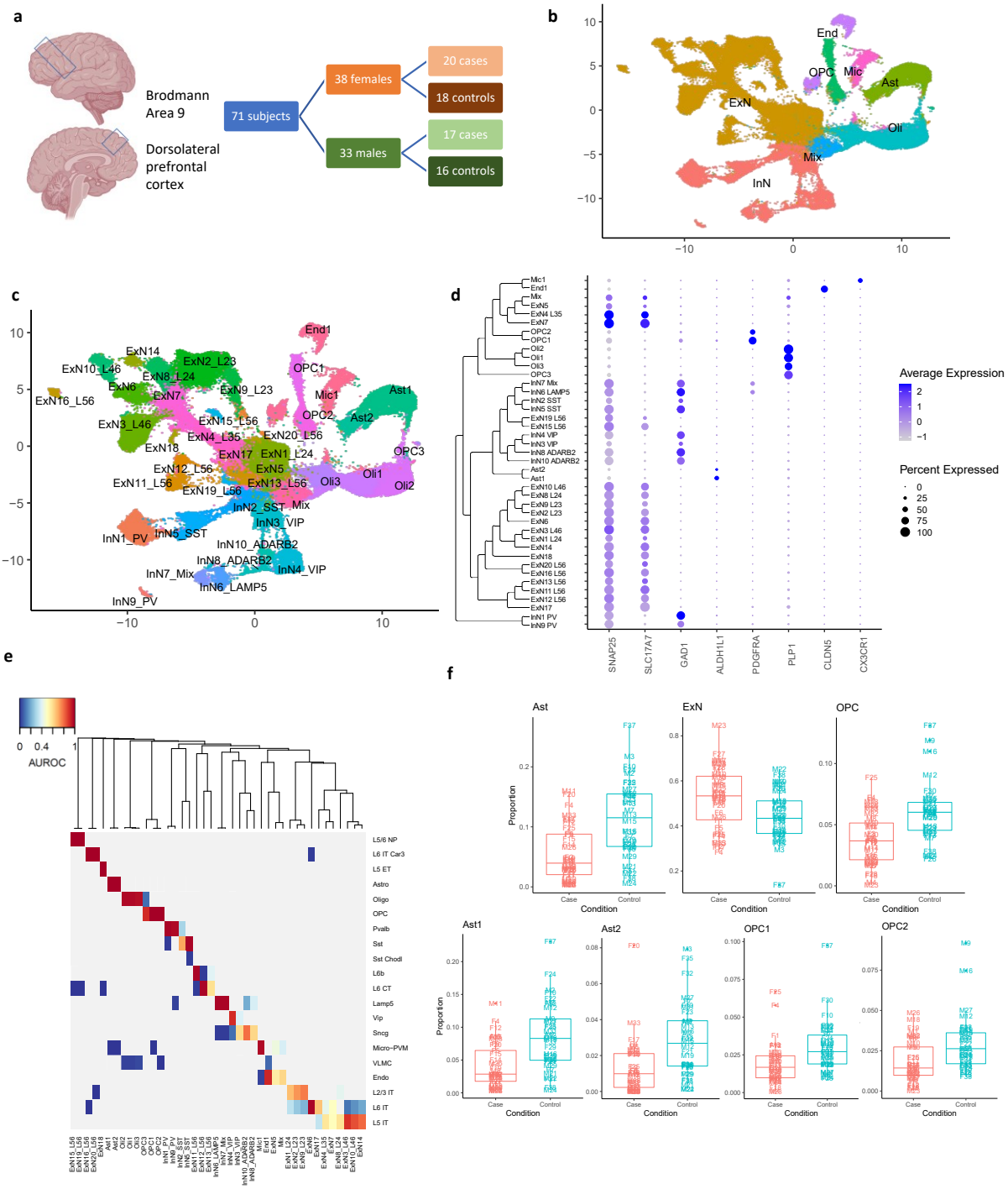


Figure 1: Overview of cell types characterized in the dlPFC. a) Schematic of study design. Diagrams depict the brain region of interest, Brodmann area 9, corresponding to the dlPFC. **b)**

UMAP plot colored by the broad cell types. **c)** UMAP plot colored by the individual clusters identified and annotated. For UMAP plots, the x and y-axes represent the first and second UMAP co-ordinates respectively. **d)** DotPlot depicting the expression of marker genes (*SNAP25* – neurons, *SLC17A7* – excitatory neurons, *GAD1* – inhibitory neurons, *ALDH1L1* – astrocytes, *PDGFRA* – oligodendrocyte precursor cells, *PLP1* – oligodendrocytes, *CLDN5* – endothelial cells, *CX3CR1* – microglia). The dendrogram next to the cluster names shows the relationship between the clusters by using the distance based on average expression of highly variable genes. **e)** Best hits heatmap from MetaNeighbor showing the correspondence between the clusters in our dataset (columns) and the broad categories of cells identified in the Allen Brain Institute human motor cortex snRNA-seq dataset[20] (rows). **f)** Boxplots showing the proportion of nuclei in each cluster for each subject split by cases and controls for the broad OPC, astrocyte, and excitatory neuron cell types and the Ast1, Ast2, OPC1, and OPC2 clusters (n= 37 cases, 34 controls, representing biologically independent samples for each cluster or broad cell type). The middle line is the median. The lower and upper hinges correspond to the 25th and 75th percentiles. Upper and lower whiskers extend from the upper or lower hinges to the largest or smallest value no further than 1.5 times the inter-quartile range from the hinge, where the inter-quartile range is the distance between the first and third quartiles. Points beyond the end of the whiskers are plotted individually. In Figures 1c-e, excitatory neuronal cluster names contain approximate layer annotations and inhibitory neuronal cluster names contain MGE or CGE specific marker information as a suffix where applicable, as described in methods: Cluster annotation. For example, ExN10_L46 denotes a cluster of excitatory neurons with enrichment of marker genes from layer 4 to layer 6 of the cortex and InN1_PV denotes a cluster on inhibitory neurons with

enrichment of the MGE specific marker PV. This convention is used throughout the paper. Brain diagram in 1a was created with BioRender.com. Source data are provided as a Source Data file.

Figure 2

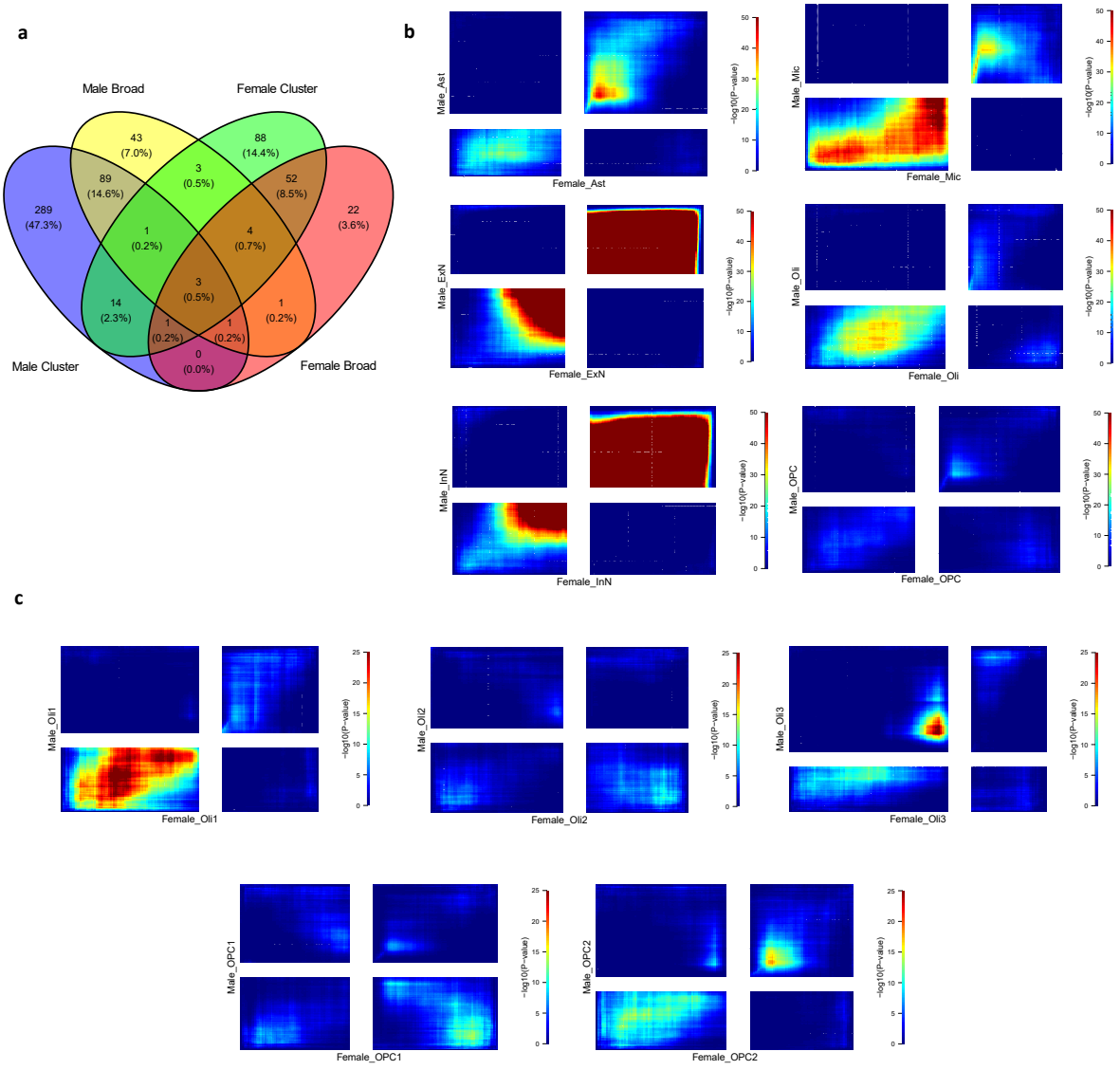


Figure 2: Overall comparison of cell type specific MDD-associated gene expression changes in males and females. a) Venn diagram showing the overlap of DEGs between the male and female datasets at the broad cell type and cluster levels. **b)** RRHO2 plots for correspondence between

differential expression results for broad cell types in the female (x-axis) and male (y-axis) datasets. Warm colors in the bottom left and top right quadrants reflect overlap in genes with increased expression or decreased expression respectively, in cases versus controls between the male and female datasets. Warm colors in the top left and bottom right quadrants reflect overlaps in genes with the opposite direction of effects between the male and female datasets. For each dataset, genes were ranked according to the value of the log of the fold change multiplied by the negative base 10 logarithm of the uncorrected p-value from differential expression analysis. **c)** RRHO2 plots similar to (b) but for oligodendrocyte lineage clusters. For RRHO2 plots comparing broad cell types the color scale maximum was set to a -log10(p-value) of 50, and for RRHO2 plots comparing clusters the color scale maximum was set to a -log10(p-value) of 25 for ease of comparison. RRHO2 uses one-sided hypergeometric tests, the p-values plotted here are uncorrected. Source data are provided as a Source Data file.

Figure 3

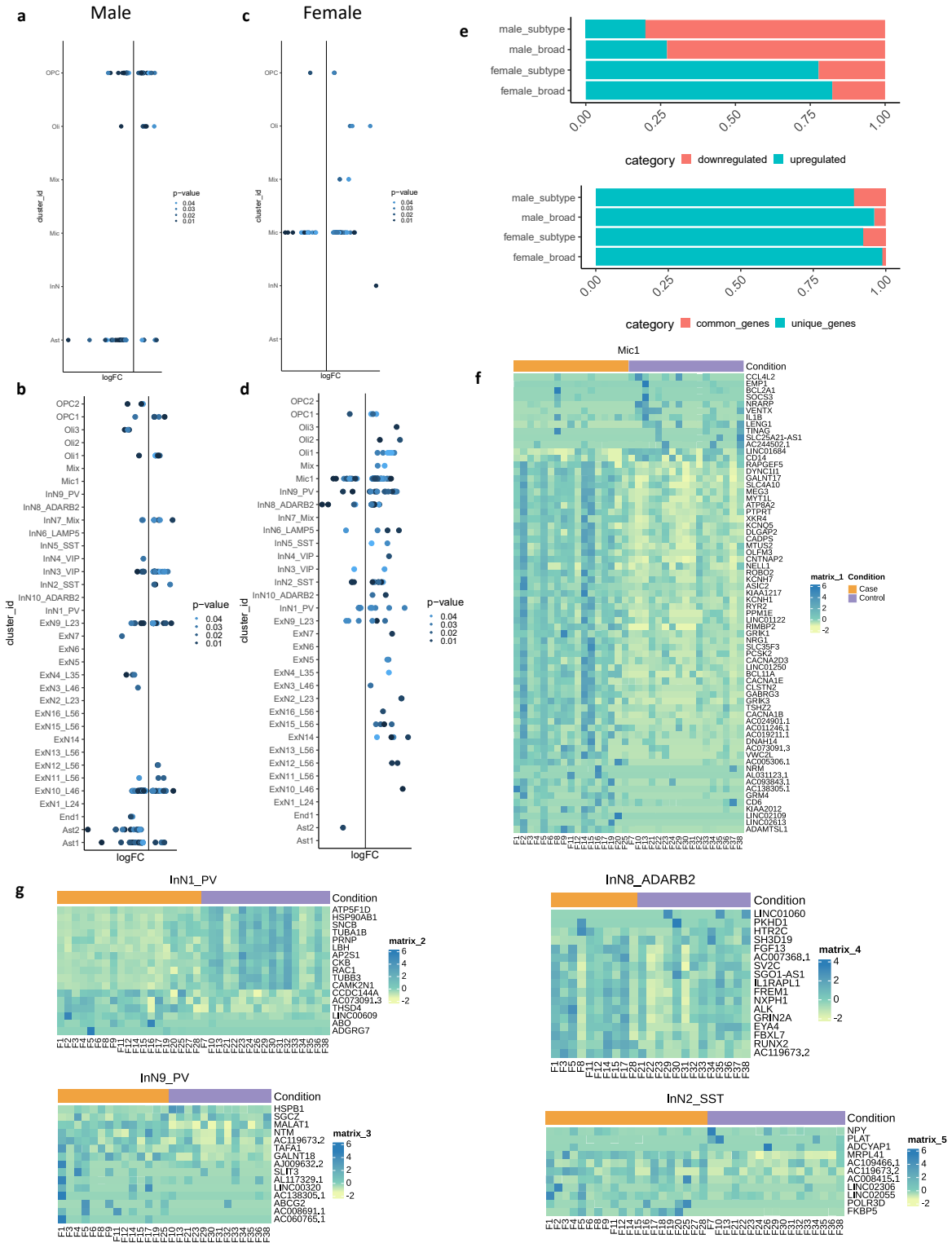

a
Male
c
Female
e
male_subtype
male_broad
female_subtype
female_broad
category downregulated upregulated
category common_genes unique_genes
b
d
cluster_id
logFC
p-value
f
Mic1
Condition
matrix_1
Case
Control
g
InN1_PV
InN8_ADARB2
InN9_PV
InN2_SST
matrix_2
matrix_3
matrix_4
matrix_5

Figure 3: Cell type specific differential gene expression in males and females with MDD. a-b) Distribution of differentially expressed genes in (a) broad cell types and (b) clusters with increased and decreased expression in male cases compared to controls. **c-d)** Distribution of differentially expressed genes in (c) broad cell types and (d) clusters with increased and decreased expression in female cases compared to controls. For a-d, points are colored by the corrected p-value for differential expression, and upregulated genes are plotted to the right of the midline while downregulated genes are plotted to the left. **e)** Barplots showing proportions of up and downregulated genes and unique and shared genes. For males, the majority of DEGs were decreased in expression in cases compared to controls both at the broad (110/151, 73%) and cluster levels (358/447, 80%) and most DEGs were cell type specific both at the broad (145/151, 96% unique DEGs) and cluster (398/447, 89% unique DEGs) level. For females, the majority of DEGs were upregulated both at the broad (70/85, 82%) and the cluster level (140/180, 78%) and most DEGs were cell type specific both at the broad (84/85, 99% unique DEGs) and cluster (166/180, 92% unique DEGs) level. **f-g)** Heatmaps showing the pseudobulk expression of differentially expressed genes in top clusters with highest number of DEGs in the female cluster level analysis – (f) microglia, (g) inhibitory neuronal clusters. For f-g, the plotted values are pseudobulk CPMs (counts per million) calculated with edgeR and muscat and scaled per row (by gene). For all heatmaps (f-g), the annotation bar at the top is colored orange for cases and purple for controls, and rows and columns are not clustered. Statistical testing corresponding to figures 3a-d were performed with the edgeR (glmQLFit, glmQLFtest), FDR (Benjamini & Hochberg) corrected p-values are plotted. Source data are provided as a Source Data file.

Figure 4

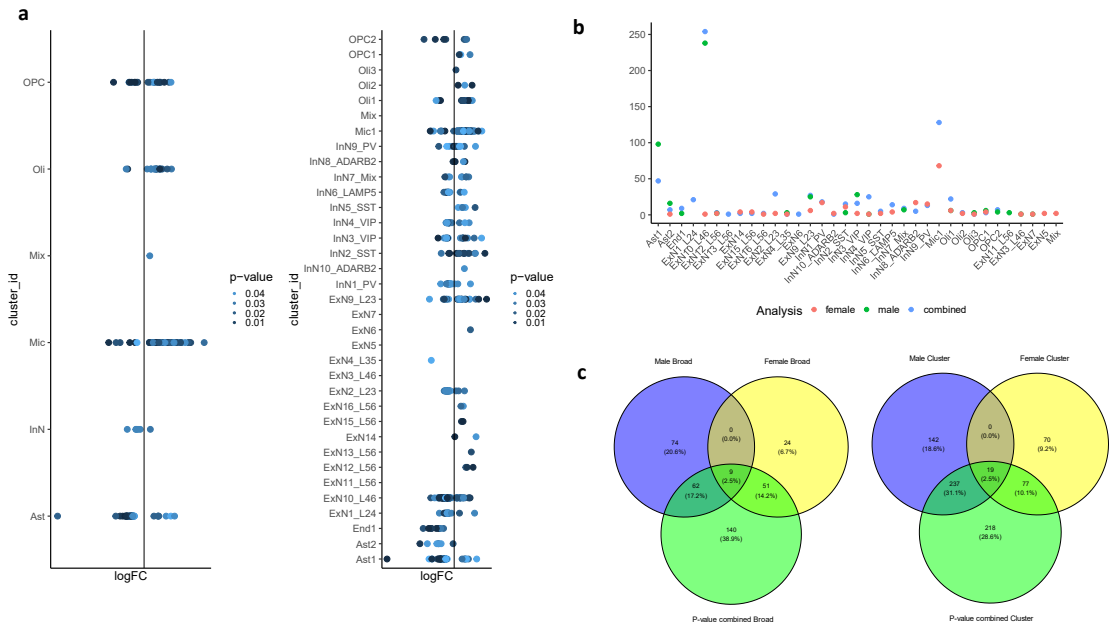


Figure 4: p-value combination meta-analysis results. a-b) Distribution of DEGs across the (left) broad cell types and (right) clusters after p-value combination meta-analysis. b) Numbers of DEGs (y-axis) in each cluster for the male analysis, female analysis, and meta-analysis. **c-d)** Overlap of meta-analysis DEGs with the individual analyses of the male and female datasets for (c) broad cell types and (d) clusters. The statistical test performed is Fisher combination of p-values as implemented in metaRNAseq, the test is one-sided, and the p-values were Benjamini Hochberg corrected. Source data are provided as a Source Data file.

Figure 5

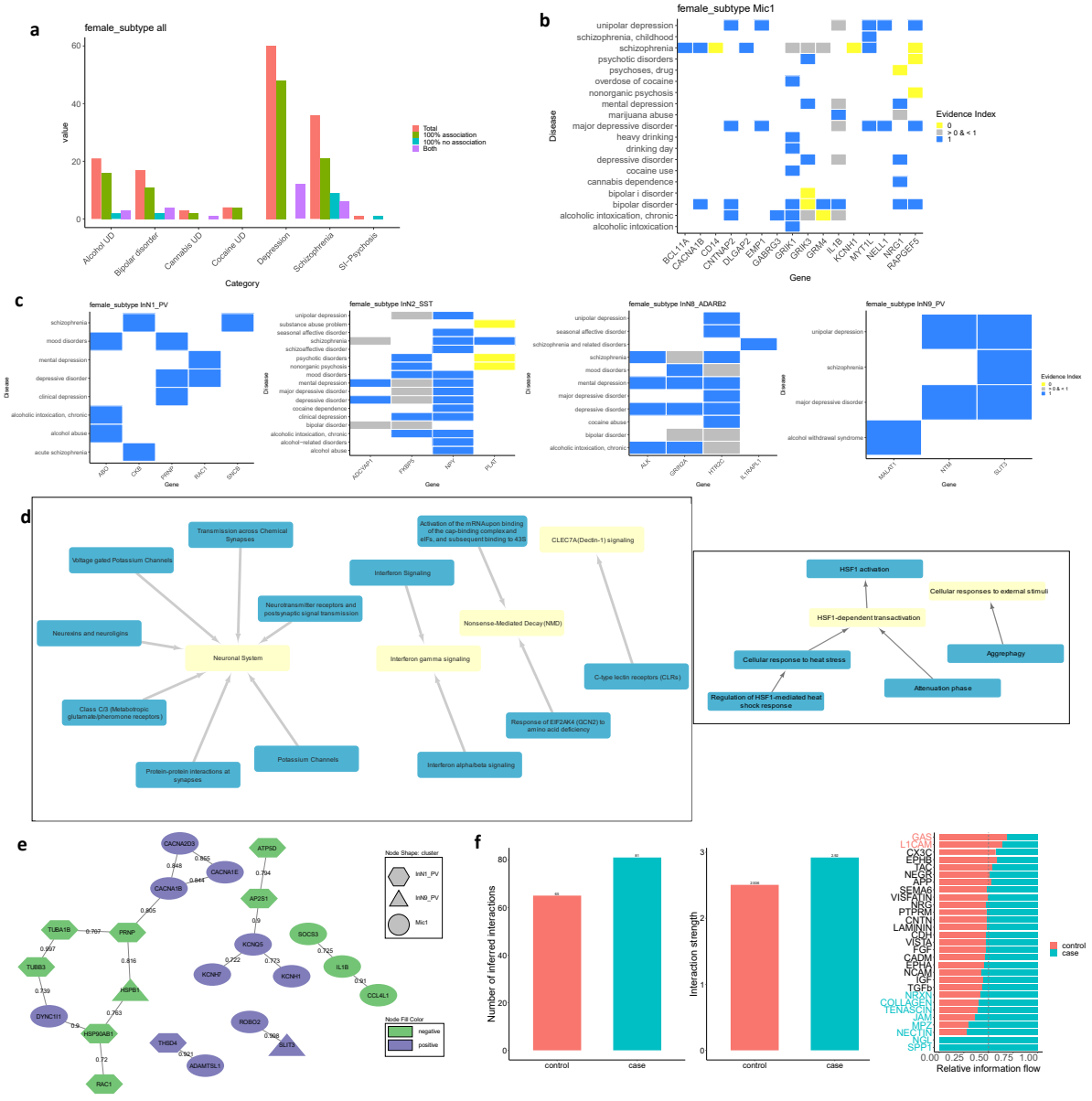


Figure 5: Characterization of cell type specific DEGs in females with MDD. a) PsyGeNET literature reported gene-disease association bar plot for all DEGs in the female cluster level analysis. The y-axis shows the number of gene-disease associations. “100% association” indicates all evidence is in support, “100% no association” indicates the opposite, while “Both” indicates

mixed support. **b-c)** Gene-disease association heatmaps for 5 clusters with the highest numbers of DEGs in females: (b) microglia, (c) inhibitory neuron clusters. Evidence index of 1 indicates that all literature supports the association, while 0 indicates that there is no support for the association. Values in between indicate partial support. **d)** Networks showing the relationship between main gene sets (yellow) and all gene sets (blue) with enrichment in pre-ranked GSEA with Reactome pathways in Mic1 (left) and InN9_PV (right) in females. Controlling for the overlap between gene sets, the main gene sets are independently enriched. **e)** STRING network showing DEGs in female microglia and PV interneurons whose protein products have reported interactions. The shape of the node represents the cluster in which the DEG was detected, and the color represents the direction of fold change in cases compared to controls. The numbers on the edges represent the confidence scores for the interactions. **f)** (left) Bar plots showing the number and strength of ligand-receptor communications within and between PV interneurons and microglia in cases and controls. (right) Relative strength of communication in different signaling pathways for cases and controls. Source data are provided as a Source Data file.

Figure 6

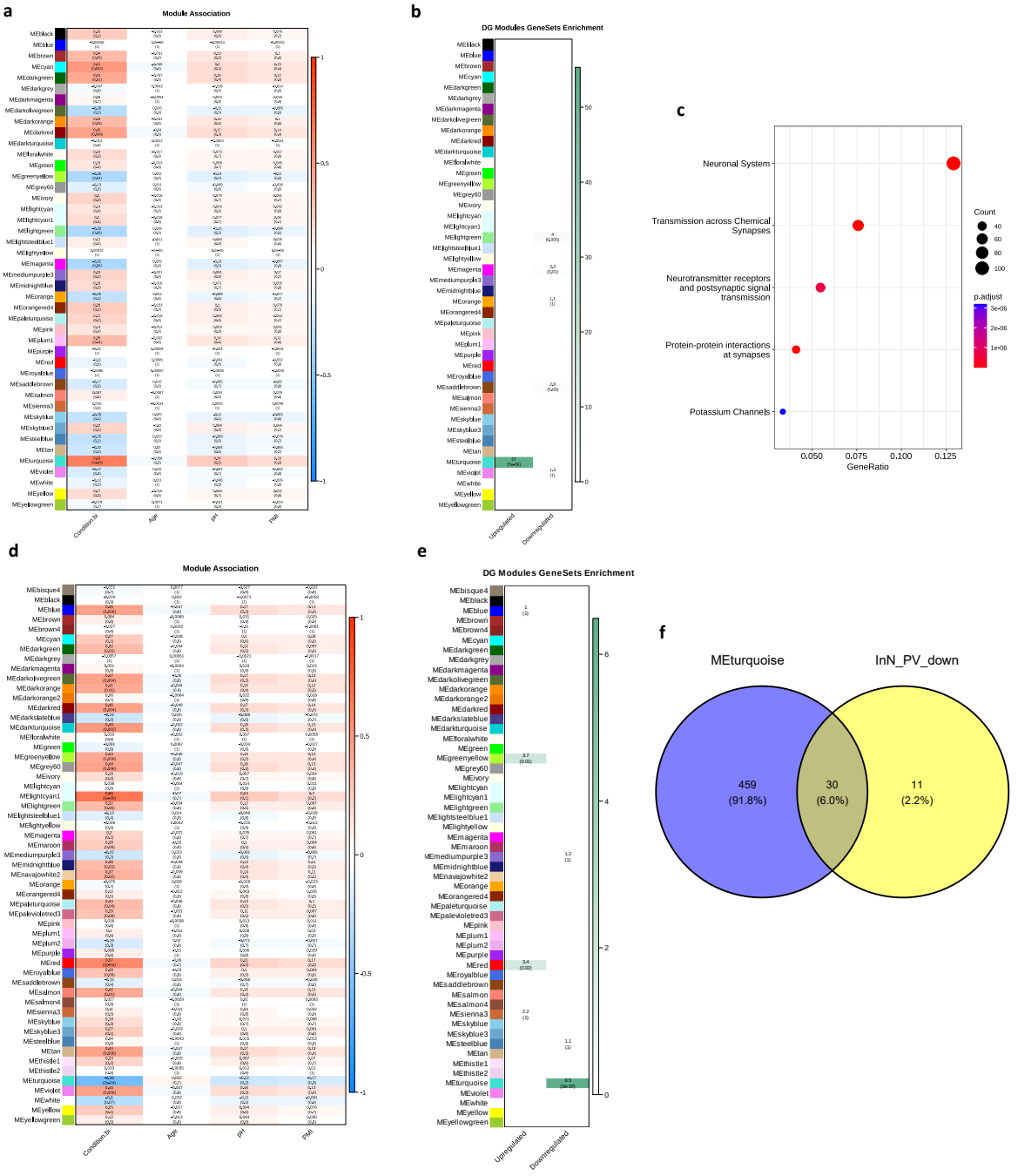
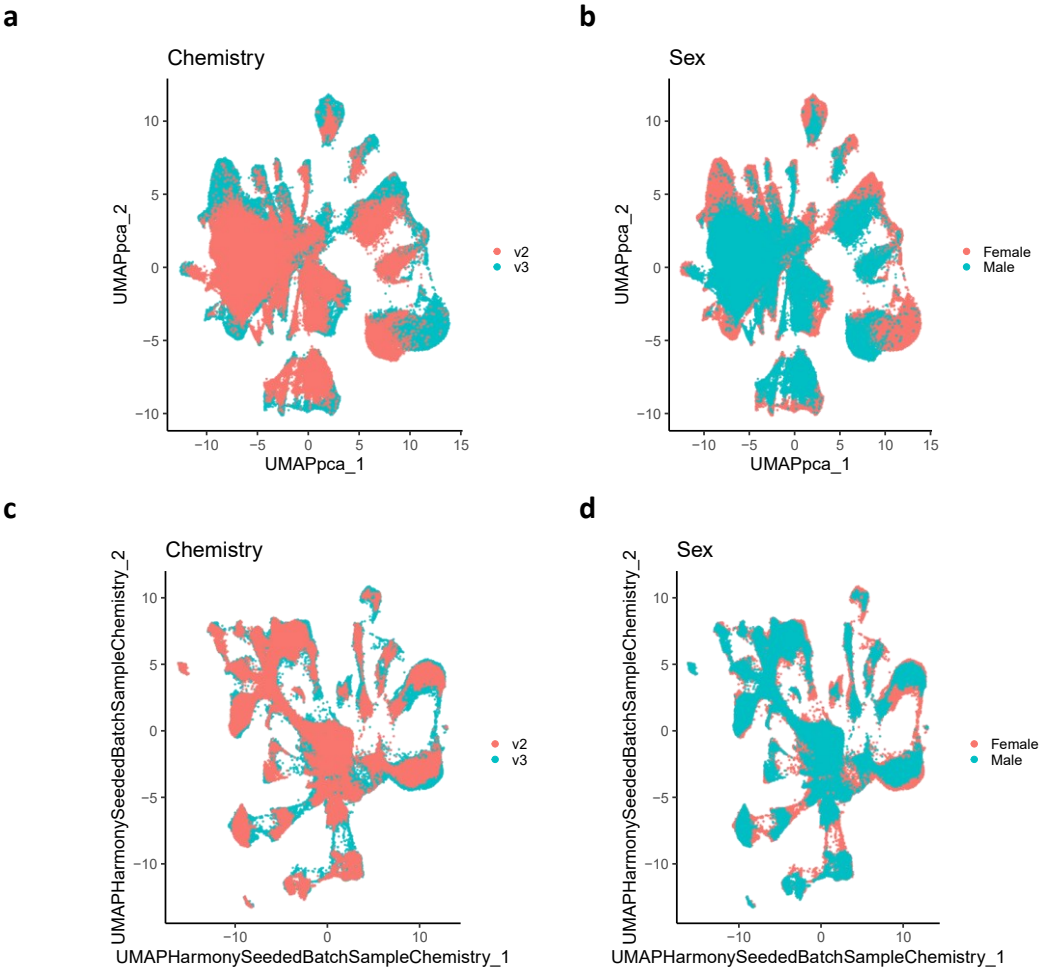


Figure 6: WGCNA results for microglia and PV interneurons in females. a) Heatmap showing the correlation and associated p-value, in parentheses, of Mic1 WGCNA module eigengenes with case-control status and covariates (age, pH, PMI). **b)** Heatmap showing the test-statistic and FDR

corrected p-value, in parentheses, for one-sided Fisher tests of overlap between the Mic1 WGCNA module member genes and DEGs in females in Mic1. **c)** Top Reactome pathway gene sets over-represented in Mic1 WGCNA, in the MEturqouise module using one-tailed hypergeometric testing. Uncorrected p-values are plotted. **d)** Heatmap showing the correlation and associated p-value, in parentheses, of InN_PV WGCNA module eigengenes with case-control status and covariates. **e)** Heatmap showing the test-statistic and FDR corrected p-value, in parentheses, for one-sided Fisher tests of overlap between the InN_PV WGCNA module member genes and DEGs in females the InN1_PV or InN9_PV clusters. **f)** Venn diagram showing the overlap of Reactome pathway gene sets enriched in the InN_PV WGCNA module MEturquoise (associated negatively with case status) and downregulated via GSEA in cases within InN1_PV or InN9_PV. For 6a, 6d the statistical test performed was a Pearson correlation as implemented in the WGCNA package and p-values are one-sided and uncorrected. Source data are provided as a Source Data file.

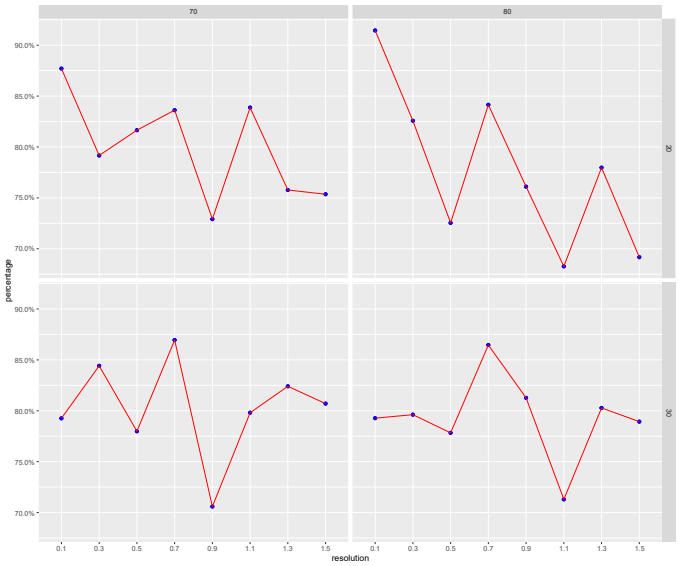
Supplementary Information

Supplementary Figures

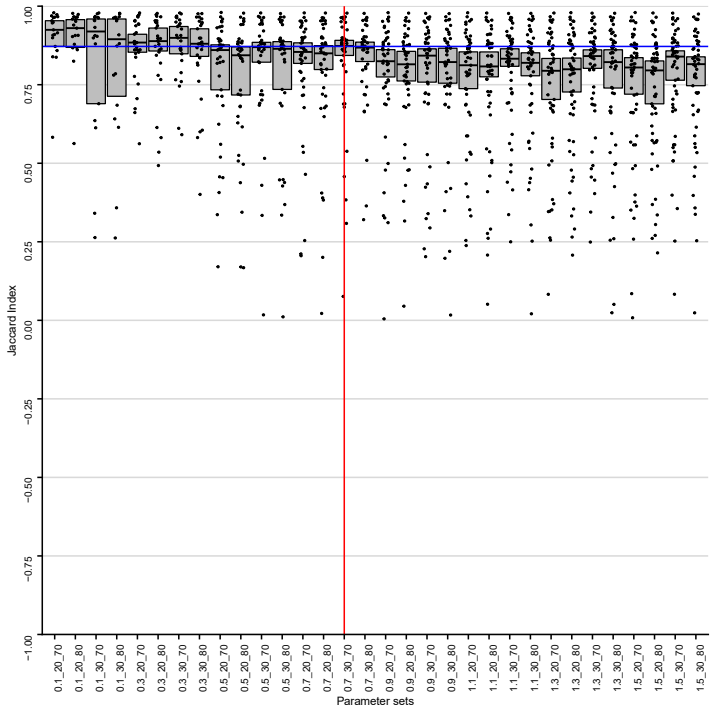


Supplementary Figure 1: UMAP plots using uncorrected and Harmony corrected PCA components. a-b) UMAP plot using uncorrected PCA components colored by chemistry and sex. **c-d)** UMAP plot using Harmony corrected PCA components colored by chemistry and sex. For UMAPs we used all 100 PC components or Harmony corrected PC components with other parameters set to default. Source data are provided as a Source Data file.

a

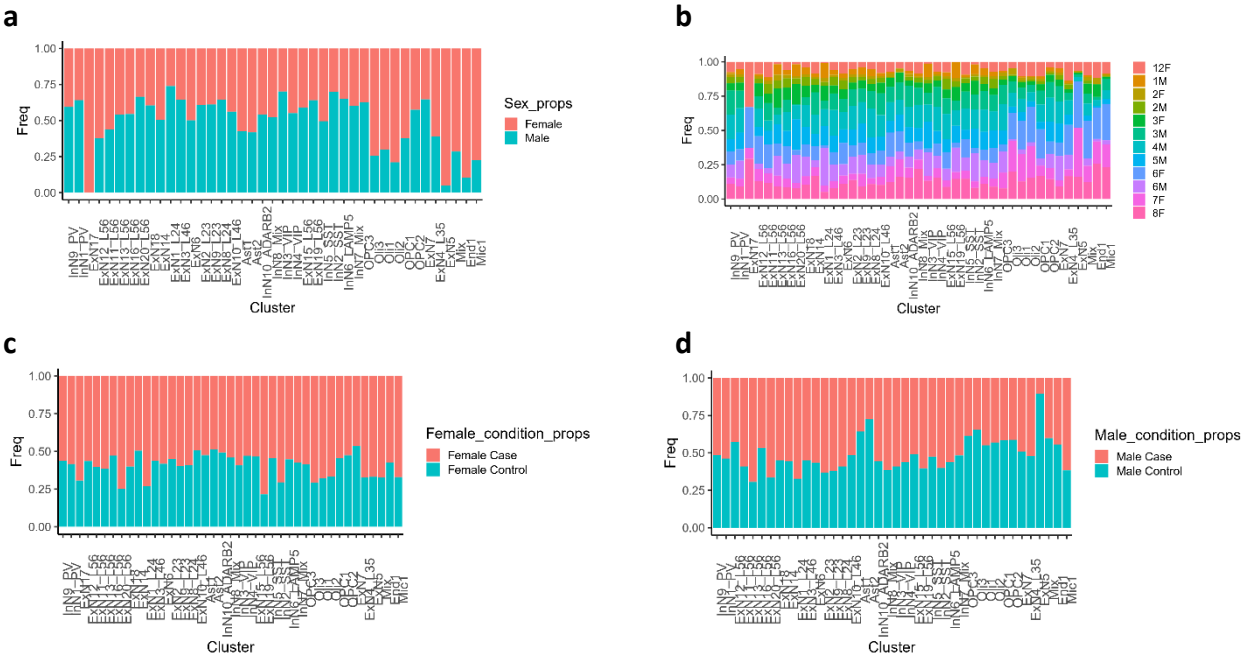


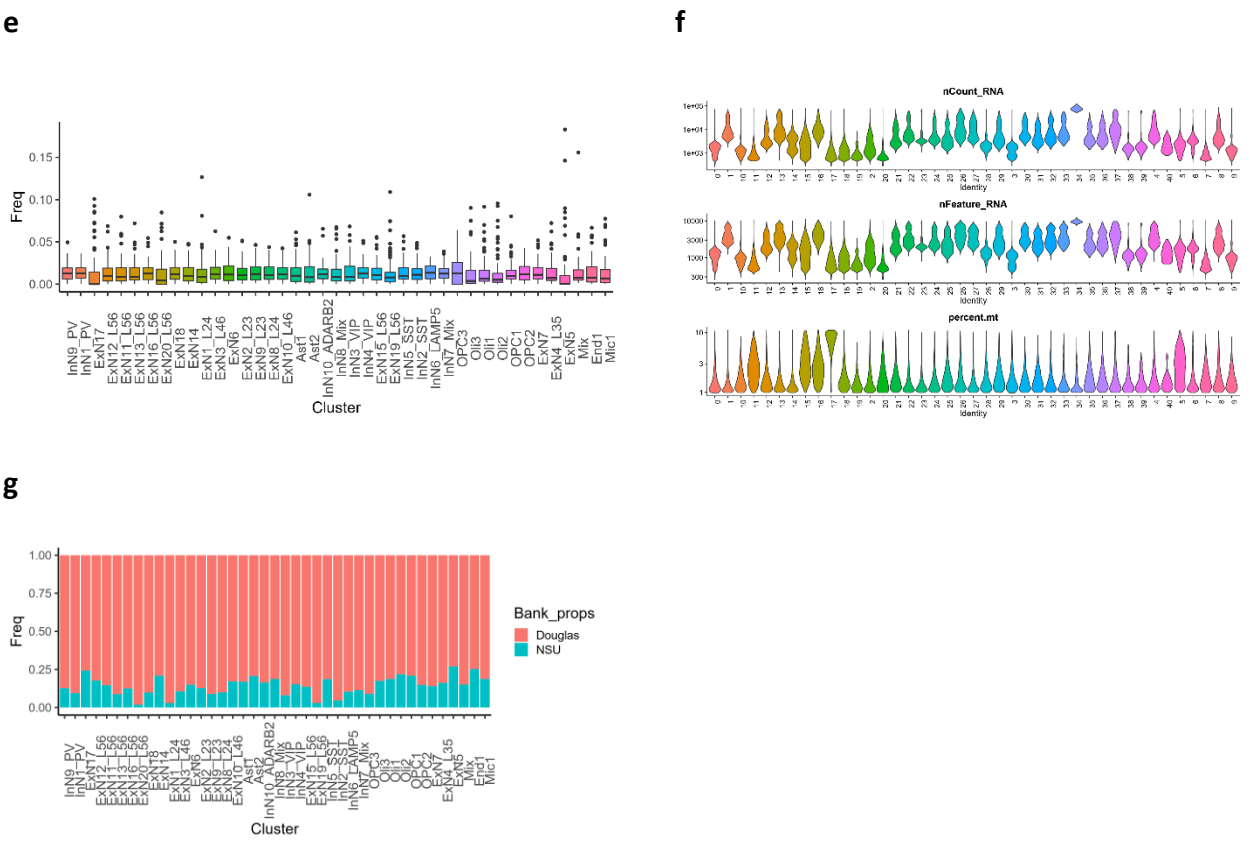
b



Supplementary Figure 2: Assessment of clustering parameters. a) scclusteval output showing the percentage of nuclei in stable clusters, as assessed by sub-sampling and Jaccard index calculation, using a range of clustering parameters. **b)** Boxplots showing the median Jaccard index for

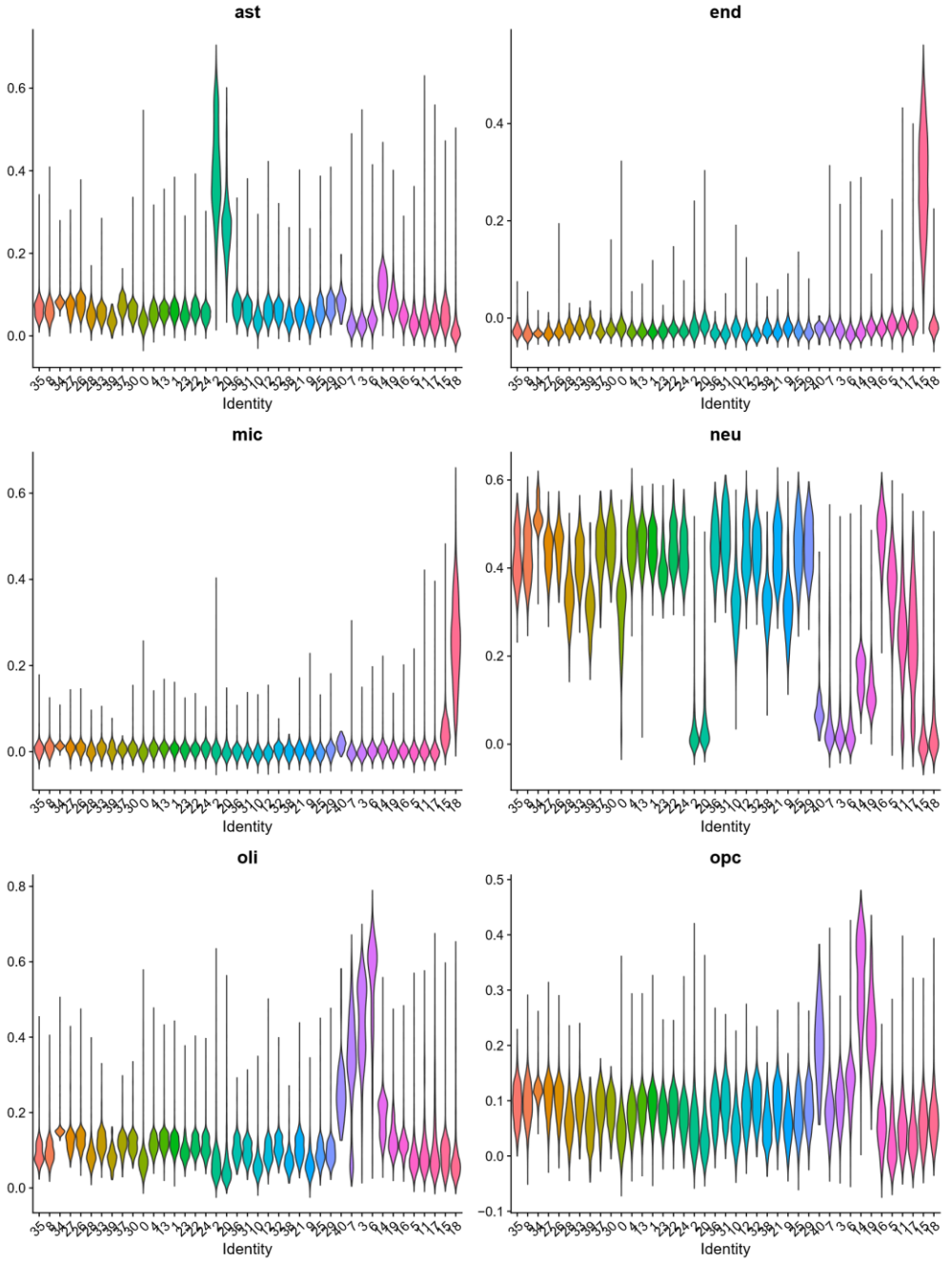
each cluster across sub-sampling with different parameter combinations for clustering. The individual dots correspond to the median Jaccard index, across all sub-samplings, for each cluster obtained using the given set of parameters. The boxes represent the bootstrapped 95% confidence interval for the median of the median Jaccard index across all clusters, for the given set of clustering parameters. The box starts at the lower bound of the confidence interval and extends to the higher bound, while the line in the center represents the actual median. The clustering parameters that provide the highest number of clusters, while crossing a bootstrapped threshold of median Jaccard index across clusters, are highlighted. Source data are provided as a Source Data file.



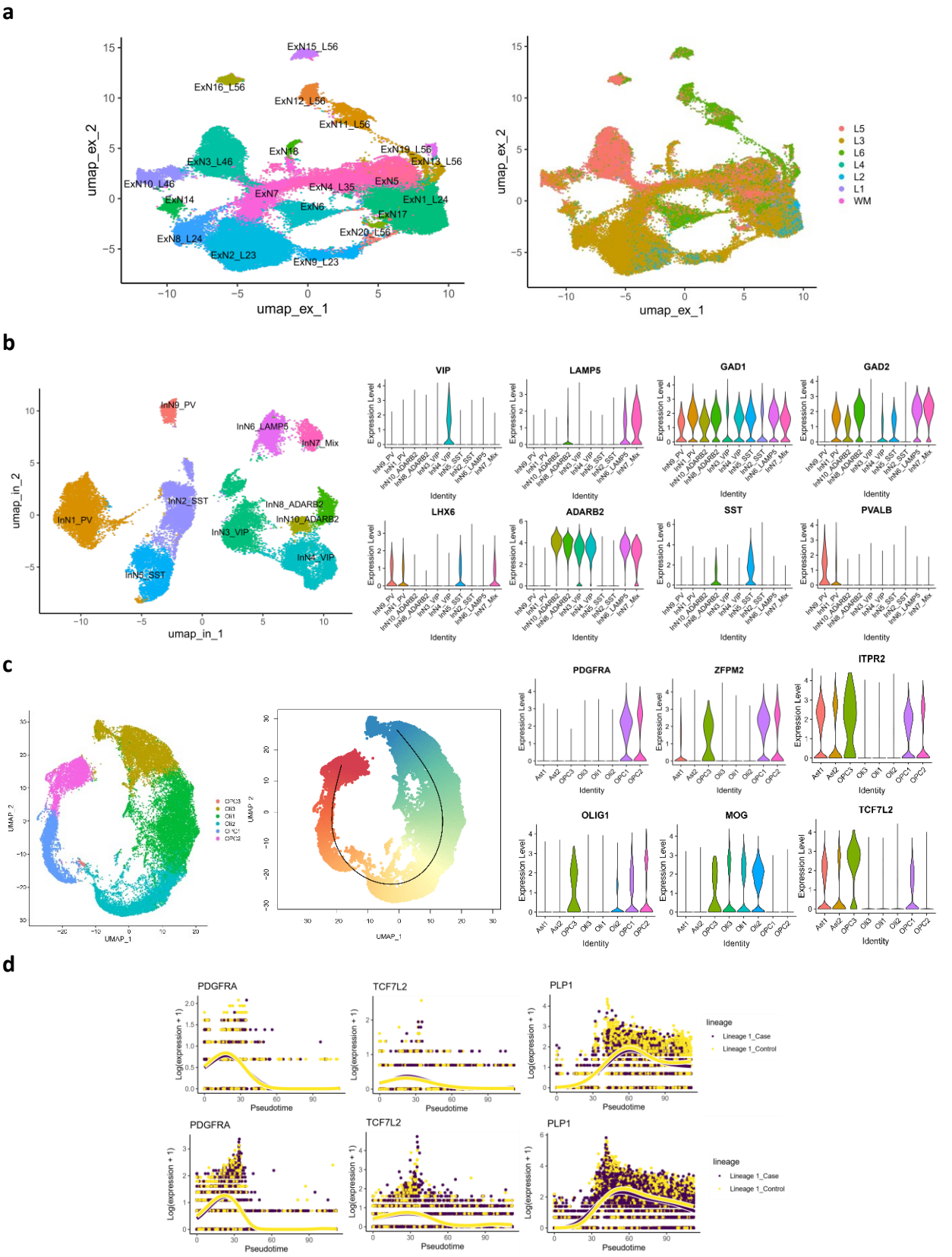


Supplementary Figure 3: Evaluation of clusters based on technical parameters. a) Proportions of nuclei from each sex in each cluster. **b)** Proportion of nuclei from each batch in each cluster. **c)** Proportion of nuclei from cases and controls among the female nuclei in each cluster. **d)** Proportion of nuclei from cases and controls among the male nuclei in each cluster. **e)** Proportion of nuclei from each library (mostly corresponding to subject) in each cluster (for each cluster n= 72 samples, i.e. libraries, from 71 biologically independent subjects). The middle line is the median. The lower and upper hinges correspond to the 25th and 75th percentiles. Upper and lower whiskers extend from the upper or lower hinges to the largest or smallest value no further than 1.5 times the inter-quartile range from the hinge, where the inter-quartile range is the distance between the first and third quartiles. Points beyond the end of the whiskers are plotted individually. **f)** Violin plots for number of molecules detected, number of genes detected, and

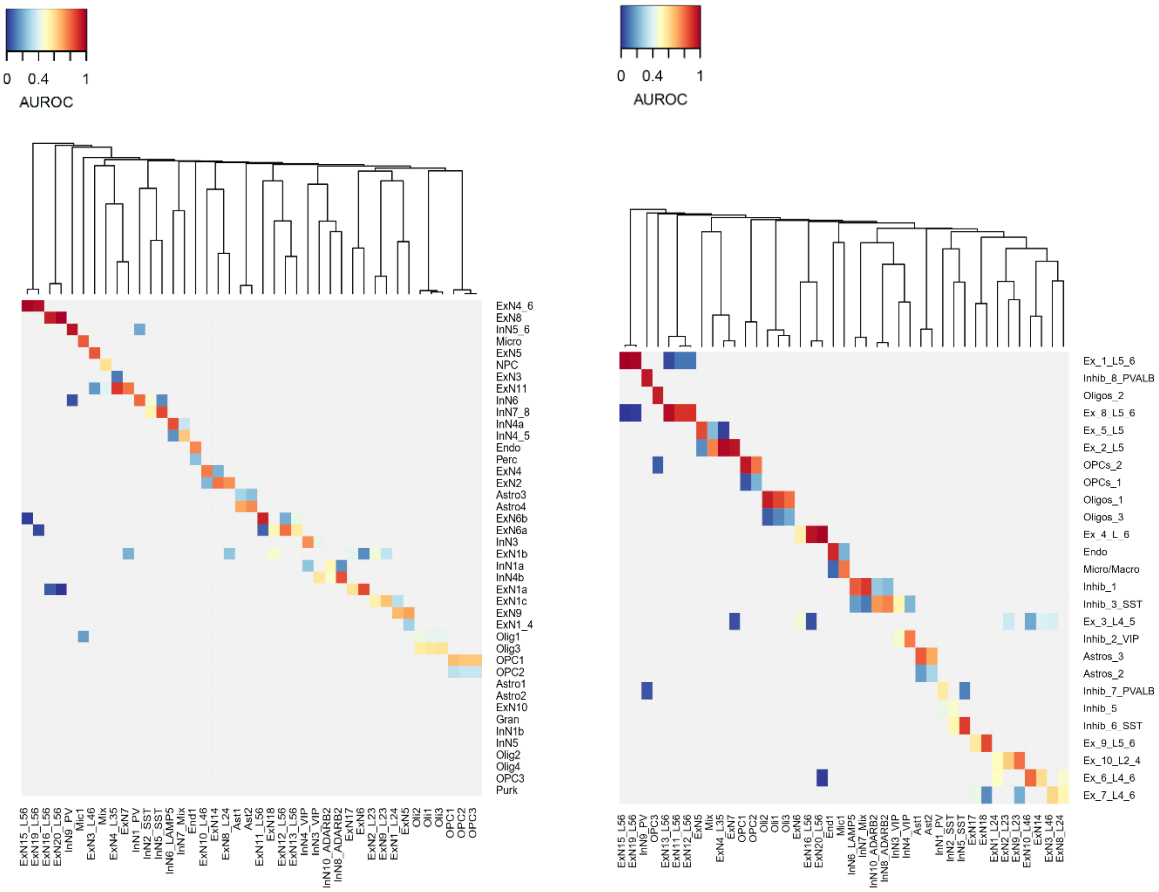
percentage of mitochondrial reads per nuclei split by cluster. Mitochondrial gene counts were removed for downstream analysis, after calculating the mitochondrial read percentage. Y-axis is in log-scale. **g)** Proportions of nuclei from each brain bank in each cluster. For 3a-e, and g, "Freq" stands for frequency representing the proportion of cells. Correspondence between numbered clusters and cluster names in 3f: 0 - ExN1_L24, 1 - ExN2_L23, 2 - Ast1, 3 - Oli1, 4 - ExN3_L46, 5 - ExN4_L35, 6 - Oli2, 7 - Oli3, 8 - InN1_PV, 9 - InN2_SST, 10 - InN3_VIP, 11 - ExN5, 12 - InN4_VIP, 13 - ExN6, 14 - OPC1, 15 - End1, 16 - ExN7, 17 - Mix, 18 - Mic1, 19 - OPC2, 20 - Ast2, 21 - InN5_SST, 22 - ExN8_L24, 23 - ExN9_L23, 24 - ExN10_L46, 25 - InN6_LAMP5, 26 - ExN11_L56, 27 - ExN12_L56, 28 - ExN13_L56, 29 - InN7_Mix, 30 - ExN14, 31 - InN8_ADARB2, 32 - ExN15_L56, 33 - ExN16_L56, 34 - ExN17, 35 - InN9_PV, 36 - InN10_ADARB2, 37 - ExN18, 38 - ExN19_L56, 39 - ExN20_L56, 40 - OPC3. Source data are provided as a Source Data file.



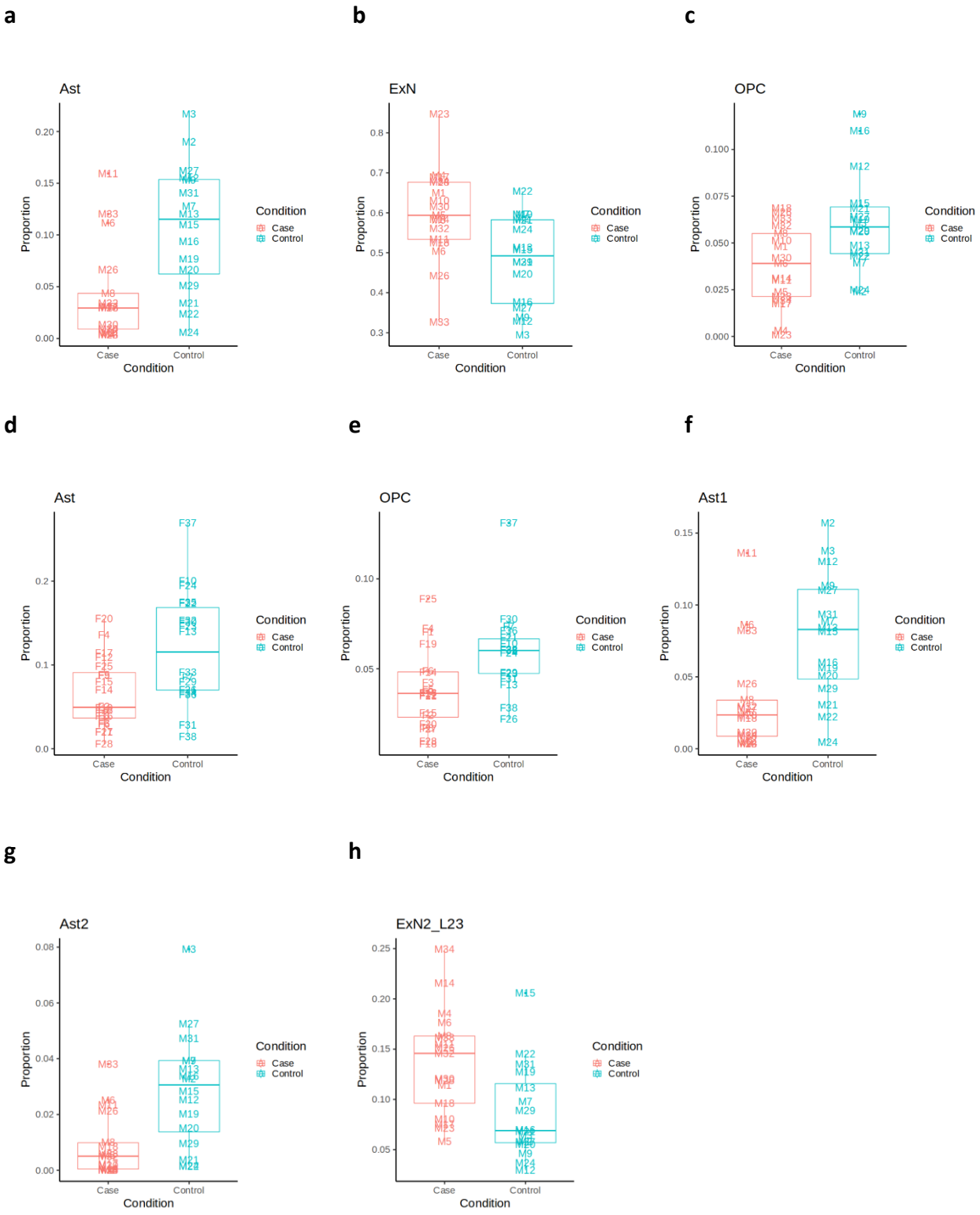
Supplementary Figure 4: Assessment of cell type scores per cluster. Violin plots showing module scores for major brain cell type marker genes from BRETIGEA in each cluster. Correspondence between numbered clusters and cluster names same as in Supplementary Figure 3. Source data are provided as a Source Data file.


a
ExN15_L56
ExN16_L56
ExN12_L56
ExN11_L56
ExN18
ExN19_L56
ExN13_L56
ExN3_L46
ExN10_L46
ExN14
ExN7
ExN4_L35
ExN5
ExN1_L24
ExN6
ExN17
ExN20_L56
ExN8_L24
ExN2_L23
ExN9_L23
umap_ex_1
umap_ex_2
L5
L3
L6
L4
L2
L1
WM
b
InN9_PV
InN6_LAMP5
InN7_Mix
InN2_SST
InN8_ADARB2
InN1_PV
InN10_ADARB2
InN3_VIP
InN4_VIP
InN5_SST
umap_in_1
umap_in_2
VIP
LAMP5
GAD1
GAD2
LHX6
ADARB2
SST
PVALB
Expression Level
Identity
c
UMAP_1
UMAP_2
PDGFRA
ZFPM2
ITPR2
OLIG1
MOG
TCF7L2
Ast1
Ast2
OPC3
Oli3
Oli1
Oli2
OPC1
OPC2
d
PDGFRA
TCF7L2
PLP1
Log(expression + 1)
Pseudotime
lineage
Lineage 1_Case
Lineage 1_Control

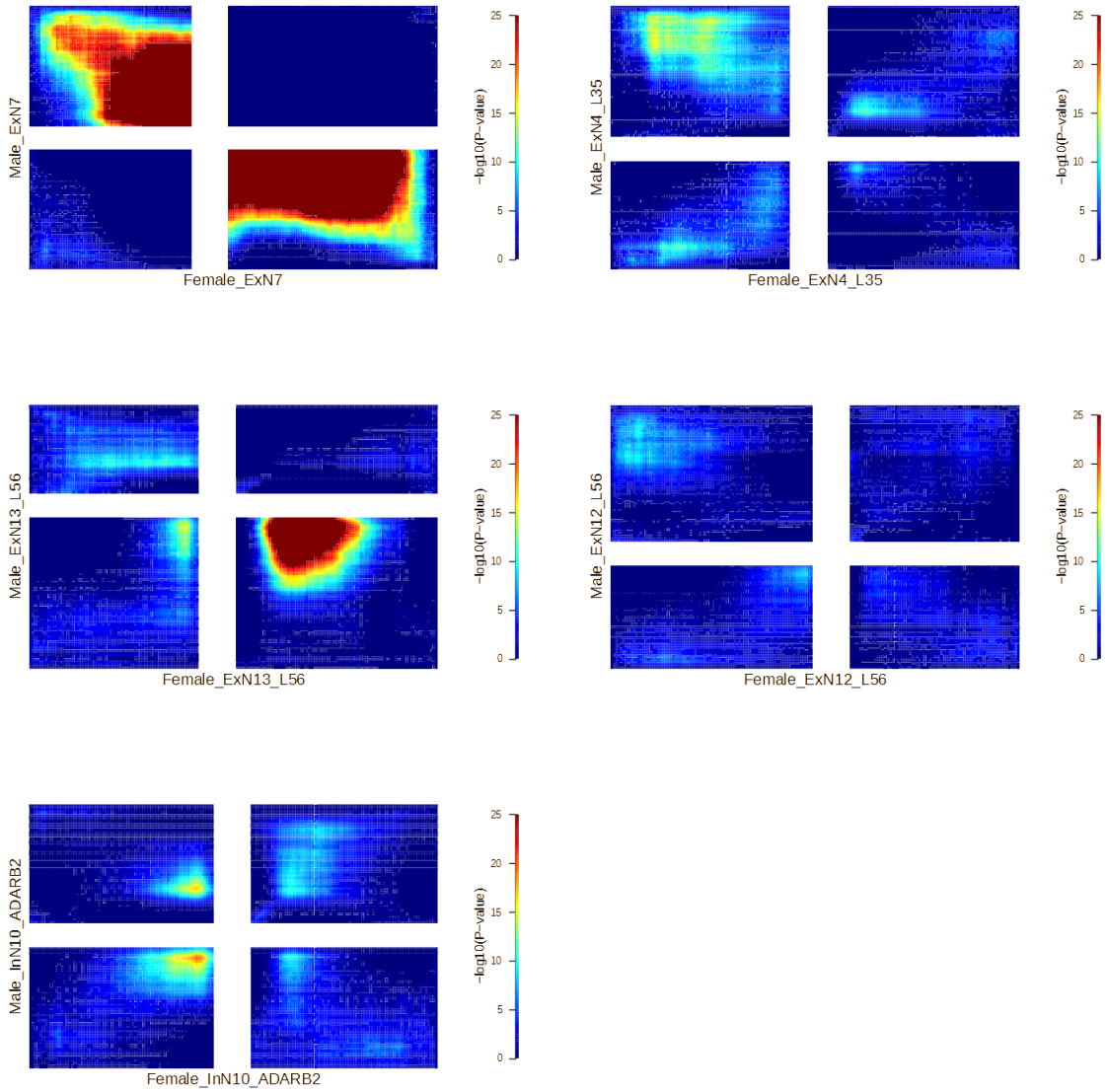
Supplementary Figure 5: Annotation of cell sub-types among clusters. a) (left) UMAP plot showing the 20 clusters of excitatory neurons identified. (right) UMAP plot showing the predicted layer labels for excitatory neurons using Seurat label transfer from a spatial transcriptomics dataset of the human dlPFC[1]. **b)** (left) UMAP plot showing the 10 clusters of inhibitory neurons identified. (right) Violin plots showing the expression of maker genes of inhibitory neurons and their known subtypes. **c)** (left) UMAP plot showing the 6 oligodendrocyte lineage (OL) clusters. (middle) Same UMAP plot as colored according to the pseudotime trajectory calculated (early to late pseudotime points colored from red to blue). (right) Violin plots showing expression of selected markers of the OL in the respective clusters. **d)** Smoothed expression fit to depict the variation in expression of selected OL genes along pseudotime using GAMS in males (top) and females (bottom). For UMAPs in (a) and (b) we used the first 50 Harmony components, and n.neighbors = 20. Source data are provided as a Source Data file.



Supplementary Figure 6: Comparison of clustering to published datasets. Best hits plot from MetaNeighbor showing the correspondence between clusters defined in our dataset and clusters defined in (left) the STAB dataset and (right) in our previous analysis of the male cohort alone. Clusters defined in this study are represented on the x-axis while clusters defined in the STAB study and the previous analysis of the male data are represented along the y-axis. Source data are provided as a Source Data file.

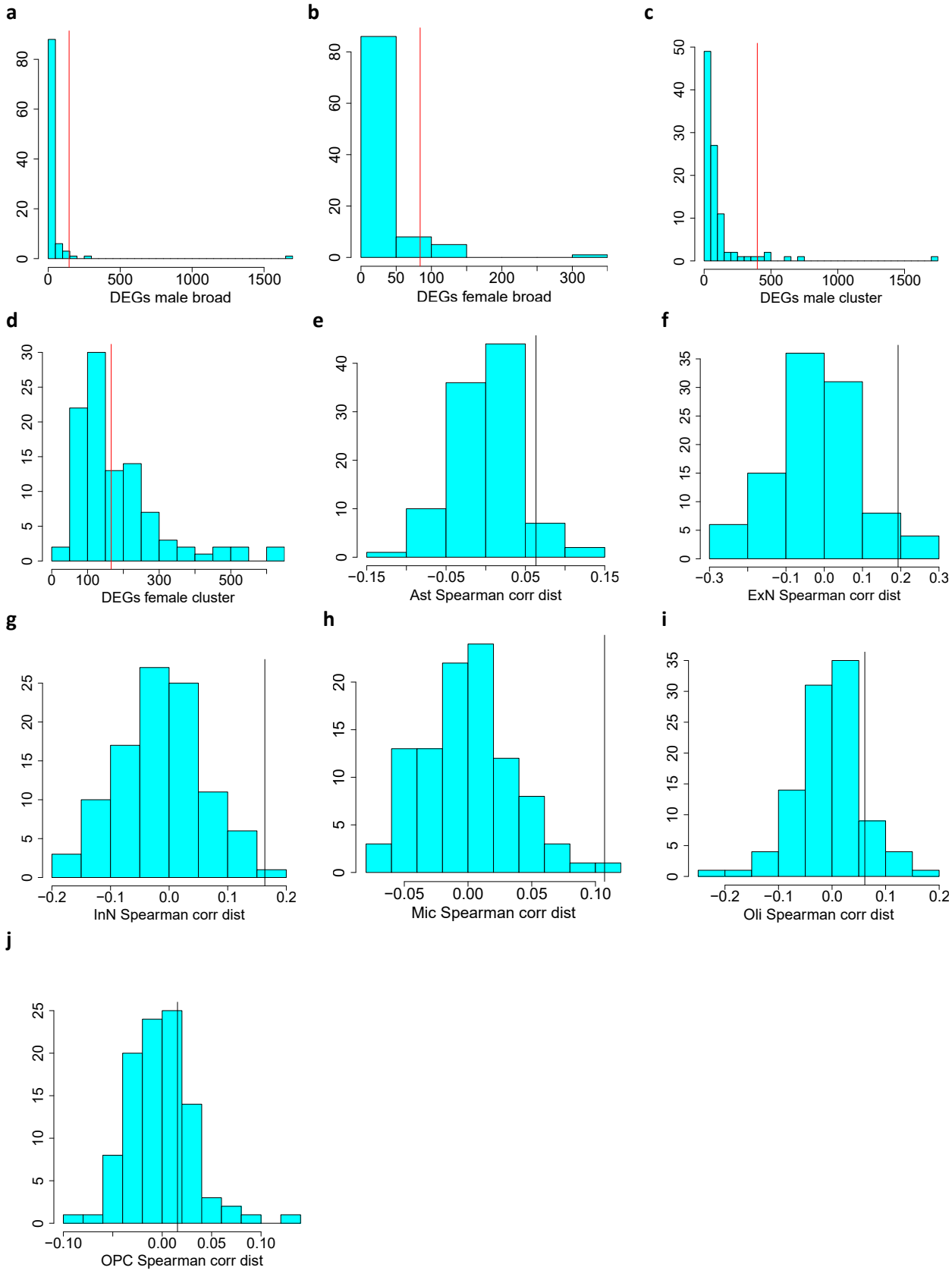

a
Ast
b
ExN
c
OPC
d
Ast
e
OPC
f
Ast1
g
Ast2
h
ExN2_L23
Proportion
Condition
Case
Control

Supplementary Figure 7: Differences in cell type proportion between cases and controls in each sex. Boxplots showing cell type proportions with significant differences between MDD cases and controls, using Wilcoxon tests performed on the male and female datasets separately (n= 17 cases, 16 controls, for males and n=20 cases, 18 controls for females, representing biologically independent samples for each cluster). At the broad level, there was a significant decrease in males in **a)** Ast (FDR 0.0130) and **b)** OPC (FDR 0.0325) and an increase in **c)** ExN (FDR 0.0325). In females, at the broad level the decrease in **d)** Ast was close to significance (FDR 0.0536) and the decrease in **e)** OPC (0.0466) was significant. At the cluster level there were no significant differences in the female dataset, but there was a decrease in **f)** Ast1 (FDR 0.0364) and **g)** Ast2 (FDR 0.0364) in males and an increase in **h)** ExN2_L23 (FDR 0.0492). Two-tailed Wilcoxon tests were performed. The middle line is the median. The lower and upper hinges correspond to the 25^{th} and 75^{th} percentiles. Upper and lower whiskers extend from the upper or lower hinges to the largest or smallest value no further than 1.5 times the inter-quartile range from the hinge, where the inter-quartile range is the distance between the first and third quartiles. Points beyond the end of the whiskers are plotted individually. Source data are provided as a Source Data file.

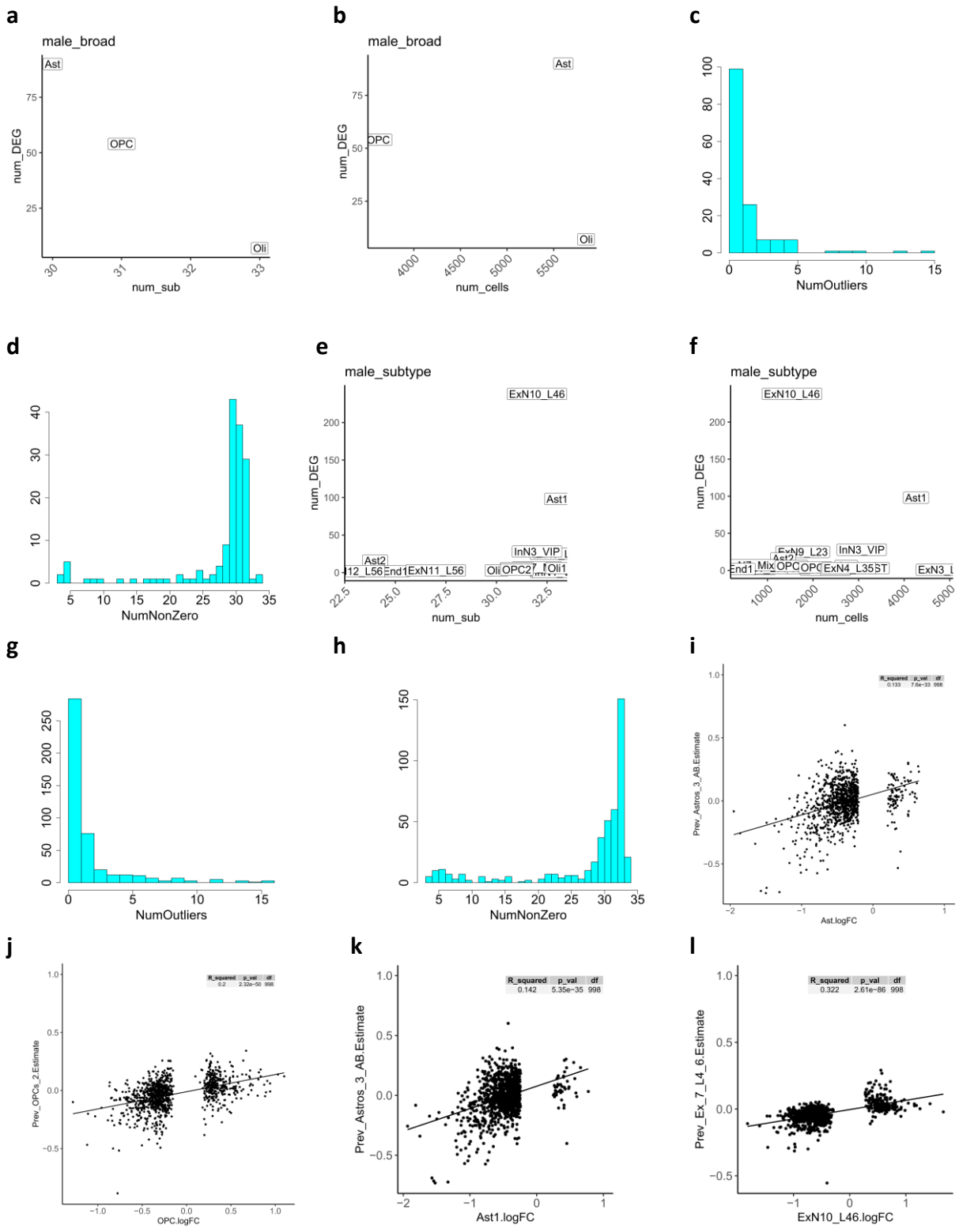


Supplementary Figure 8: RRHO plots for discordant clusters. RRHO plots showing the discordant relationship between males and females for patterns in depression-associated gene expression difference in several excitatory and inhibitory neuronal clusters. Interpretation of the plots is similar to Figure 2. For RRHO plots comparing broad cell types the color scale maximum was set to a -log10(p-value) of 50, and for RRHO plots comparing clusters the color scale maximum was

set to a -log10(p-value) of 25 for ease of comparison. RRHO2 uses one-sided hypergeometric tests, the p-values plotted here are uncorrected. Source data are provided as a Source Data file.

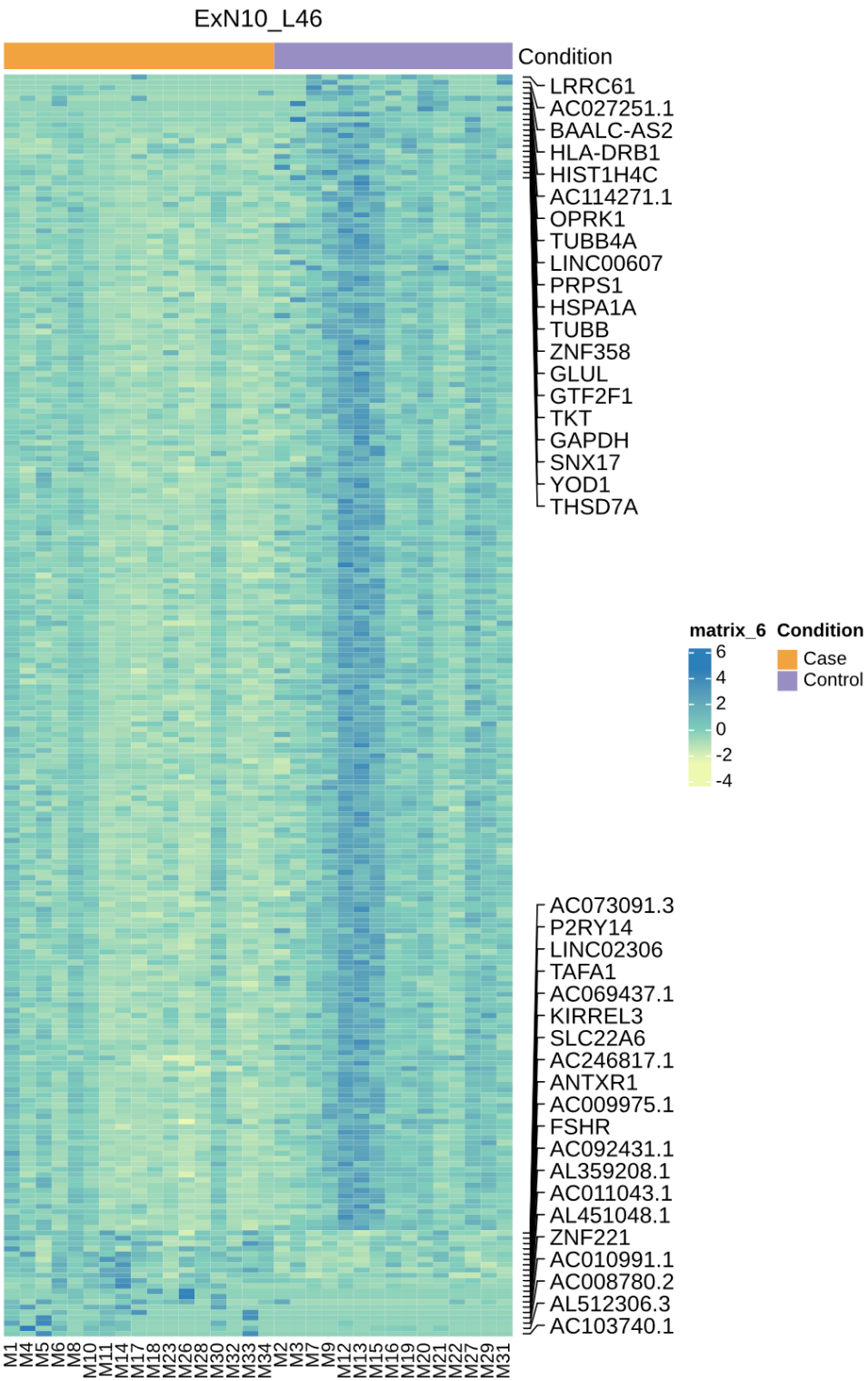

a
DEGs male broad
b
DEGs female broad
c
DEGs male cluster
d
DEGs female cluster
e
Ast Spearman corr dist
f
ExN Spearman corr dist
g
InN Spearman corr dist
h
Mic Spearman corr dist
i
Oli Spearman corr dist
j
OPC Spearman corr dist

Supplementary Figure 9: Permutation assessment for differential expression. Distribution of numbers of unique DEGs using permuted case versus controls labels in: **a)** males at the broad level, **b)** females at the broad level, **c)** males at the cluster level, and **d)** females at the cluster level. Red vertical lines indicate the number of unique DEGs obtained with real labels and show that the real labels result in higher number of unique DEGs than expected by chance with random case versus control annotations for both sexes at the broad level, and for males at the cluster level. **e-j)** Distribution of Spearman correlation coefficients when correlating gene scores based on male and female dataset differential gene expression analyses with permuted labels, for broad cell types. The black vertical line denotes the Spearman correlation coefficient obtained with the real case control labels and is higher than the majority of coefficients obtained with randomly permuted cases versus control labels for Ast, ExN, InN, and Mic and to a lesser extent for Oli and OPC. In all plots, the y-axis represents the number of iterations of permutations. Source data are provided as a Source Data file.

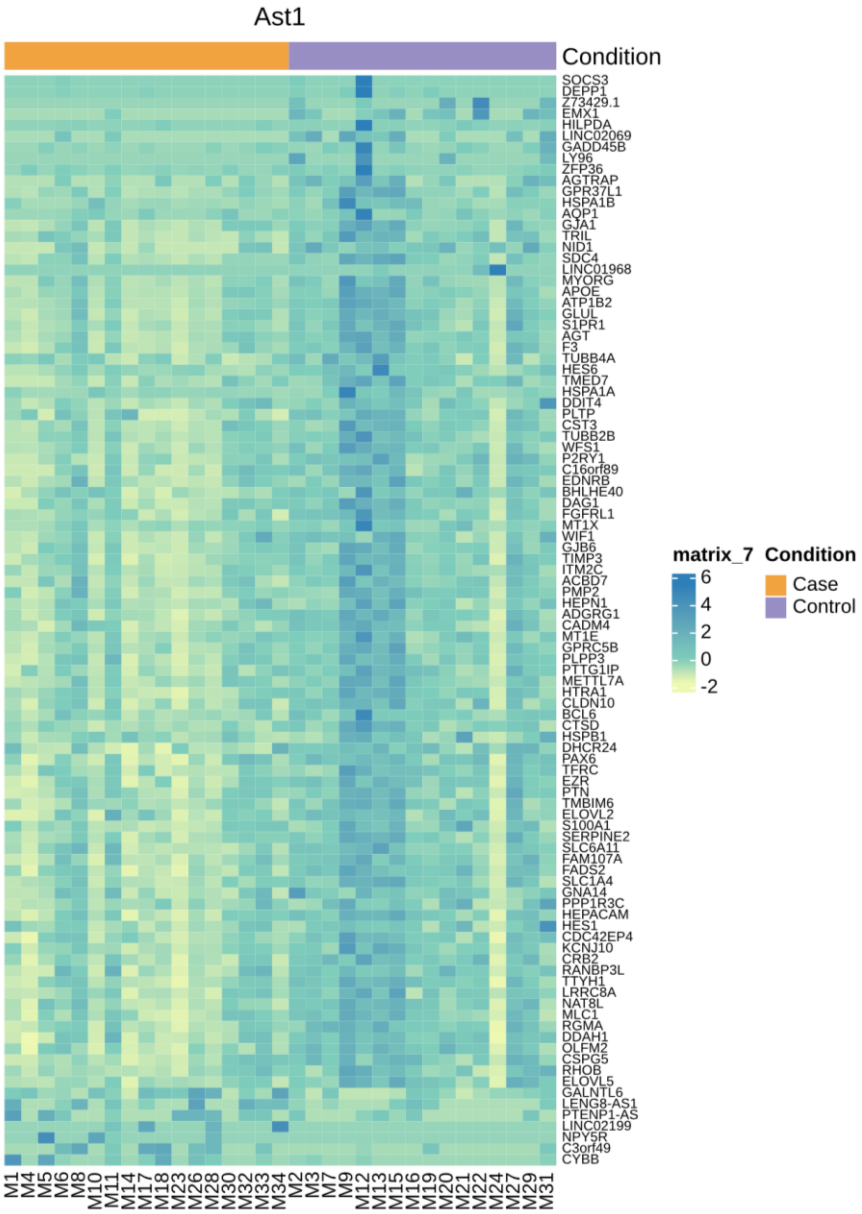

a
male_broad
Ast
OPC
Oli
num_DEG
num_sub
b
male_broad
Ast
OPC
Oli
num_DEG
num_cells
c
NumOutliers
d
NumNonZero
e
male_subtype
ExN10_L46
Ast1
Ast2
InN3_VIP
num_DEG
num_sub
f
male_subtype
ExN10_L46
Ast1
InN3_VIP
num_DEG
num_cells
g
NumOutliers
h
NumNonZero
i
Prev_Astros_3_AB.Estimate
Ast.logFC
j
Prev_OPCs_2.Estimate
OPC.logFC
k
R_squared p_val df
0.142 5.35e−35 998
Prev_Astros_3_AB.Estimate
Ast1.logFC
l
R_squared p_val df
0.322 2.61e−86 998
Prev_Ex_7_L4_6.Estimate
ExN10_L46.logFC

Supplementary Figure 10: Evaluation of male differential expression results. Number of DEGs in each broad cell type for males plotted against **(a)** the number of subjects included in DEG analysis and **(b)** the total number of nuclei in the broad cell type. For DEGs in males at the broad cell type level **(c)** the distribution of number of subjects flagged as possible outliers and **(d)** the distribution of subjects with non-zero expression. Number of DEGs in each cluster for males plotted against **(e)** the number of subjects included in DEG analysis and **(f)** the total number of nuclei in the cluster. For DEGs in males at the cluster level **(g)** the distribution of number of subjects flagged as possible outliers and **(h)** the distribution of subjects with non-zero expression. Possible outliers were not removed from the analyses but were assessed as a quality metric for DEG analysis. The plots include the Mix cluster results. **i-l)** Scatter plots showing the relationship between (linear regression and corresponding statistics) the estimated effects per gene from our previous analysis of the male data and the log fold changes from our current analysis for cell type specific case-control differences in males for similar pairs of clusters. Source data are provided as a Source Data file.

a

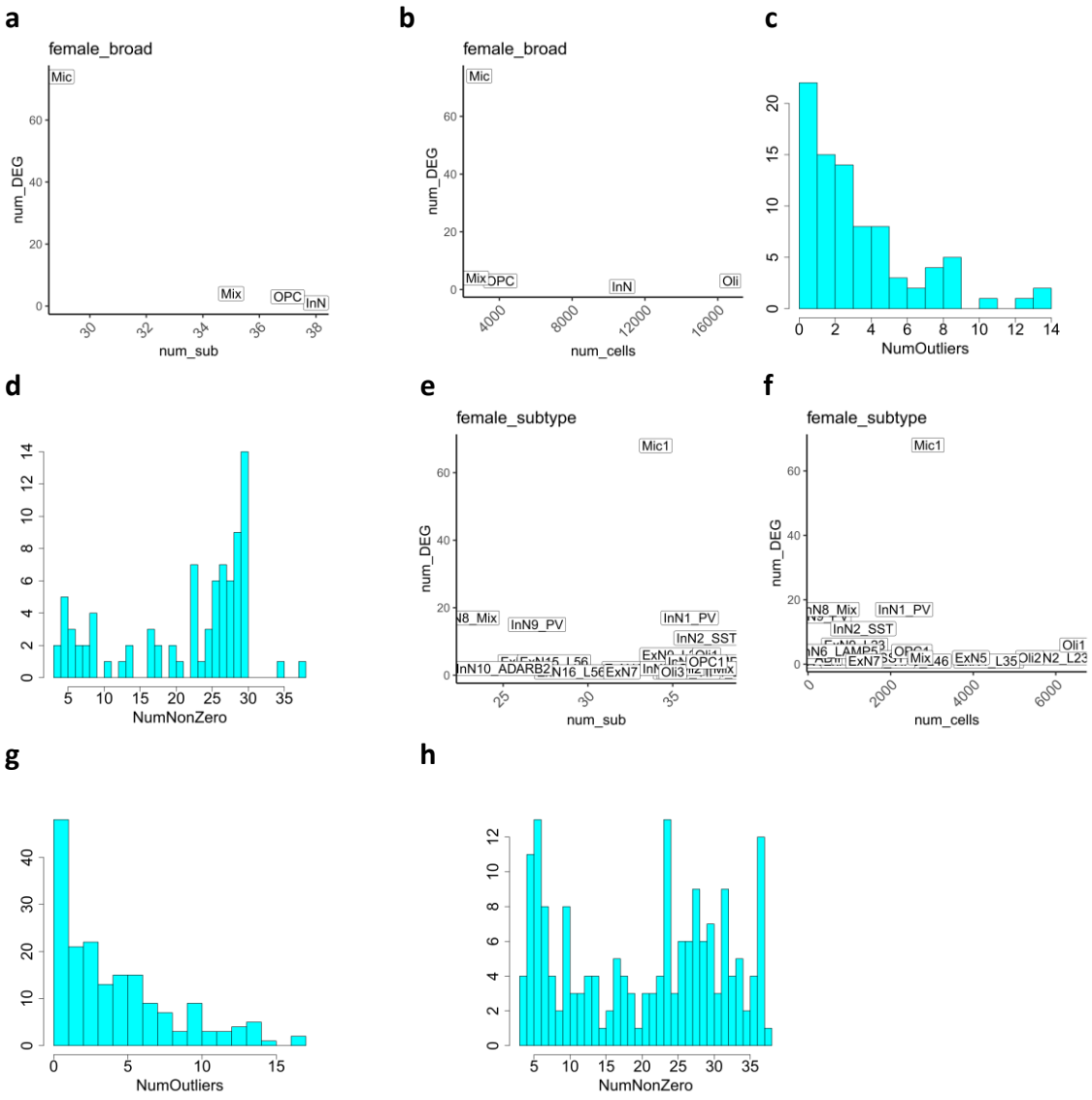

ExN10_L46
Condition
LRRC61
AC027251.1
BAALC-AS2
HLA-DRB1
HIST1H4C
AC114271.1
OPRK1
TUBB4A
LINC00607
PRPS1
HSPA1A
TUBB
ZNF358
GLUL
GTF2F1
TKT
GAPDH
SNX17
YOD1
THSD7A
matrix_6
6
4
2
0
-2
-4
Condition
Case
Control
AC073091.3
P2RY14
LINC02306
TAFA1
AC069437.1
KIRREL3
SLC22A6
AC246817.1
ANTXR1
AC009975.1
FSHR
AC092431.1
AL359208.1
AC011043.1
AL451048.1
ZNF221
AC010991.1
AC008780.2
AL512306.3
AC103740.1
M1 M4 M5 M6 M8 M10 M11 M14 M17 M18 M23 M26 M28 M30 M32 M33 M34 M2 M3 M7 M9 M12 M13 M15 M16 M19 M20 M21 M22 M27 M29 M31

b

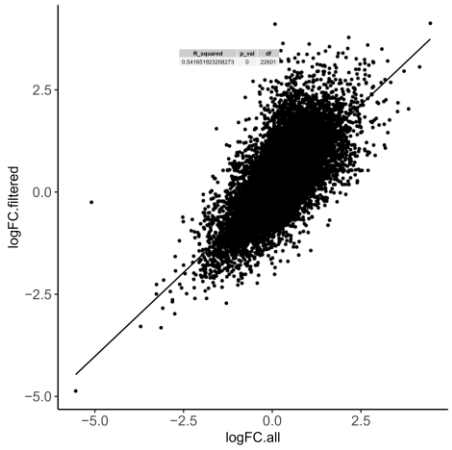


Supplementary Figure 11: DEGs in top clusters for males. a) Heatmaps showing the pseudobulk expression of differentially expressed genes in top clusters with highest number of DEGs in the male cluster level analysis – **(a)** ExN10_L46, **(b)** Ast1. The plotted values are pseudobulk CPMs calculated with edgeR and muscat and scaled per row (by gene). The annotation bar at the top is colored orange for cases and purple for controls, and rows and columns are not clustered. For

ExN10_L46 the DEGs are in ascending order of log fold change from top to bottom. Source data are provided as a Source Data file.



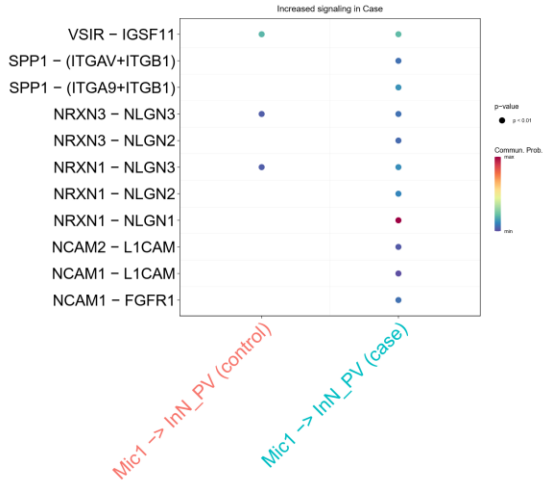
i



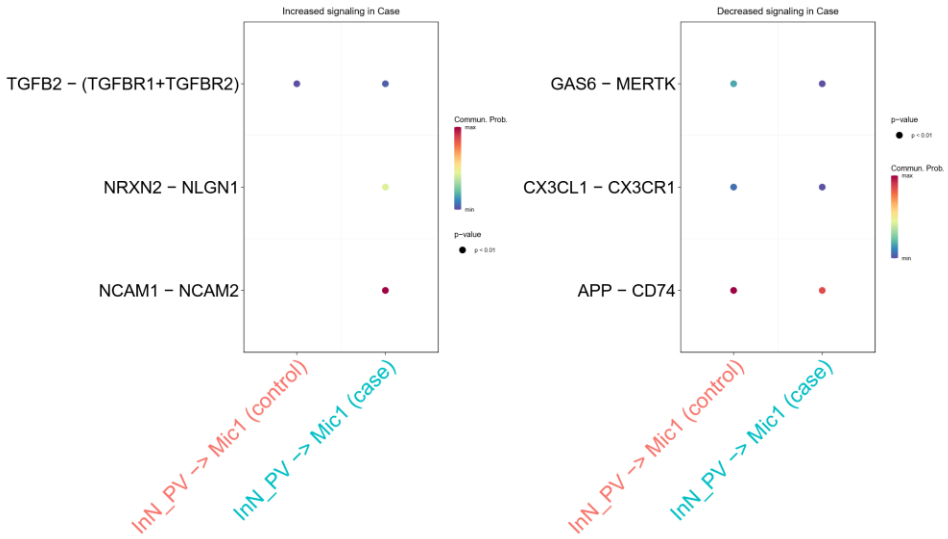
Supplementary Figure 12: Evaluation of female differential expression results. Number of DEGs in each broad cell type for females plotted against **(a)** the number of subjects included in DEG analysis and **(b)** the total number of nuclei in the broad cell type. For DEGs in females at the broad cell type level **(c)** the distribution of number of subjects flagged as possible outliers and **(d)** the distribution of subjects with non-zero expression. Number of DEGs in each cluster for females plotted against **(e)** the number of subjects included in DEG analysis and **(f)** the total number of nuclei in the cluster. For DEGs in females at the cluster level **(g)** the distribution of number of subjects flagged as possible outliers and **(h)** the distribution of subjects with non-zero expression. Possible outliers were not removed from the analyses but were assessed as a quality metric for DEG analysis. The plots include the Mix cluster results. **i)** Scatter plot of log fold changes per gene in DEG analysis of the female microglia data at the cluster level with or without including nuclei that express oligodendroglial markers and cluster close to oligodendroglia on the UMAP plot (methods: Differential expression analysis - Sub-clustering of microglia for differential expression analysis in females). The majority of DEGs obtained with the full microglia cluster were retained in the subsetted microglia cluster (40/68, 59%) and the log fold changes for DEGs were strongly

positively related (linear regression) with an R-squared of 0.54. Source data are provided as a Source Data file.

a



b



Supplementary Figure 13: Exploratory visualization of ligand-receptor interaction comparison in female microglia and PV interneurons. a) Ligand-receptor pairs with increased signaling from microglia to PV interneurons in cases compared to controls. **b)** Ligand receptor pairs with increased (left) and decreased (right) signaling from PV interneurons to microglia in cases compared to controls. These results are based on a preliminary assessment using CellChat. The

p-values, corresponding to communication between individual ligand-receptor pairs in cases and controls separately, are based on permutation and are uncorrected. Source data are provided as a Source Data file.

Supplementary References:

1. Maynard, K.R.*, et al.* Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nature neuroscience* 24, 425-436 (2021).

Chapter 4: Discussion

Preface to Chapter 4

In this chapter we summarize our findings from the two studies of cell type specific gene expression differences in depression, including the findings from the male and female cohort, and discuss them in the context of previous studies of brain cell type contributions to depression. Finally, we mention some of the limitations of the research presented in this thesis and indicate possible future directions of research which might build upon the results and address the gaps remaining in our knowledge. We end with a brief conclusion connecting our work back to the rationale and need for molecular studies of the brain in depression.

Discussion

Summary of the research

We have produced novel cell type specific gene expression and depression-associated gene expression datasets across cortical cell types from the human dlPFC. The datasets have been made publicly available and can be used by other researchers to generate further insights and hypotheses or to verify results obtained with independent cohorts. These data and results are contributing to the growing body of knowledge regarding cell type specific molecular profiling of the human body and brain. Our combined dataset compares favorably to the current publicly available disease-specific single-nucleus datasets of the human brain in terms of number of individuals included.

Our results provide a first glimpse at cell-type specific gene expression differences in males and females with depression compared to psychiatrically healthy controls and highlight the importance of deep layer excitatory neurons, astrocytes, oligodendrocyte precursor cells, and microglia. The results implicate a variety of cell types, both neuronal and glial, which is in line with findings from previous bulk tissue molecular and histological studies. Further, the findings that cell types with the most differentially expressed genes were distinct in males and females is in line with previous studies in humans and in animal models[18,23,26,32,44] indicating differences in the molecular pathophysiology of MDD between the sexes.

Additionally, our meta-analyses comparing the male and female datasets suggest that within individual cell types the transcriptomic patterns associated with depression are more similar than opposite in males and females. This highlights the strength of cell type specific approaches over

bulk tissue approaches. In previous bulk studies the discordant MDD-associated transcriptomic patterns between males and females[18,23] could have arisen due to the prominence of separate cell types in each sex.

Further, the key results of our studies are in many ways consistent with previous cell type specific examinations of molecular changes in depression, which generally focused on a few cell types at a time, as discussed below.

Comparison to previous findings

Excitatory and inhibitory neurons

The dysregulation of a subtype of deep layer excitatory neurons of the dlPFC in the male dataset was consistent between our first analysis and reanalysis of these data using quite distinct approaches, and thus is likely to reflect a robust feature of depression-associated transcriptomic differences in this cohort. Our finding of prominent differential expression in inhibitory interneurons in females is consistent with past literature demonstrating contributions of cortical interneuron subtypes to depression[41] although we found more changes in PV interneurons than in SST interneurons and the previous results were not necessarily specific to females. Together the above two results provide support for theories that suggest an imbalance in excitatory and inhibitory signaling in depression.

The threshold-free meta-analyses suggested concordance in changes between males and females in the deep layer excitatory neuron and PV interneuron clusters which had high numbers of DEGs in males and female respectively. Additionally, in the p-value combination meta-analyses, most of the DEGs in the male dataset for ExN10_L46 and in the female dataset for InN1_PV were recapitulated. We might speculate that the above cell types are involved in excitatory and

inhibitory disbalance in both sexes, but the contributions of deep layer excitatory neurons are greater in males and those of PV interneurons are greater in females.

Based on bioinformatic comparison, ExN10_L46 matched closely with layer 5 intratelencephalic neurons from the Allen Brain Institute motor cortex dataset. Although motor cortex cell types are not directly equivalent to dlPFC cell types, layer 5 intratelencephalic neurons typically send projections to other regions of the telencephalon – the forebrain, mainly the cortex and striatum, in contrast to extratelencephalic neurons which send signals outside the cortex and striatum, for example to the thalamus and pons[93]. Interestingly, intratelencephalic neurons in the PFC can respond to a variety of neuromodulators including dopamine, norepinephrine, and serotonin which are important in mood regulation[93].

PV interneurons synapse onto the soma and axons of excitatory neurons in the cortex and modulate their output, in contrast to SST interneurons which synapse onto the dendrites of cortical excitatory neurons and modulate the inputs to these neurons[94]. The dlPFC can contribute to the shift from executive network to DMN that has often been implicated in MDD and changes in excitation and inhibition in this brain region have been proposed to contribute to the altered connectivity in brain imaging studies[41]. We might speculate based on our results, that in males and females, different transcriptomic mechanisms – PV interneuron versus excitatory neuron gene expression differences – could be contributing to similar circuit level changes in the dlPFC resulting in MDD phenotypes.

The oligodendrocyte lineage

Our finding of prominent differential expression in OPCs was common to both analyses of the male dataset, however while in our first approach we found DEGs mainly in a potentially less

differentiated subcluster of OPCs, in the second approach we found DEGs in the broad OPC cluster. The implication of both OPCs and astrocytes is consistent with a previous study which found evidence for altered interaction between astrocytes and OPCs in depressed individuals[54]. Although we did not explore cell-cell interactions between astrocytes and OPCs, this could be an interesting avenue for future research both using bioinformatic and experimental approaches.

We did not find prominent differential expression signatures in oligodendrocytes either in males or in females and this in contrast with previous work which identified molecular differences implicating oligodendrocytes in depression[53,55]. However, the previous studies focused on a different brain region, the ACC, and the differences identified were specific to those with depression and a history of experiencing childhood abuse. On the other hand, our cohorts included individuals with a history of depression, not separated by the presence or absence of experienced childhood abuse in their history. Further, we are limited to examining gene expression whereas some of the findings in these studies were in DNA methylation differences[53] or differences in levels of lipid components of myelin[55].

Other glial cells

Astrocytes and microglia are two classes of glial cells that are highly responsive to the surrounding environment. These dynamic cells are thus involved in many disease pathologies in the brain. In our results, we found evidence for greater involvement of astrocytes in depression pathology in males and greater involvement of microglia in depression pathology for females. One possible explanation for this is that similar external cues or risk factors for depression (genetic susceptibility, chronic stress, early life adversity) may be triggering a response in a different glial cell type in each sex. While we cannot verify this hypothesis based on our dataset, it provides an

interesting avenue of future exploration of sex specific mechanisms of depression using animal models.

Our findings of altered gene expression in astrocytes in the reanalysis of the male cohort, including downregulation of connexin genes such as *GJA1* and other astrocytic genes such as *AQP4,* is consistent with previous studies[46,47]. This is not surprising since individuals were selected from the same brain bank and a subset of individuals were in common with these previous studies. While we had detected DEGs in astrocytes in the previous analysis of the male dataset, and some of the genes are also recapitulated in the reanalysis (such as *FADS2*), astrocytes were not as prominently dysregulated as deep layer excitatory neurons and OPCs in the previous assessment. This could reflect better ability to correctly identify astrocytes when combining the male and female datasets for analysis and better power for detecting differential expression at the pseudobulk level.

Our findings in microglia in females and the dysregulation of inflammation related pathways is consistent with previous literature[57-60] however there is no clear pattern of pro-inflammatory or anti-inflammatory. The female specificity of these findings is more difficult to verify against previous literature since not all studies performed sex-stratified analyses and even for studies that did, female cohorts were relatively small[59]. The downregulation of *IL1B* in depression is consistent with previous findings from the choroid plexus, although the previous study was not cell type specific[95]. Additionally, a few of our DEG findings are consistent with the Scheepstra et al. (2023) study which was specific to microglia but in a different cortical region, namely decrease in *CD14* and increase in *CNTNAP2*[61].

On the other hand, we did not detect changes in endothelial cells either in males or females,

which is not consistent with the previous literature in animal models and in human brains[62]. This could be a limitation of our snRNA-seq approach as several studies have shown that specific enrichment of endothelial cells needs to be performed at the wet lab level to fully profile this and associated cell types from the post-mortem brain[91].

Limitations

Despite the strengths of our research and its consistency with previous literature, there are also weaknesses that need to be considered. There are many reasons to be cautious when interpreting results from high-throughput transcriptomic studies of disease relying on postmortem brain tissue. From the potential for lack of specificity of disease associated gene expression signatures in high-throughput transcriptomic studies[96] to the unavoidable biases when using postmortem tissue for molecular research[97], there are pitfalls that deserve consideration. However, limited though we may be by available tools, techniques, and resources we can and should continue to strive towards acquiring more information, knowledge, and understanding of brain disorders at a molecular level. This work represents one such attempt to make sense of the brain in depression. Our results represent the outcome of a specific set of decisions regarding analysis and the details of the results would inevitably change to some extent if the methods applied were changed. However, for our male cohort we saw that the prominence of gene expression changes in a subtype of deep layer excitatory neurons, and generally in oligodendrocyte precursor cells, was preserved even after a very different analytical approach was implemented. This increases our confidence that the broad conclusions derived from our data and analysis should be robust to analytical approaches.

We are limited by the generation of data in males and females in separate studies, which did not

allow a direct comparison between the sexes, or an assessment of sex by phenotype interaction, and future work must address this by simultaneously generating and analyzing cell type specific gene expression profiles associated with depression in both sexes. Moreover, the cohort of individuals used in this study was mostly comprised of people of European ancestry. While the genetic background of individuals is more important in genetic studies than in transcriptomic studies, we cannot rule out the possibility that findings in more diverse cohorts will differ.

When interpreting these data, we must keep in mind that these results pertain to a single region of the cortex, albeit a region which is relevant to depression pathology. Thus, future work should test the extensibility of our findings to other cortical and sub-cortical brain regions relevant to depression. In fact, in the future, combining brain imaging network analysis in living patients with postmortem snRNA-seq across brain regions may begin to tease apart the similarities and differences of depression associated changes in anatomically and functionally connected brain regions with cell type resolution.

We did not account for the presence or absence of a history of childhood abuse in our analysis, which previous studies have demonstrated can in some cases produce a more specific molecular phenotype compared to depression without a history of experienced childhood abuse. Additionally, our depressed individuals all died by suicide, and we cannot distinguish specific effects of suicide from the effects of depression in our results.

A limitation of our work is that, while we attempted technical validation of a few of our results in the initial male analysis with partial success, the process is technically challenging and very resource intensive, thus we did not perform equivalent experiments in the female cohort.

Lastly, we have not attempted to replicate our findings in an independent cohort from a different

brain bank. Thus, we do not know if these results are generalizable to other cohorts. It will be important for future work to replicate our major findings in an independent cohort, which is a non-trivial endeavor when working with postmortem brain samples pertaining to a specific disease state which are challenging to collect. Moreover, any molecular studies of the postmortem brain cannot draw causal conclusions. Follow-up studies in animal models of stress can be helpful in this regard.

Future directions

The comparison of cell type specific gene expression differences between cases and controls is a first step in generating insights from the data generated. Many other assessments of the data are possible. For example, construction of cell type specific gene regulatory networks (GRN) in cases and controls, followed by identification of GRNs associated with depression, by applying methods such as SCENIC[98] which identifies co-regulated "regulons" of transcription factors and their target genes, could be informative for expanding our molecular understanding of depression.

Further, it would be interesting to explore if the depression-associated gene expression modules identified in female microglia and PV interneurons via WGCNA are preserved in males in these cell types. Equivalent comparisons could be performed for prominently affected cell types in males – astrocytes and deep layer excitatory neurons.

Moreover, cell type specific prioritized lists of genes, such as DEGs in top cell types or cell type specific depression-associated WGCNA modules hub genes, can be used to augment or be augmented by genetic findings in MDD. For example, it may be possible to construct cell type specific expression-based polygenic risk scores (ePRS)[99] or transcriptome-based polygenic score (T-PRS)[15] by leveraging cell type specific expression quantitative trait loci (eQTLs) that may

distinguish individuals with genetic risk for depression from those without. In fact, the gene expression profiles of the cell types in our datasets could be used to perform cell type specific "deconvolution" of eQTLs, leveraging the statistical power of larger bulk datasets[100]. Further, it may be possible to prioritize GWAS variants by assessing their potential to affect cell type specific MDD-associated genes.

Some cell type specific gene expression studies have sought to link gene expression within given cell types to specific symptomatology of the disease under investigation[101,102]. We did not perform such analyses with our dataset, but the possibilities for linking molecular differences in specific cell types of the brain to different dimensions of depression are intriguing and should be explored in future work if sufficient characterization of the individuals is available.

While we measured only one molecular modality of cell type specific information, gene expression, many of the more recent single-cell resolution studies of the brain are leaning towards incorporating multiple modalities such as gene expression and chromatin accessibility within the same cells. Future studies can extend our work by profiling the cell type specific epigenetic differences associated with depression, including chromatin accessibility, histone modifications, and DNA-methylation, which no doubt function in concert with gene expression differences to produce depression pathology. Using other modalities to corroborate our results will reveal the points of convergence and divergence of these inter-related molecular mechanisms in determining depression pathology.

Conclusion

From the perspective of public health, the problem of depression is complex and difficult to solve, even while it is very important to address, and deserves our attention and dedicated efforts

towards a solution. The work presented here, although examining a disease state in the postmortem human brain, represents basic science rather than translational research. We attempted to better understand the human brain at a molecular level, and we attempted to better understand how the molecular state of the brain differs in depression. Our results both concurred with and built upon previous findings in molecular and cellular brain pathology in MDD and suggest several new avenues for exploring the sex and cell type specific mechanisms that contribute to the development of depression.

References

1. American Psychiatric, A. *Diagnostic and statistical manual of mental disorders : DSM-5*, (American Psychiatric Association, Arlington, VA
Washington, D.C., 2013).
2. Mental Disorders Collaborators, G. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet Psychiatry* **9**, 137-150 (2022).
3. Bachmann, S. Epidemiology of Suicide and the Psychiatric Perspective. *International Journal of Environmental Research and Public Health* **15**, 1425 (2018).
4. Howard, D.M.*, et al.* Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat Neurosci* **22**, 343-352 (2019).
5. Levey, D.F.*, et al.* Bi-ancestral depression GWAS in the Million Veteran Program and meta-analysis in >1.2 million individuals highlight new therapeutic directions. *Nat Neurosci* **24**, 954-963 (2021).
6. Mena, F. & Benoit, L. Molecular programs underlying differences in the expression of mood disorders in males and females. *Brain Res* **1719**, 89-103 (2019).
7. Malhi, G.S. & Mann, J.J. Depression. *Lancet* **392**, 2299-2312 (2018).
8. Godoy, L.D., Rossignoli, M.T., Delfino-Pereira, P., Garcia-Cairasco, N. & de Lima Umeoka, E.H. A Comprehensive Overview on Stress Neurobiology: Basic Concepts and Clinical Implications. *Frontiers in Behavioral Neuroscience* **12**(2018).
9. Breit, S., Kupferberg, A., Rogler, G. & Hasler, G. Vagus Nerve as Modulator of the Brain–Gut Axis in Psychiatric and Inflammatory Disorders. *Frontiers in Psychiatry* **9**(2018).
10. Ibrahim, P., Almeida, D., Nagy, C. & Turecki, G. Molecular impacts of childhood abuse on the human brain. *Neurobiol Stress* **15**, 100343 (2021).
11. Kessler, R.C. & Bromet, E.J. The epidemiology of depression across cultures. *Annu Rev Public Health* **34**, 119-138 (2013).
12. Schuch, J.J., Roest, A.M., Nolen, W.A., Penninx, B.W. & de Jonge, P. Gender differences in major depressive disorder: results from the Netherlands study of depression and anxiety. *J Affect Disord* **156**, 156-163 (2014).
13. Silveira, P.P., Pokhvisneva, I., Howard, D.M. & Meaney, M.J. A sex-specific genome-wide association study of depression phenotypes in UK Biobank. *Molecular Psychiatry* (2023).
14. Talishinsky, A.*, et al.* Regional gene expression signatures are associated with sex-specific functional connectivity changes in depression. *Nat Commun* **13**, 5692 (2022).
15. Miles, A.E.*, et al.* Transcriptome-based polygenic score links depression-related corticolimbic gene expression changes to sex-specific brain morphology and depression risk. *Neuropsychopharmacology* **46**, 2304-2311 (2021).
16. Labaka, A., Goñi-Balentziaga, O., Lebeña, A. & Pérez-Tejada, J. Biological Sex Differences in Depression: A Systematic Review. *Biol Res Nurs* **20**, 383-392 (2018).
17. Rainville, J.R., Lipuma, T. & Hodes, G.E. Translating the Transcriptome: Sex Differences in the Mechanisms of Depression and Stress, Revisited. *Biol Psychiatry* **91**, 25-35 (2022).
18. Labonté, B.*, et al.* Sex-specific transcriptional signatures in human depression. *Nat Med*

23, 1102-1111 (2017).
19. Issler, O*., et al.* Sex-Specific Role for the Long Non-coding RNA LINC00473 in Depression. *Neuron* **106**, 912-926.e915 (2020).
20. Issler, O*., et al.* The long noncoding RNA FEDORA is a cell type– and sex-specific regulator of depression. *Science Advances* **8**, eabn9494 (2022).
21. Girgenti, M.J*., et al.* Transcriptomic organization of the human brain in post-traumatic stress disorder. *Nat Neurosci* **24**, 24-33 (2021).
22. Ding, Y*., et al.* Molecular and Genetic Characterization of Depression: Overlap with other Psychiatric Disorders and Aging. *Mol Neuropsychiatry* **1**, 1-12 (2015).
23. Seney, M.L*., et al.* Opposite Molecular Signatures of Depression in Men and Women. *Biol Psychiatry* **84**, 18-27 (2018).
24. Wu, W., Howard, D., Sibille, E. & French, L. Differential and spatial expression meta-analysis of genes identified in genome-wide association studies of depression. *Transl Psychiatry* **11**, 8 (2021).
25. Seney, M.L., Glausier, J. & Sibille, E. Large-Scale Transcriptomics Studies Provide Insight Into Sex Differences in Depression. *Biol Psychiatry* **91**, 14-24 (2022).
26. Bollinger, J.L. Uncovering microglial pathways driving sex-specific neurobiological effects in stress and depression. *Brain Behav Immun Health* **16**, 100320 (2021).
27. Touchant, M. & Labonté, B. Sex-Specific Brain Transcriptional Signatures in Human MDD and Their Correlates in Mouse Models of Depression. *Front Behav Neurosci* **16**, 845491 (2022).
28. Caradonna, S.G*., et al.* Corticosterone induces discrete epigenetic signatures in the dorsal and ventral hippocampus that depend upon sex and genotype: focus on methylated Nr3c1 gene. *Translational Psychiatry* **12**, 109 (2022).
29. Deonaraine, K.K*., et al.* Sex-specific peripheral and central responses to stress-induced depression and treatment in a mouse model. *J Neurosci Res* **98**, 2541-2553 (2020).
30. Olave, F.A*., et al.* Chronic restraint stress produces sex-specific behavioral and molecular outcomes in the dorsal and ventral rat hippocampus. *Neurobiol Stress* **17**, 100440 (2022).
31. Bollinger, J.L., Bergeon Burns, C.M. & Wellman, C.L. Differential effects of stress on microglial cell activation in male and female medial prefrontal cortex. *Brain Behav Immun* **52**, 88-97 (2016).
32. Bollinger, J.L., Collins, K.E., Patel, R. & Wellman, C.L. Behavioral stress alters corticolimbic microglia in a sex- and brain region-specific manner. *PLoS One* **12**, e0187631 (2017).
33. Woodburn, S.C., Bollinger, J.L. & Wohleb, E.S. Synaptic and behavioral effects of chronic stress are linked to dynamic and sex-specific changes in microglia function and astrocyte dystrophy. *Neurobiol Stress* **14**, 100312 (2021).
34. Paden, W*., et al.* Sex differences in adult mood and in stress-induced transcriptional coherence across mesocorticolimbic circuitry. *Translational Psychiatry* **10**, 59 (2020).
35. Stockmeier, C.A. & Rajkowska, G. Cellular abnormalities in depression: evidence from postmortem brain tissue. *Dialogues in Clinical Neuroscience* **6**, 185-197 (2004).
36. Kang, H.J*., et al.* Decreased expression of synapse-related genes and loss of synapses in major depressive disorder. *Nature medicine* **18**, 1413-1417 (2012).
37. Chang, L.-C*., et al.* A conserved BDNF, glutamate- and GABA-enriched gene module related

to human depression identified by coexpression meta-analysis and DNA variant genome-wide association studies. *PloS one* **9**, e90980 (2014).
38. Cruceanu, C*., et al.* H3K4 tri-methylation in synapsin genes leads to different expression patterns in bipolar disorder and major depression. *International Journal of Neuropsychopharmacology* **16**, 289-299 (2012).
39. Cruceanu, C*., et al.* DNA hypomethylation of Synapsin II CpG islands associates with increased gene expression in bipolar disorder and major depression. *BMC Psychiatry* **16**, 286 (2016).
40. Duman, R.S., Sanacora, G. & Krystal, J.H. Altered Connectivity in Depression: GABA and Glutamate Neurotransmitter Deficits and Reversal by Novel Treatments. *Neuron* **102**, 75-90 (2019).
41. Northoff, G. & Sibille, E. Why are cortical GABA neurons relevant to internal focus in depression? A cross-level model linking cellular, biochemical and neural network findings. *Molecular psychiatry* **19**, 966-977 (2014).
42. Gray, A.L., Hyde, T.M., Deep-Soboslay, A., Kleinman, J.E. & Sodhi, M.S. Sex differences in glutamate receptor gene expression in major depression and suicide. *Molecular psychiatry* **20**, 1057-1068 (2015).
43. Goswami, D.B*., et al.* Gene expression analysis of novel genes in the prefrontal cortex of major depressive disorder subjects. *Progress in Neuropsychopharmacology & Biological Psychiatry* **43**, 126-133 (2013).
44. Girgenti, M.J*., et al.* Prefrontal cortex interneurons display dynamic sex-specific stress-induced transcriptomes. *Translational Psychiatry* **9**, 292 (2019).
45. Tanti, A*., et al.* Child abuse associates with increased recruitment of perineuronal nets in the ventromedial prefrontal cortex: a possible implication of oligodendrocyte progenitor cells. *Molecular Psychiatry* **27**, 1552-1561 (2022).
46. Ernst, C*., et al.* Dysfunction of Astrocyte Connexins 30 and 43 in Dorsal Lateral Prefrontal Cortex of Suicide Completers. *Biological Psychiatry* **70**, 312-319 (2011).
47. Nagy, C*., et al.* Astrocytic abnormalities and global DNA methylation patterns in depression and suicide. *Molecular Psychiatry* **20**, 320-328 (2015).
48. Nagy, C., Torres-Platas, S.G., Mechawar, N. & Turecki, G. Repression of Astrocytic Connexins in Cortical and Subcortical Brain Regions and Prefrontal Enrichment of H3K9me3 in Depression and Suicide. *International Journal of Neuropsychopharmacology* **20**, 50-57 (2016).
49. Torres-Platas, S.G., Nagy, C., Wakid, M., Turecki, G. & Mechawar, N. Glial fibrillary acidic protein is differentially expressed across cortical and subcortical regions in healthy brains and downregulated in the thalamus and caudate nucleus of depressed suicides. *Molecular Psychiatry* **21**, 509-515 (2016).
50. Torres-Platas, S.G*., et al.* Astrocytic Hypertrophy in Anterior Cingulate White Matter of Depressed Suicides. *Neuropsychopharmacology* **36**, 2650-2658 (2011).
51. O'Leary, L.A*., et al.* Widespread Decrease of Cerebral Vimentin-Immunoreactive Astrocytes in Depressed Suicides. *Frontiers in Psychiatry* **12**(2021).
52. Aston, C., Jiang, L. & Sokolov, B.P. Transcriptional profiling reveals evidence for signaling and oligodendroglial abnormalities in the temporal cortex from patients with major

depressive disorder. *Molecular Psychiatry* **10**, 309-322 (2004).
53. Pierre-Eric Lutz, M.D., Ph.D. *„ et al.* Association of a History of Child Abuse With Impaired Myelination in the Anterior Cingulate Cortex: Convergent Epigenetic, Transcriptional, and Morphological Evidence. *American Journal of Psychiatry* **174**, 1185-1194 (2017).
54. Tanti, A.*, et al.* Evidence of decreased gap junction coupling between astrocytes and oligodendrocytes in the anterior cingulate cortex of depressed suicides. *Neuropsychopharmacology* **44**, 2099-2111 (2019).
55. Perlman, K.*, et al.* Fatty acid dysregulation in the anterior cingulate cortex of depressed suicides with a history of child abuse. *Translational Psychiatry* **11**, 535 (2021).
56. Barnes, J., Mondelli, V. & Pariante, C.M. Genetic Contributions of Inflammation to Depression. *Neuropsychopharmacology* **42**, 81-98 (2017).
57. Morrison, F.G.*, et al.* Reduced interleukin 1A gene expression in the dorsolateral prefrontal cortex of individuals with PTSD and depression. *Neuroscience Letters* **692**, 204-209 (2019).
58. Qingzhong Wang, Ph.D. ,, Bhaskar Roy, Ph.D. ,, Gustavo Turecki, M.D., Ph.D. ,, Richard C. Shelton, M.D. , & Yogesh Dwivedi, Ph.D. Role of Complex Epigenetic Switching in Tumor Necrosis Factor-α Upregulation in the Prefrontal Cortex of Suicide Subjects. *American Journal of Psychiatry* **175**, 262-274 (2018).
59. Mahajan, G.J.*, et al.* Altered neuro-inflammatory gene expression in hippocampus in major depressive disorder. *Progress in Neuro-Psychopharmacology and Biological Psychiatry* **82**, 177-186 (2018).
60. Torres-Platas, S.G., Cruceanu, C., Chen, G.G., Turecki, G. & Mechawar, N. Evidence for increased microglial priming and macrophage recruitment in the dorsal anterior cingulate white matter of depressed suicides. *Brain, Behavior, and Immunity* **42**, 50-59 (2014).
61. Scheepstra, K.W.F.*, et al.* Microglia transcriptional profiling in major depressive disorder shows inhibition of cortical grey matter microglia. *medRxiv*, 2023.2001.2011.23284393 (2023).
62. Dion-Albert, L.*, et al.* Vascular and blood-brain barrier-related changes underlie stress responses and resilience in female mice and depression in human tissue. *Nature Communications* **13**, 164 (2022).
63. Velmeshev, D.*, et al.* Single-cell genomics identifies cell type–specific molecular changes in autism. *Science* **364**, 685 (2019).
64. Mathys, H.*, et al.* Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332-337 (2019).
65. Jäkel, S.*, et al.* Altered human oligodendrocyte heterogeneity in multiple sclerosis. *Nature* **566**, 543-547 (2019).
66. Bartosovic, M., Kabbe, M. & Castelo-Branco, G. Single-cell CUT&Tag profiles histone modifications and transcription factors in complex tissues. *Nat Biotechnol* **39**, 825-835 (2021).
67. Zhang, D.*, et al.* Spatial epigenome–transcriptome co-profiling of mammalian tissues. *Nature* **616**, 113-122 (2023).
68. Lun, A.T.L.*, et al.* EmptyDrops: distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. *Genome Biology* **20**, 63 (2019).
69. Neavin, D.*, et al.* Demuxafy: Improvement in droplet assignment by

integrating multiple single-cell demultiplexing and doublet detection methods. *bioRxiv*, 2022.2003.2007.483367 (2022).
70. Xi, N.M. & Li, J.J. Benchmarking Computational Doublet-Detection Methods for Single-Cell RNA Sequencing Data. *Cell Syst* **12**, 176-194.e176 (2021).
71. Luecken, M.D*., et al.* Benchmarking atlas-level data integration in single-cell genomics. *Nature Methods* **19**, 41-50 (2022).
72. Tang, M*., et al.* Evaluating single-cell cluster stability using the Jaccard similarity index. *Bioinformatics* **37**, 2212-2214 (2021).
73. Patterson-Cross, R.B., Levine, A.J. & Menon, V. Selecting single cell clustering parameter values using subsampling-based robustness metrics. *BMC Bioinformatics* **22**, 39 (2021).
74. Crowell, H.L*., et al.* muscat detects subpopulation-specific state transitions from multi-sample multi-condition single-cell transcriptomics data. *Nature Communications* **11**, 6077 (2020).
75. Squair, J.W*., et al.* Confronting false discoveries in single-cell differential expression. *Nature Communications* **12**, 5692 (2021).
76. Zhang, M.J*., et al.* Polygenic enrichment distinguishes disease associations of individual cells in single-cell RNA-seq data. *Nature Genetics* **54**, 1572-1580 (2022).
77. Liu, B., Li, Y. & Zhang, L. Analysis and Visualization of Spatial Transcriptomic Data. *Frontiers in Genetics* **12**(2022).
78. Baek, S. & Lee, I. Single-cell ATAC sequencing analysis: From data preprocessing to hypothesis generation. *Computational and Structural Biotechnology Journal* **18**, 1429-1439 (2020).
79. Dimitrov, D*., et al.* Comparison of methods and resources for cell-cell communication inference from single-cell RNA-Seq data. *Nat Commun* **13**, 3224 (2022).
80. Almet, A.A., Cang, Z., Jin, S. & Nie, Q. The landscape of cell-cell communication through single-cell transcriptomics. *Curr Opin Syst Biol* **26**, 12-23 (2021).
81. Morabito, S*., et al.* Single-nucleus chromatin accessibility and transcriptomic characterization of Alzheimer's disease. *Nature Genetics* **53**, 1143-1155 (2021).
82. Lau, S.-F., Cao, H., Fu, A.K.Y. & Ip, N.Y. Single-nucleus transcriptome analysis reveals dysregulation of angiogenic endothelial cells and neuroprotective glia in Alzheimer's disease. *Proceedings of the National Academy of Sciences* **117**, 25800-25809 (2020).
83. Lee, H*., et al.* Cell Type-Specific Transcriptomics Reveals that Mutant Huntingtin Leads to Mitochondrial RNA Release and Neuronal Innate Immune Activation. *Neuron* **107**, 891-908.e898 (2020).
84. Smajić, S*., et al.* Single-cell sequencing of human midbrain reveals glial activation and a Parkinson-specific neuronal state. *Brain* **145**, 964-978 (2022).
85. Ruzicka, W.B*., et al.* Single-cell multi-cohort dissection of the schizophrenia transcriptome. *medRxiv*, 2022.2008.2031.22279406 (2022).
86. Reiner, B.C*., et al.* Single-nuclei transcriptomics of schizophrenia prefrontal cortex primarily implicates neuronal subtypes. *bioRxiv*, 2020.2007.2029.227355 (2021).
87. Batiuk, M.Y*., et al.* Upper cortical layer–driven network impairment in schizophrenia. *Science Advances* **8**, eabn8367 (2022).
88. Tran, M.N*., et al.* Single-nucleus transcriptome analysis reveals cell-type-specific

molecular signatures across reward circuitry in the human brain. *Neuron* **109**, 3088-3103.e3085 (2021).
89. Herring, C.A.*, et al.* Human prefrontal cortex gene regulatory dynamics from gestation to adulthood at single-cell resolution. *Cell* **185**, 4428-4447.e4428 (2022).
90. Geirsdottir, L.*, et al.* Cross-Species Single-Cell Analysis Reveals Divergence of the Primate Microglia Program. *Cell* **179**, 1609-1622.e1616 (2019).
91. Garcia, F.J.*, et al.* Single-cell dissection of the human brain vasculature. *Nature* **603**, 893-899 (2022).
92. Zappia, L., Phipson, B. & Oshlack, A. Exploring the single-cell RNA-seq analysis landscape with the scRNA-tools database. *PLOS Computational Biology* **14**, e1006245 (2018).
93. Moberg, S. & Takahashi, N. Neocortical layer 5 subclasses: From cellular properties to roles in behavior. *Frontiers in Synaptic Neuroscience* **14**(2022).
94. Hu, H. & Vervaeke, K. Synaptic Integration in Cortical Inhibitory Neuron Dendrites. *Neuroscience* **368**, 115-131 (2018).
95. Devorak, J.*, et al.* Cellular and Molecular Inflammatory Profile of the Choroid Plexus in Depression and Suicide. *Front Psychiatry* **6**, 138 (2015).
96. Crow, M., Lim, N., Ballouz, S., Pavlidis, P. & Gillis, J. Predictability of human differential gene expression. *Proc Natl Acad Sci U S A* **116**, 6491-6500 (2019).
97. Liharska, L.E.*, et al.* A study of gene expression in the living human brain. *medRxiv*, 2023.2004.2021.23288916 (2023).
98. Aibar, S.*, et al.* SCENIC: single-cell regulatory network inference and clustering. *Nature Methods* **14**, 1083-1086 (2017).
99. Restrepo-Lozano, J.M.*, et al.* Corticolimbic DCC gene co-expression networks as predictors of impulsivity in children. *Mol Psychiatry* **27**, 2742-2750 (2022).
100. Park, Y.*, et al.* Single-cell deconvolution of 3,000 post-mortem brain samples for eQTL and GWAS dissection in mental disorders. *bioRxiv*, 2021.2001.2021.426000 (2021).
101. Mathys, H.*, et al.* Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332-337 (2019).
102. Velmeshev, D.*, et al.* Single-cell genomics identifies cell type-specific molecular changes in autism. *Science* **364**, 685-689 (2019).

Appendices

Appendix 1: Extraction of nuclei from archived post-mortem tissues for single-nucleus sequencing applications

Malosree Maitra[1,2], Corina Nagy[1,2], Anjali Chawla[2], Yu Chang Wang[4], Camila Nascimento[5], Mathew Suderman[6,7], Jean-Francois Théroux[2], Naguib Mechawar[2,3], Jiannis Ragoussis[4,8], Gustavo Turecki[2,3,4]*

[1]These authors contributed equally

[2]McGill Group for Suicide Studies, Douglas Mental Health University Institute, [3]Department of Psychiatry, McGill University, [4]McGill University and Genome Quebec Innovation Centre, Department of Human Genetics, McGill University, [5]Department of Psychiatry, University of Sao Paolo, [6]MRC Integrative Epidemiology Unit, University of Bristol, UK, [7]Population Health Sciences, Bristol Medical School, Bristol, UK, [8]Department of Bioengineering, McGill University

***Correspondence to**: Dr. Gustavo Turecki (gustavo.turecki@mcgill.ca)
6875 LaSalle Blvd, Verdun, Québec, Canada H4H 1R3
Tel.: (514) 761-6131 (ext. 2370)

Related links
Key references using this protocol
Jessa, S. et al. Nat Genet 51, 1702–1713 (2019): https://doi.org/10.1038/s41588-019-0531-7
Reiner, B. C., et al. bioRxiv, 2020.2007.2029.227355 (2020): https://doi.org/10.1101/2020.07.29.227355
Nagy, C. et al. Nat Neurosci 23, 771–781 (2020): https://doi.org/10.1038/s41593-020-0621-y

Abstract

Single-cell and single-nucleus sequencing techniques are a burgeoning field with various biological, biomedical, and clinical applications. Numerous high and low-throughput methods have been developed for sequencing the RNA and DNA content of single cells. However, for all these methods the key requirement is high quality input of a single-cell or single-nucleus suspension. Preparing such a suspension is the limiting step when working with fragile, archived tissues of variable quality. This hurdle can prevent such tissues from being extensively investigated with single-cell technologies. We describe a protocol for preparing single-nucleus suspensions within the span of a few hours that reliably works for multiple post-mortem and archived tissue types using standard lab equipment. The stages of the protocol include tissue preparation and dissociation, nuclei extraction, and nuclei concentration assessment and capture. The protocol is comparable to other published protocols but does not require fluorescence assisted nuclei sorting or ultracentrifugation. The protocol can be carried out by a competent graduate student familiar with basic laboratory techniques and equipment. Moreover, these preparations are compatible with single-nucleus RNA-seq and ATAC-seq using the 10X Genomics' Chromium system. The protocol reliably results in efficient capture of single nuclei for high-quality single-nucleus RNA- seq libraries.

Keywords: nuclei extraction, single nucleus suspension, snRNA-seq, snATAC-seq

Introduction

Advancements in technology have allowed researchers to preform large-scale transcriptomic studies at the level of a single-cell. Droplet-based cell isolation has become a favorite in the field for its scalability and simplicity of use with either in-house fluidic set-ups[1,2], or commercially available equipment (10x Genomics)[3]. This technique is particularly interesting for tissues with highly heterogeneous cellular compositions like intestine[4], lung[5], spinal cord[6], and brain[7,8]. There has been particular interest in deconvoluting brain architecture and function, which at its base, starts by accurately identifying all the cells types present[9-11]. However, truly harnessing the power of individual cellular transcriptomes comes with assessing differences between those transcriptomes in different physiological states. This is of particular value for complex diseases where multiple genes contribute with additive effects, making it difficult to identify changes in tissues homogenates[12]. The chemical dissociation of tightly interconnected brain cells and other cell-types has been found to alter transcription profiles[13,14]. Given that nuclear transcriptomes closely reflect the cell's cytosolic profile[15,16], isolating the nuclei from brain tissue has proven to be an excellent strategy for single-cell level studies. Likewise, other tissues that have either been frozen for long term storage or that are formed by syncytium, such as in skeletal muscles[17,18], could benefit from this approach.

Development of the Protocol

Numerous protocols for isolating nuclei from brain cells have been published[10,11,19-23]; some rely on the additional purification by fluorescence assisted cell sorting (FACS)[20,23], which is costly, time consuming and not readily available for all researchers, while others have made adjustments to the microfluidics component used to isolate and capture single cells or nuclei[10],

which can also be limiting to labs. Our protocol has been developed for use with the commercially available Chromium™ Single Cell Controller. This is a highly optimized system that allows scalable single-cell capture. We have adapted our protocol to allow the Chromium system to efficiently capture nuclei from archived post-mortem tissue. Our preparation produces stable and easily quantifiable nuclear suspension even when using archived brain tissue. We have used this approach to successfully compare the gene expression differences in the post-mortem prefrontal cortex of depressed patients who died by suicide compared to psychiatrically healthy controls[24]. The protocol has also been successfully applied, with minor modifications, to collect single-nucleus transcriptomic data from surgical samples of glioblastoma[25] and in a recent study of post-mortem brain in schizoprenia[26]. Thus, gene expression and chromatin accessibility can be measured from post-mortem brain tissue using this protocol. We anticipate that newer techniques from 10x Genomics such as those that combine scATAC-seq and snRNA-seq to study them simultaneously in frozen tissues could also take advantage of this protocol.

Overview of the procedure

The experimental workflow (**Fig. 1**) begins with cellular lysis by dounce-homogenization in low concentration detergent . Integral to the protocol are numerous wash steps to reduce ambient nucleic acid contamination, in a buffer containing a high percentage of bovine serum albumin to prevent nuclei aggregation. The suspension is repeatedly filtered to remove large debris. Most centrifugation steps are performed at low speeds to prevent damage to the nuclei. Finally, an iodixanol cushion is used to purify the nuclei. The nuclei numbers and concentration are assessed by a cell counter or hemocytometer. In addition, Hoechst or DAPI can be used to stain DNA for assessing the nuclei concentration by fluorescence microscopy. The concentration

of the nuclei suspension is important to reduce aggregation, particularly in tissues that have undergone long-term storage and are thus more likely to be damaged, fragile and inclined to aggregate. Generally, a concentration of 500-1000 nuclei per microliter is sufficient for nuclei capture and should not result in excessive aggregation.

Comparison with other methods

As previously mentioned, existing single-nucleus RNA sequencing (snRNA-seq) protocols either rely on FACS[20,23] which is harder to scale, or on droplet-based approaches which use in house set-ups[10]. Early protocols used relatively fresh frozen tissue, which is not available in most tissue banks where samples are likely to have undergone long-term storage[9-11]. Moreover, when studying specific phenotypes for which it is harder to obtain tissues, it is not always possible to select for short post-mortem intervals (PMIs) and archival times. Early protocols were also limited to high-quality tissue which may not be an option for answering certain types of research questions.

As with several more recently published potocols[19-22], we have been able to adapt our nuclear prep to be compatible with the 10X Chromium system which is becoming increasingly available as a service platform. Furthermore, the wet-lab aspect of the protocol will produce nuclei suitable for multiple post-nuclei capture applications such as whole genome sequencing for the study of somatic mutations or single-nucleus Assay for Transposase-Accessible Chromatin (snATAC-seq) as supported by preliminary results from our lab.

Each of the more recently developed protocols have their own strengths and weaknesses and, in some cases, adaptations for specific tissue types such as macro-dissections for white matter regions[21]. The strength of our protocol is that it is mostly unaffected by variations in post

mortem interval (PMI) or archival times of the samples **(Fig 2)**. Some of these protocols also incorporate ultra-centrifugation[22,27], which is time consuming and requires specialized equipment and could be damaging to fragile nuclei. We are able to circumvent the additional challenges that arise with archived tissue such as the fragility of the cells and organelles upon freeze-thawing which typically results in large amounts of debris and ambient RNA than can either interfere with droplet formation or be integrated into droplets, increasing background sequencing noise. Here, we show that brain tissue which has been stored at -80^{o}C for as long as 22 years, can produce high quality single-nuclei suspensions.

Directly applying either the cell preparation protocol or the demonstration protocol for nuclei developed from 10x Genomics did not produce useable results in our hands with archived post-mortem brain tissue (**Fig. 3**), although other labs have been able to successfully use this protocol for nuclei extraction for snRNA-seq. Moreover, our attempts to use nuclei isolated by fluorescent assisted nuclei sorting (FANS) as input to the 10X Genomics protocol did not yield acceptable results, although this approach has been adopted successfully by other groups[20]. The modifications made here are primarily for use with post-mortem brain tissue that has been archived for long periods of time, but can also be applied to any frozen post-mortem sample. Similar to previous studies[10,11], we applied a few modifications to the standard bioinformatic analysis with the CellRanger pipeline from 10X Gennomics to address a number of issues which arise with droplet-based single-nucleus sequencing. First, we assembled a pre-mRNA reference to account for unprocessed transcripts found in the nucleus[28]. Second, given that previous studies have consistently shown fewer identifiable transcripts in glial cells[10,11] we performed customized barcode filtering to include cells with a wider range of unique molecular identifiers (UMIs) while

removing noise. With these minor modifications to the analysis[24] our isolation approach for our tissue type, i.e. archived post-mortem brain, produced much improved data compared to the available 10X Genomics nuclei preparation protocols in our hands.

Experimental design

The most important factor to take into consideration while designing single-cell or single-nucleus RNA-seq experiments is the potential batch effects. Given that the Chromium system only allows for the capture of 8 samples at a time and that for many experiments that total number of samples to be analyzed may be greater than eight, it may be preferable to create a balanced experimental design if possible. This will help limit the effects of batch to batch variability. For example, if two phenotypic or treatment groups are to be compared, it would be ideal to include equal numbers of samples from each group in every batch. Moreover, other potential co-variates to take into consideration include age, PMI, and sex. It may be possible to account for the effects of these variables by matching samples by these parameters within each batch.

In cases where cell-type specific gene-expression data has been previously published, or single-cell or nucleus gene expression datasets are available, these data can be used for comparison to help determine whether the cell-types identified and single-nucleus transcriptomic profiles detected are comparable to previously published literature. In cases where such datasets are not available it may be informative to prepare bulk-tissue samples in parallel or to perform sequencing of fluorescence assisted nuclei sorting (FANS) purified populations of expected cell-types based on known genetic markers to validate the cell-type identification from the single-nucleus transcriptomic data[29]. High-throughput *in situ* hybridization (ISH)[30] and ISH based nuclei sorting[29] have also been used to confirm experimentally determined cell-types from snRNA-seq.

In the case of complex tissues, it can be useful to perform careful dissection and even to cryosection the tissue before preparing nuclei to ensure that the approximate cell-type composition for each sample will be comparable[11,21].

Another strategy which has been recently applied to increase cost-effectiveness as well as to aid in batch effect correction is combining male and female samples in a single capture followed by using the expression of sex-specific, X-chromosome genes such as *XIST* and Y-chromosome genes such as *SRY*[31], or the chromosome accessibility ratios for sex-chromosome versus autosomes[32] to separate the cells from each sample. Since both samples are captured on the same lane of the microfluidic chip, it may be possible to account for lane to lane variability using this approach. Moreover, the use of cryosections of histological grade tissue blocks may be a good strategy to account for uniform input from a micro-anatomically heterogeneous regions such as the cerebral cortex[11,22].

Expertise needed to implement the protocol

This protocol requires access to a 10X Genomics' Chromium system and corresponding reagents, or an in-house droplet-based single-nucleus sequencing system. A hemocytometer or cell-counting microscope will be required for determining proper loading concentration. Wet-lab work will require familiarity with standard molecular biology approaches such as cDNA synthesis and sequencing library preparation.

Advantages and Limitations

We are unable to get information about cytoplasmic transcription which may be limiting for obtaining data for some cell types[33]. Some tissue types, such as spinal cord or intestine, may require additional processing, such as through FACS or collagenase treatment. Representation of

all cell types may not be uniform as different cell types are differentially susceptible to lysis during the isolation procedure.

We cannot rule out the possibility that the multiple rounds of washing and centrifugation incorporated into our protocol may result in damage to fragile tissue or unacceptable levels of loss of material if starting with small amounts of precious tissue. Moreover, we cannot rule out the possibility that use of the iodixanol gradient may cause biases in the types of nuclei recovered and this may require empirical assessment for different tissue types. We have not systematically assessed this bias.

Materials

BIOLOGICAL MATERIALS:

- Tissue samples: This protocol was successfully applied for processing frozen archived post-mortem prefrontal cortex tissue obtained from the Douglas Bell Canada Brain Bank, post-mortem intestinal tissue (with modifications such as collagenase treatment), and surgical samples of tumor tissue[25]. CAUTION All experiments involving the use of human samples must be performed in accordance with the relevant institutional and national regulations. Use of post-mortem tissues was approved by the Institutional Review Board of the Douglas Hospital.

REAGENTS:

- NP-40 detergent at 10% (vol/vol) concentration (Abcam, cat. no. ab142227)
- Bovine Serum Albumin Fraction V (BioShop, cat. no. ALB001.25)
- Tris (BioShop, cat. no. TRS003.5)
- NaCl (BioShop, cat. no. SOD001.1)
- $MgCl_2.6H_2O$ (BioShop, cat. no. MAG510)
- HCl (BioShop, cat. no. HCL333)
  CAUTION: Concentrated HCl is highly corrosive and should be handled inside a fume hood while wearing PPE.
- KCl (BioShop, cat. no. POC308)
- KOH (BioShop, cat. no. PHY202)
  CAUTION: Concentrated KOH is highly corrosive and should be handed inside a fume hood while wearing PPE.
- Tricine (BioShop, cat. no. TRI001)

- Glycerol (BioShop, cat. no. GLY001)
- Protector RNAse Inhibitor (Millipore Sigma, 3335399001)

CRITICAL: Other RNAse inhibitors may not be compatible with the protocol and may result in low yield of nuclei.

- Optiprep™ Density Gradient Medium, 60% weight/volume iodixanol (Millipore Sigma, D1556-250)
- Gibco™ PBS, pH 7.4 (1X), (Thermo Fisher Scientific, cat. no. 100100203)
- Deionized water
- Ethanol 100% (Sigma, cat. no. 459836-500ML)
- Trypan Blue Stain (0.4%) (Thermo Fisher Scientific, cat. no. T10282)
- Hoechst stain (Invitrogen, cat. no. H1399)
- Chromium Single Cell 3' Library & Gel Bead Kit v2 or newer (10X Genomics Inc, cat. no. 120237)
- Chromium Single Cell A Chip Kit (10X Genomics Inc, cat. no. 120236)
- SPRIselect Reagent Kit (Beckman Coulter, cat. no. B23318)
- Tween 20 (Bio-Rad, cat. no. 1610781)
- Buffer EB (250mL) (Qiagen, cat. no. 19086)
- Glycerin (glycerol), 50% (v/v) Aqueous Solution (Ricca Chemical Company (or other), cat. no. 3290-32)
- DynaBeads MyOne™ Silane Beads (5mL) (Thermo Fisher, cat. no. 37002D) (may be included in newer 10X Genomics snRNA-seq kits)
- Low TE Buffer (10 mM Tris-HCl pH 8.0, 0.1 mM EDTA, 100mL) (Thermo Fisher, cat. no. 12090-015)
- Nuclease-Free Water (Thermo Fisher, cat. no. AM9937)
- TapeStation High Sensitivity D1000 Sample Buffer (Agilent, cat. no. 5067-5603) or Tape Station High Sensitivity D5000 Sample Buffer & Ladder (Agilent, cat. no. 5067-5593)
- TapeStation High Sensitivity D1000 ScreenTape (Agilent, cat. no. 5067-5584) or High Sensitivity D5000 ScreenTapes (Agilent, cat. not. 5067-5592)

EQUIPMENT:

Lab equipment

- Scalpel
- Spatula
- Weighing boat
- Weighing scale
- Refrigerated bench-top centrifuge for 5mL tubes (Eppendorf, model 5430R)
- Refrigerated bench-top centrifuge for 15 mL tubes (Beckman Coulter, model Allegra X-14R)

- Countess® II FL Automated Cell Counter (Thermo Fisher Scientific, cat. no. AMAQAF1000)
- Countess® II FL Automated Cell Counter Chamber Slides (Thermo Fisher Scientific, cat. no. C10228)
- Flowmi™ Cell Strainer, 40 µm (Bel-Art, cat. no. H13680-0040)
- MACS® SmartStrainers, 30 µm (Miltenyi Biotec, cat. no. 130-098-458)
- 7 ml Tissue Grinder, Dounce (Wheaton, cat. no. 357542)
- 15 ml centrifuge tubes (Corning, cat.no. 430791)
- Centrifuge tube, 50 mL screw cap (Sarstedt, SAR62547205)
- DNA LoBind Microcentrifuge Tubes 1.5 mL (Eppendorf™, cat. no. 022431021)
- 250 mL glass bottles
- DNA LoBind Microcentrifuge Tubes 5.0 ml (Eppendorf, cat. no. 30108310)
- INCYTO C-Chip Disposable Hemocytometers (SKC Films Inc., cat. no. DHCN012 or DHCN015)
- TempAssure PCR 8-tube strip (USA Scientific, cat. no. 1402-4700)
- 10mL serological pipette
- Invitrogen EVOS FL Auto Cell Imaging System (Thermo Fisher Scientific)
- Chromium Controller (10X Genomics)
- Divided Polystyrene Reservoirs (25mL, 50) (VWR, cat. no. 41428-960)
- 200UL Filter Tips (Rainin, cat. no. 17007961)
- Pipet-Lite Multi Pipette L8-200XLS+ (Rainin, cat. no. 17013805)
- TapeStation 2200 (Agilent) or equivalent equipment

Software for sequence alignment and gene-barcode counting

- CellRanger version 2.1.0 CRITICAL Linux OS must meet the minimum requirements for running CellRanger as described on the 10X Genomics' webpage (https://support.10xgenomics.com/single-cell-gene-expression/software/overview/system-requirements).

- bcl2fastq2, version 2.19 (https://support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html)

Software for secondary analysis in R

- Seurat, version 2.3.0 or higher[34]
- mixtools (1.1.0)[35]
- R, version 3.4 or higher[36]

Software for analysis of snATAC-seq

- scATAC-pro version 1.1.4[37]

REAGENT SETUP

Stock solutions

CRITICAL The following reagents should be prepared ahead of time:

- Prepare 1M NaCl, 100mM $MgCl_2.6H_2O$, 1M $MgCl_2.6H_2O$, 1M KOH solutions in ddH_2O in separate 50 mL centrifuge tubes and store at room temperature (21-22^oC for our laboratory) for up to 6 months..
- Prepare 250 mL of 10% (weight/volume) BSA solution in a glass bottle by dissolving crystalline BSA in ddH_2O at room temperature. Store at 4^oC for up to 1 week. Keep crystalline BSA at -20^oC.
- Prepare 250 mL of 1M Tris HCl buffer in a glass bottle by dissolving Tris in ddH_2O. Adjust pH to 7.4 by adding HCl dropwise. Store at room temperature for up to 6 months. This is a time-consuming step.
- Prepare 250 mL of 0.5M Tricine KOH buffer in a glass bottle by dissolving Tricine in ddH_2O and adjusting pH to 7.8 by adding 1M KOH dropwise. This is a time-consuming step. Store at room temperature for up to 6 months.

CRITICAL Buffer recipes provided are calculated assuming preparation of 8 samples for capture using a full Chromium chip.

Optiprep™ diluent (altered as per Kriaucionis et al., 2009)[38]: Combine the following in a 250 mL glass bottle. Store at room temperature.

Component	Volume (in mL)	Final concentration
1M KCl	15	150 mM
1M $MgCl_2.6H_2O$	0.5	5 mM
0.5M Tricine-KOH (pH 7.8)	4	20 mM
Deionized water	80.5	–
Total volume	100	–

Optiprep™ solutions: Using Optiprep™ diluent solution dilute the Optiprep™ reagent to make 50% weight/volume iodixanol and 29% weight/volume iodixanol solutions from Optiprep™ solution in separate 50 mL centrifuge tubes. Protect from light and store at room temperature for up to 6 months.

Lysis buffer (LB): Combine the listed components in a 50 mL centrifuge tube. This buffer should be made fresh and kept at 4^oC or on ice.

Component	Volume (in µL)	Final concentration

1M Tris-HCl pH 7.4	200	10 mM
1M NaCl	200	10 mM
100 mM $MgCl_2.6H_2O$	600	3 mM
NP-40 (10%)	100	0.05% (v/v)
Deionized water	18900	–
Total volume	20000	–

Nuclei wash buffer (NWB): Combine the listed components in a 250 mL glass bottle. This buffer should be made fresh and kept at 4°C or on ice.

Component	Volume (in mL)	Final concentration
10% BSA	100	5% (w/v)
Glycerol	0.5	0.25% (v/v)
Protector RNAse inhibitor	0.2	40 units/mL
1X PBS	to 200 (~100)	0.5X
Total volume	200	–

EQUIPMENT SETUP

- Precool both centrifuges to 4°C.
- Set up EVOS FL Auto microscope with 10X magnification, bright field and DAPI channels.

Procedure

Tissue preparation (Timing 1-4 hours)

1. Cut tissue using a scalpel and weigh 30-50 mg of frozen tissue per sample. Keep tissue on dry ice while cutting to minimize degradation. Transfer to a 1.5mL microcentrifuge tube

using a spatula and place back on dry ice. Clean scalpel and spatula with 70% ethanol (v/v) between samples. Use a fresh weigh boat for each sample. Alternatively, this step could be replaced by cryosectioning a fresh frozen histology grade dissection of tissue and collecting several sections such that the total weight is between 30-50 mg.

CAUTION: Post-mortem human tissue can contain pathogens. Take precautions including wearing PPE and seek medical attention if the scalpel breaks your skin.

Nuclei extraction (Timing 2-3 hours)

2. Transfer tissue using spatula to douncing tube on ice. Add 3mL of ice-cold lysis buffer and dounce with loose pestle 10 times and 5 more times with the tight pestle.

   CRITICAL STEP: Use proper douncing technique to ensure proper mechanical breakdown of tissue. Proceed slowly and avoid bubbles. Grind tissue against the bottom of the tube using the douncer with each stroke.

3. Transfer homogenized tissue to a 15mL centrifuge tube by pouring and add 2 mL of chilled lysis buffer. Incubate on ice for 5 minutes, gently swirling to mix 2 times during incubation.
4. Add 5 ml of chilled wash buffer to lysed tissue to quench lysis. Swirl to mix.
5. Place 30 µm MACs SmartStrainer on a 15 mL centrifuge tube. Pipette lysed tissue suspension on top of filter to remove cell debris and large clumps. In case of blocked flow through the filter, tap filter gently to encourage the suspension to flow through.
6. In the precooled Allegra-14X centrifuge, spin down the lysed tissue suspensions at 500g for 5 minutes at 4°C.
7. Decant supernatant into a waste beaker without disrupting the nuclei pellet.

   CAUTION: The supernatant should be treated as biohazardous waste and treated with bleach before disposal.

   CRITICAL STEP: Pour out supernatant in a single motion as repeated pouring motions can dislodge the pellet. If the pellet dislodges during decanting slowly remove the supernatant using a pipette.
8. Using a 10 mL serological pipette add 10mL of nuclei wash buffer to the pelleted nuclei and gently pipet 8-10 times to mix.
9. Repeat step 5-7 using the resuspended nuclei.
10. Using a 10 mL serological pipette add 5mL of nuclei wash buffer to the pelleted nuclei and gently pipet 8-10 times to mix.
11. Repeat steps 6-7 using the resuspended nuclei.
12. Using a 1000 µL pipette tip, add 1 mL of nuclei wash buffer to pelleted nuclei and gently pipet 8-10 times to mix.
13. Add 1 mL of 50% (w/v) working solution of iodixanol (Optiprep™) to the nuclei and mix well to obtain 2 mL of 25% (w/v) iodixanol solution containing nuclei.

14. Prepare an iodixanol cushion of 2 ml of 29% (w/v) iodixanol solution in a 5mL Eppendorf centrifuge tube.
15. Gently add the 2 mL nuclei suspension on top of the iodixanol cushion by pipetting slowly against the wall of the tube to avoid mixing.
16. In the precooled Eppendorf centrifuge, spin the tubes containing nuclei layered over iodixanol cushion at 10,000 g for 30 minutes at 4°C.
17. Carefully pour out the supernatant leaving the least possible amount of volume in the tube without disrupting the pellet. **?TROUBLESHOOTING**
18. Using a 1000 µL pipette tip, resuspend the nuclei pellets in 500 µL or less of nuclei wash buffer. Gently pipette 8-10 times or until nuclei are resuspended.
19. For a quick estimate of nuclei concentration, mix 10 µL of the nuclei suspension with 10 µL of Trypan blue in a separate tube. Load 10 µL of the mixture onto a Countess hemocytometer slide. Count nuclei on the Countess hemocytometer and measure range of sizes. For human nuclei from archived post-mortem cortical tissue we have observed that the average diameter is around 10 µm. However, nuclei may have a range of sizes and it is only concerning if a long tail of particles of more than 30 µm are detected at this may indicate debris and aggregation. Trypan blue is a live dead stain and properly isolated nuclei should be marked as dead cells. **?TROUBLESHOOTING**
20. Using the estimated count from the Countess, dilute nuclei to around 500,000 cells/ mL or 500 cells/ µL by adding an appropriate volume of nuclei wash buffer. It may be possible to increase these concentrations for capturing more nuclei.

    CRITICAL STEP: If the concentration of nuclei is too high it can result in aggregation which will prevent efficient capture of single nuclei in subsequent steps. We have achieved good suspensions and capture with up to 1000 nuclei/ µL, but if aggregation is observed, lower concentrations (as low as 500 nuclei/ µL) may be better.

21. Add Hoechst stain to the resuspended nuclei at a 1:2000 dilution to obtain counts using fluorescence microscopy.

Nuclei concentration assessment and capture (Timing 45 mins)

CRITICAL Immediately prior to loading the Hoechst stained nuclei on the Chromium system, check to make sure the nuclei are well segregated (nuclei may clump) and recheck sample concentration using a fluorescent microscope like the Evos FL Auto (Thermofisher)

22. OPTIONAL: Use a 1mL pipette to take a minimum of 200 µL of the sample and filter it through a 40µm Flowmi pipette tip filter before counting and loading. This will get rid of the clumped nuclei and large debris that can clog the microfluidics of Chromium chips.

23. Load 10µL of sample onto a disposable hemocytometer slide and into the Evos.

CRITICAL STEP: Make sure to pipette up and down the full volume of the nuclei suspension several times to avoid settling of the pellet at the bottom of the tube before loading for accurate counts.

24. Set the Evos to 10X magnification and image the entire hemocytometer grid field of view in DAPI and bright field. This image makes counting nuclei easier and serves as a record as well (**Fig. 3g**). If a fluorescence microscope is not available, it may be sufficient to use the counts based on Trypan blue staining, but unfortunately it will not be possible to distinguish between debris and nuclei using this approach. On the other hand, using a fluorescence microscopy may be more time-consuming for nuclei counting when processing many samples for capture. User's discretion and considering the state and type of tissue are important parameters when determining cell counting strategy.
25. In parallel to steps 22-24, make the RT Master Mix and aliquot appropriate volumes of master mix and water into PCR tubes according to the number of nuclei to be targeted for capture, referring to the Chromium protocol CG00052 Rev. D or later. **?TROUBLESHOOTING**

CRITICAL STEP: The Chromium capture rate for nuclei from archived tissue is lower (~20% less) than the capture rate for cells. To account for this, it is necessary to adjust the count used to determine loading volume. We empirically determined that choosing the loading volume by using a count that is 30% less than the observed count worked best for our samples. This adjustment may vary from tissue type to tissue type. For example, if the sample has a concentration of 500 nuclei/µL, the sample volume should be loaded as if it has 350 cells/µL (70% of 500 cells) in order to recover the targeted number of nuclei.

CRITICAL STEP: Resuspend nuclei by pipette, mixing the full volume several times immediately before loading to prevent aggregation of nuclei.

26. Load the Chromium Chip and harvest the nuclei captured in droplets (i.e. GEMs) according to the Chromium protocol CG00052 Rev. D

Library preparation and sequencing (Timing as per the Chromium protocol, ~8 h split over 2 days)

27. Perform reverse transcription, cDNA amplification, and library preparation according to the Chromium protocol CG00052 Rev. D. Libraries can be sequenced on an Illumina sequencer. Sequencing two samples per lane of a HiSeq 4000 machine can yield 150,000,000, reads per sample. This can translate to ~50,000 reads per nucleus if capturing 3000 nuclei per sample based on default CellRanger parameters and provides sufficient information for cell-type identification and differential expression analysis. However, the exact number of reads per cell will depend on how many nuclei are loaded and on the algorithm used to call cells.

Downstream analysis options for the sequencing results are described in Box 1.

Box 1: Downstream data analysis

Alignment, Demultiplexing, and Generation of Counts Matrix (Time: variable)

Since our experiments utilized human nuclei, we built a pre-mRNA reference using the cellranger mkref (Cellranger version 2.0.1) command. Default parameters were used, starting with the refdata-cellranger-GRCh38-1.2.0 transcriptome and as per the instructions provided on the 10X Genomics website. For mouse tissue the corresponding pre-mRNA reference would need to be created for the mouse genome. We demultiplexed reads by sample index using the cellranger mkfastq command (Cellranger v2.1.0), aligned FASTQ files to the custom transcriptome, demultiplexed cell barcodes, counted the UMIs corresponding to genes using the cellranger count command and default parameters. These steps may be performed with custom code if desired.

Custom Filtering to Recover Low Transcript Number Cell Types (Time: variable)

While there are many options for software to be used for downstream analysis of snRNA-seq data such as scater[39], SC3[40], Monocle3[41], etc., we used the Seurat R package (version 2.2.0, 2.3.0)[42]. Unfiltered gene barcode matrices for each sample were loaded into R using the Read10X function. At this step, cell names can be modified such that the subject name, batch, and biological condition are appended 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 combined sequentially using the MergeSeurat function. No filtering or normalization was performed up to this step. Since we were working with a single nucleus dataset, all mitochondrial genes that

are transcribed from the mitochondrial genome were removed, along with genes not detected in any cell. More recently several methods have been developed to align multiple datasets of snRNA-seq and other single-cell level data[30,34] which can be used for combining the data from individual subjects if inter-individual variability or batch effects are deemed to have a large influence on the results.

For preliminary filtering, some nuclei with very low number of genes detected (<110) and nuclei with very high numbers of UMIs detected (in the top 0.5%) were removed as low-quality nuclei and potential multiplets respectively. These cut-offs are arbitrary but can be based upon the distribution of the data. For example, in our dataset there was a sharp increase in the number of UMIs from 16,393 at the 99.5th percentile to 102,583 at the maximum which probably represents the multiplets in the dataset.

If the dataset contains multiple cell-types which are expected to be heterogenous in terms of the number of molecules of RNA present per nucleus (such as when the nuclei of different cell-types are known to be of very different size), the following approach can be used for removing low quality cells without unduly biasing the filtering against nuclei which biologically contain fewer molecules . For our dataset, given the known trend for higher number of RNA molecules in neuronal nuclei compared to glial nulcei[10,11,43], the distribution of number of UMIs was fit with three normal distributions using the normalmixEM function from the mixtools[35] package. The rationale is that the filtered barcodes contain a population of low quality “noise” barcodes that have a very low number of UMIs on average, a population of non-neuronal cells that have an intermediate numbersof UMIs and a population of neuronal cells that have a high number

of UMIs. After fitting 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. As an example, for a subset of 20 subjects, applying our custom filtering approximately doubled the total number of cells, as more cells which were previously discarded as empty barcodes are now included, but increased the number of non-neuronal cells by almost 6 times[24].

Of note, in our experience with the newest version of Cell Ranger (3 and above) which incorporates the EmptyDrops algorithm for cell calling, it may not be necessary to customize the process of calling cells to account for biases in number of gene and RNA molecules across cell-types.

TROUBLESHOOTING

Troubleshooting advice can be found in Table 1.

Table 1: Troubleshooting table.

Step	**Problem**	**Possible reason**	**Solution**
17	Absence of visible pellet after Optiprep™ cushion centrifugation	A large pellet may indicate presence of excessive debris rather than high nuclei yield and absence of a visible pellet is not	Continue with downstream steps assuming the location of the nuclei based on the direction in which the tube is placed within the centrifuge

		necessarily cause for concern. However, in certain cases it may indicate very low yield of nuclei.	and assess nuclei yield under the microscope. If very low nuclei yield is observed, consider increasing the amount of input material.
19	Low yield of nuclei	Too little starting material (<30 mg)	Consider eliminating one of the wash steps (9-10 in protocol) and resuspending in less volume (5 mL instead of 10 mL) in step 8.
24	Number of nuclei captured does not meet the expected number based on the table provided by the 10x loading guidelines	The capture rate for nuclei may not be the same as that for cells.	Empirically determine the difference between the capture rate expected and observed and adjust loading volume accordingly.

TIMING

Step 1, Tissue preparation: 1 hour for 8 samples if cutting pieces using a scalpel, up to 4 hours if collecting cryosections

Step 2-21, Nuclei extraction: 2-3 hours

Step 22-26, Nuclei concentration assessment and capture: 45 minutes

Step 27, Library preparation and sequencing, can be split into two 4 hours blocks on 2 days

Anticipated Results

We expect our single-nuclei extraction protocol to produce high quality single-nucleus suspensions **(Fig. 3)** from frozen archived post-mortem tissues. The nuclear suspensions are relatively free from debris and do not show substantial aggregation of nuclei even after 16 hours of refrigeration, upon visual inspection **(Fig. 3-f)**. Capture of single-nuclei using these nuclear suspensions on a microfluidic device for droplet-based snRNA-seq reproducibly produces high-quality libraries with sufficient cDNA yield for sequencing **(Fig. 4)**. The variability in sample parameters such as PMI, archival time, pH and RIN did not affect most of the quality metrics of snRNA-seq results with this nuclei extraction protocol (**Fig. 2**). The samples for which data are presented in Figure 2 were processed using two different gradients of iodixanol – a weaker gradient using 29% and 25% volume/volume dilutions of Optiprep™ reagent (majority of samples) and a stronger gradient using the 29% and 25% weight/volume dilutions of iodixanol as described in this protocol and previously[38]. We subsequently found that the stronger gradient produces cleaner nuclei preparations and yields better sequencing quality control metrics, such as higher fraction of reads in cell, and higher numbers of UMIs and genes detected, especially using the updated Cell Ranger 3 pipeline and 10X Genomics v3 single-cell sequencing chemistry. Thus, the protocol published herein utilizes the 29% and 25% weight/volume dilutions of iodixanol for the gradient. Finally, we have produced preliminary results using nuclei extracted from post-mortem brain with our extraction protocol as input to the single-nucleus ATAC-seq approach employing 10X Genomics Chromium for single-cell capture. Using MACS2[44] for peak

calling and the scATAC-pro pipeline[37] for cell calling, we achieved about 83% of total fragments uniquely mapped to genome assembly GRCh38, fraction of reads in peak (FRiP) scores of up to 23%, median fragments mapping per cell in the range of 12,000 – 15,000, and TSS (transcription start site) enrichment, according to ENCODE definition, indicating a signal-to-noise ratio of more than 3.9 (Fig 4e-f).

Data Availability

Raw sequencing data are accessible on GEO using the accession number GSE144136.

Acknowledgements
GT 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 (Josée Prud'homme, Maâmar Bouchouka and Annie Baccichet), and the technology development team at the McGill University and Genome Quebec Innovation Centre. 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

MM and C. Nagy developed nuclei extraction protocol, prepared nuclei, and wrote the manuscript. YCW preformed 10X snRNA-seq protocol. C. Nascimento performed 10X snATAC-seq

protocol. AC performed snATAC-seq data analysis. MS and JFT guided bioinformatic analysis. NM contributed to tissue processing and data interpretation. JR provided technical single-cell expertise and experimental support. GT provided general oversight, including in experimental design. All authors contributed to manuscript preparation.

Competing Interests

No, I declare the authors have no competing interests as defined by Nature Research, or other interests that might be perceived to influence the interpretation of the article.

Figures and Figure Legends

Figure 1

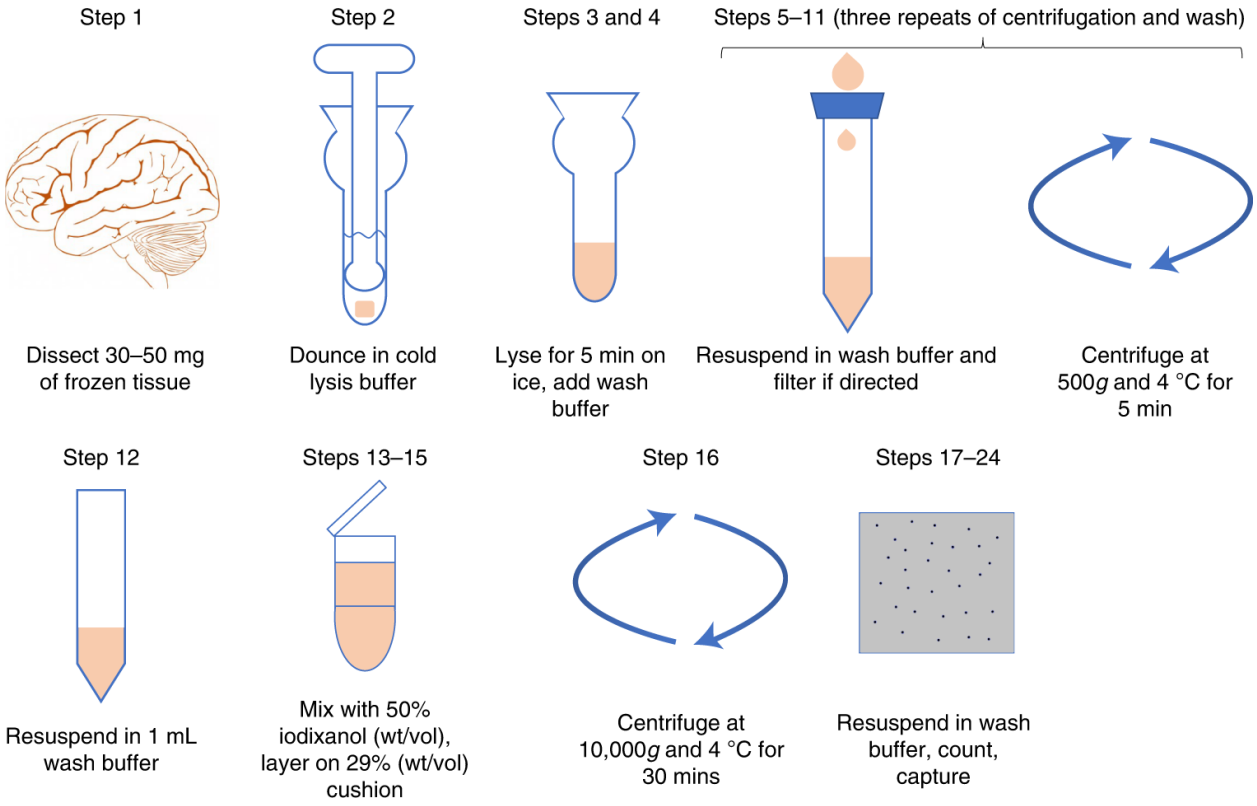


Figure 1: Schematic representation of the steps of the protocol. Frozen tissue is dissected, homogenized by douncing, lysed, and then washed, filtered, and centrifuged several times until a single-nucleus suspension is obtained.

Figure 2

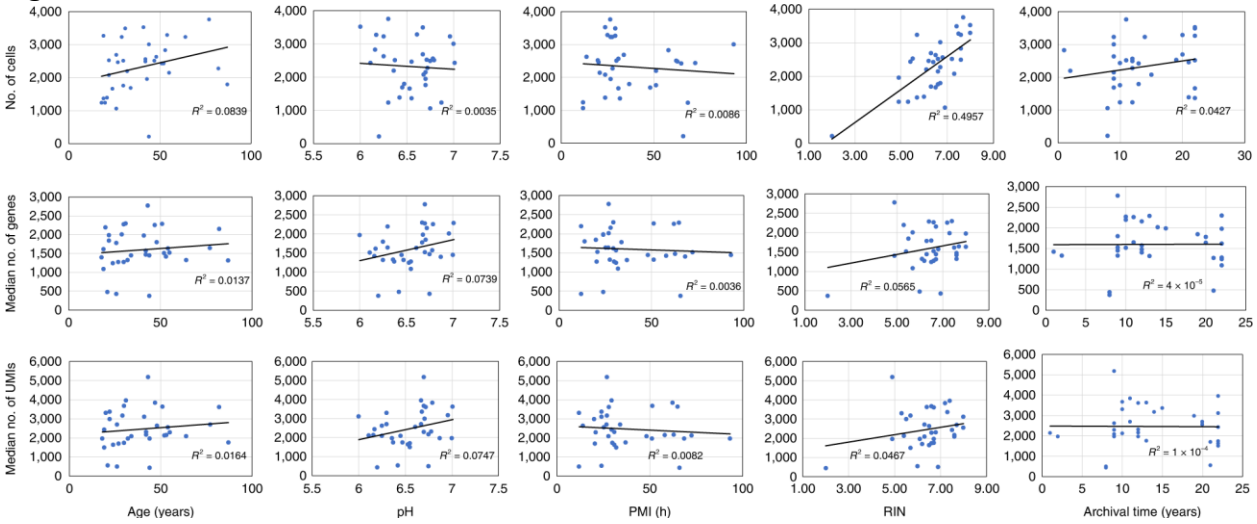


Figure 2: Effect of sample quality parameters on single-nucleus capture and sequencing metrics. The archival time, post mortem interval (PMI), age, and pH of the brains accounted for less than 10% of the variation in number of cells (nuclei) retained after filtering, median number of genes per nucleus, and median no of UMIs (unique molecular identifiers) per nucleus. The RIN of the samples had had a significant negative effect ($p = 3.4x10^{-6}$) on the number of nuclei captured, especially for RIN < 4, but did not have a large effect on the median numbers of genes or UMIs. The R^2 values based on Pearson correlations and p-values (n = 34 samples) were calculated using the cor.test function in R. Linear trendlines were added using Microsoft Excel. For two of the data points the median UMIs, median genes, and number of cells are the aggregated values of two runs for those samples. All data in this figure are from the dataset published in Nagy et al. (2020)[24].

Figure 3

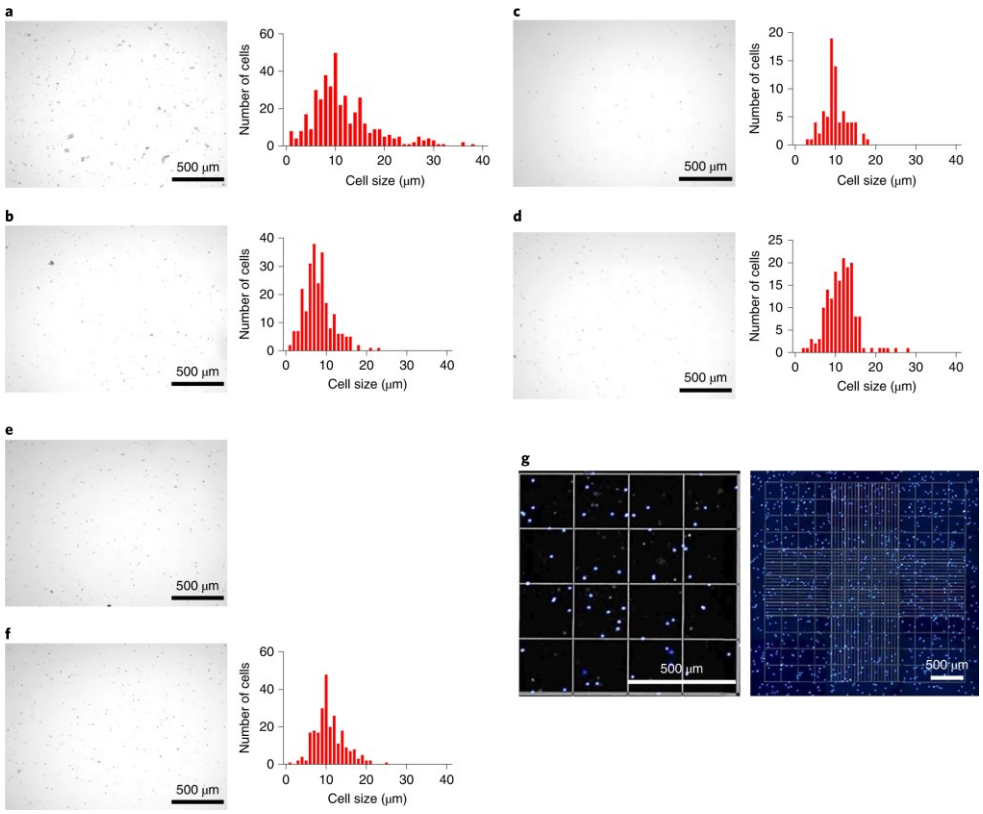
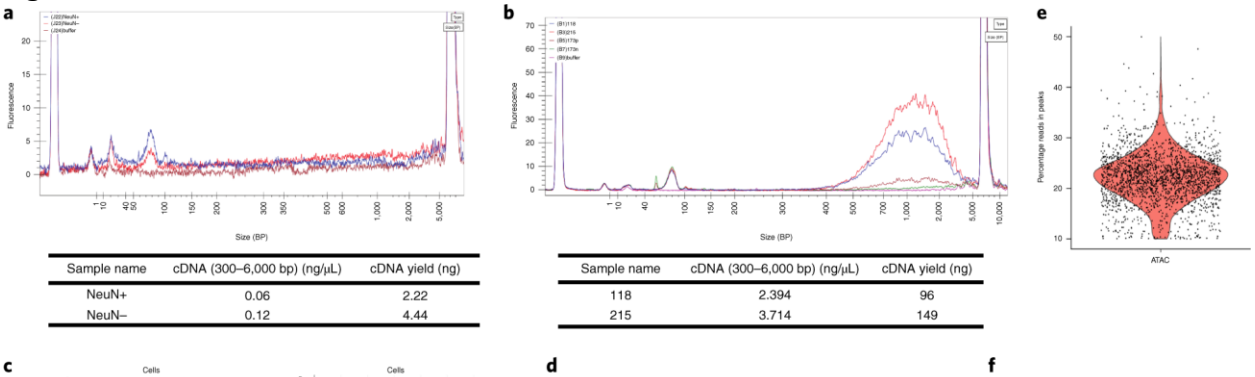


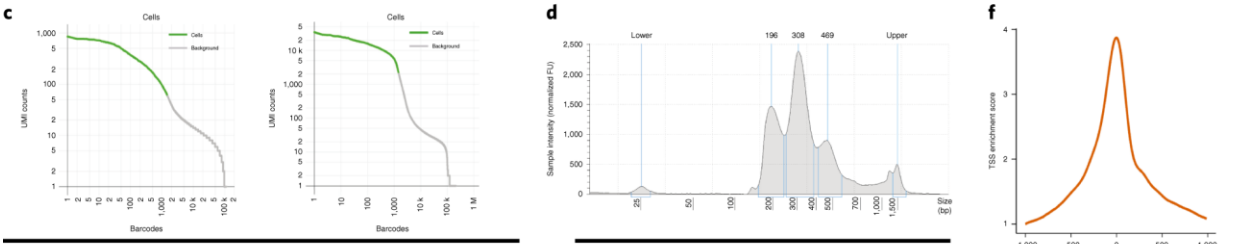
Figure 3: Images of extracted nuclei. (a-d) Before optimization, using the 10X Genomics demonstrated protocol, the extracted nuclei from two different samples (a and b) show large amounts of debris and the size distribution is skewed towards larger sizes (> 10 µm). After optimization, representative images of nuclei extracted from two samples (c and d) show much less debris and size-distributions are centered around 10 µm, within the expected range for human brain nuclei. Images were acquired with the Countess Cell Counter using Trypan blue for staining. Note that extracted nuclei should be marked as dead cells, as seen. (e-f) Extracted nuclei do not tend to aggregate even after (e) 2.5 hours or (f) 16 hours of storage at 4°C. Note that the size distribution after 16 hours is still centered around 10 µm, indicating an absence of aggregation. (g) Representative images of extracted nuclei stained with Hoechst (1:2000) acquired at 10X magnification on the Evos microscope. Figures (c), (d), and (g) correspond to samples used in Nagy et al. (2020)[24]. All scale bars represent 500 µm.

Figure 4



Sample name	cDNA (300–6,000 bp) (ng/μL)	cDNA yield (ng)
NeuN+	0.06	2.22
NeuN–	0.12	4.44

Sample name	cDNA (300–6,000 bp) (ng/μL)	cDNA yield (ng)
118	2.394	96
215	3.714	149



Sample name	Number of reads	Estimated no. of cells	Mean reads per cell	Median genes per cell	Median UMI count per cell
NeuN+	165,749,748	1,507	109,986	80	124
215	144,125,115	1,486	96,988	3,501	7,287

Sample name	cDNA (163–700 bp) (pg/μL)	cDNA yield (ng)
84	6,140	

Figure 4: cDNA traces and quality metrics for snRNA-seq libraries before and after optimization of nuclei extraction. (a) A FANS based nuclei isolation (using Millipore anti NeuN-PE FCMAB317PE antibody and DRAQ5 both at 1:300 dilution) of single-nucleus suspensions prepared as per Lutz et al. (2017)[45] resulted in very low yield cDNA libraries whereas (b) the optimized nuclei extraction protocol resulted in good yield of cDNA in the expected size range. Samples in both (a) and (b) are derived from archived post-mortem brain tissue. Perkin Elmer Caliper traces are shown for snRNA-seq cDNA libraries at a dilution of 1:6. The expected library size is between 200-9000 bp and here we performed quantification in the 300-600 bp range. The minimum yield of cDNA should be 2 ng and as can be seen the yield was much improved (>90 ng) after protocol optimization. (c) With similar numbers of sequencing reads and median reads per cell, the libraries produced using the optimized nuclei extraction protocol have much higher median numbers of genes and UMIs per cell as can be seen from the elbow plot produced by Cell Ranger

as well as the tabulated summary metrics. The NeuN+ sample (cDNA trce shown in (a)) was processed with Cell Ranger 1.3.1 and the hg19 transcriptome while sample 215 (cDNA trace shown in (b)) was processed with Cell Ranger 2.1.0 and the GRCh38-1.2.0 pre-mRNA reference. (d) Our nuclei extraction protocol is compatible with 10X Genomics' commercial snATAC-seq protocol, as can be seen from the Tapestation trace and fragment-size distribution (163-700bp) of a successfully prepared snATAC-seq library prepared from archived post-mortem human brain tissue. (e) Preliminary processing of snATAC-seq data aligned to the hg38 genome, showing the distribution of the percentage of reads in peaks (pct_reads_in_peaks) across captured nuclei and the transcription start site (TSS) enrichment score across captured nuclei for a post-mortem human brain sample. Samples 118 and 215 in (b) and sample 215 in (c) are from the Nagy et al. (2020)[24] dataset.

References:

1. Macosko, Evan Z*., et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202-1214 (2015).
2. Zilionis, R*., et al.* Single-cell barcoding and sequencing using droplet microfluidics. *Nat Protoc* **12**, 44-73 (2017).
3. Zheng, G.X.Y*., et al.* Massively parallel digital transcriptional profiling of single cells. **8**, 14049 (2017).
4. Haber, A.L*., et al.* A single-cell survey of the small intestinal epithelium. *Nature* **551**, 333 (2017).
5. Wang, Y*., et al.* Pulmonary alveolar type I cell population consists of two distinct subtypes that differ in cell fate. *Proceedings of the National Academy of Sciences* **115**, 2407-2412 (2018).
6. Sathyamurthy, A*., et al.* Massively Parallel Single Nucleus Transcriptional Profiling Defines Spinal Cord Neurons and Their Activity during Behavior. *Cell reports* **22**, 2216-2225 (2018).
7. Hrvatin, S*., et al.* Single-cell analysis of experience-dependent transcriptomic states in the mouse visual cortex. *Nature neuroscience* **21**, 120-129 (2018).
8. Hochgerner, H., Zeisel, A., Lönnerberg, P. & Linnarsson, S. Conserved properties of dentate gyrus neurogenesis across postnatal development revealed by single-cell RNA sequencing. *Nature neuroscience* **21**, 290-299 (2018).
9. Lake, B.B*., et al.* Neuronal subtypes and diversity revealed by single-nucleus RNA

sequencing of the human brain. *Science* **352**, 1586-1590 (2016).
10. Habib, N.*, et al.* Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nature Methods* **14**, 955 (2017).
11. 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).
12. Sullivan, P.F., Neale, M.C. & Kendler, K.S. Genetic epidemiology of major depression: review and meta-analysis. *The American journal of psychiatry* **157**, 1552-1562 (2000).
13. Lacar, B.*, et al.* Nuclear RNA-seq of single neurons reveals molecular signatures of activation. *Nature Communications* **7**, 11022 (2016).
14. van den Brink, S.C.*, et al.* Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. *Nature Methods* **14**, 935 (2017).
15. Grindberg, R.V.*, et al.* RNA-sequencing from single nuclei. *Proceedings of the National Academy of Sciences* **110**, 19802-19807 (2013).
16. 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. *Scientific reports* **7**, 6031 (2017).
17. Cutler, A.A., Jackson, J.B., Corbett, A.H. & Pavlath, G.K. Non-equivalence of nuclear import among nuclei in multinucleated skeletal muscle cells. *Journal of Cell Science* **131**(2018).
18. Berridge, B.R., Bolon, B. & Herman, E. Chapter 10 - Skeletal Muscle System. in *Fundamentals of Toxicologic Pathology (Third Edition)* (eds. Wallig, M.A., Haschek, W.M., Rousseaux, C.G. & Bolon, B.) 195-212 (Academic Press, 2018).
19. Mathys, H.*, et al.* Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332-337 (2019).
20. Grubman, A.*, et al.* A single-cell atlas of entorhinal cortex from individuals with Alzheimer's disease reveals cell-type-specific gene expression regulation. *Nature neuroscience* **22**, 2087-2097 (2019).
21. Jäkel, S.*, et al.* Altered human oligodendrocyte heterogeneity in multiple sclerosis. *Nature* **566**, 543-547 (2019).
22. Velmeshev, D.*, et al.* Single-cell genomics identifies cell type–specific molecular changes in autism. *Science* **364**, 685 (2019).
23. Krishnaswami, S.R.*, et al.* Using single nuclei for RNA-seq to capture the transcriptome of postmortem neurons. *Nat Protoc* **11**, 499-524 (2016).
24. Nagy, C.*, et al.* Single-nucleus transcriptomics of the prefrontal cortex in major depressive disorder implicates oligodendrocyte precursor cells and excitatory neurons. *Nature neuroscience* **23**, 771-781 (2020).
25. Jessa, S.*, et al.* Stalled developmental programs at the root of pediatric brain tumors. *Nature genetics* **51**, 1702-1713 (2019).
26. Reiner, B.C.*, et al.* Single-nuclei transcriptomics of schizophrenia prefrontal cortex primarily implicates neuronal subtypes. *bioRxiv*, 2020.2007.2029.227355 (2020).
27. Sorrells, S.F.*, et al.* Immature excitatory neurons develop during adolescence in the human amygdala. *Nature Communications* **10**, 2748 (2019).
28. Proudfoot, N.J., Furger, A. & Dye, M.J. Integrating mRNA Processing with Transcription. *Cell* **108**, 501-512 (2002).

29. Amamoto, R.*, et al.* Probe-Seq enables transcriptional profiling of specific cell types from heterogeneous tissue by RNA-based isolation. *eLife* **8**, e51452 (2019).
30. Welch, J.D.*, et al.* Single-Cell Multi-omic Integration Compares and Contrasts Features of Brain Cell Identity. *Cell* **177**, 1873-1887.e1817 (2019).
31. Skelly, D.A.*, et al.* Single-Cell Transcriptional Profiling Reveals Cellular Diversity and Intercommunication in the Mouse Heart. *Cell reports* **22**, 600-610 (2018).
32. Cusanovich, D.A.*, et al.* The cis-regulatory dynamics of embryonic development at single-cell resolution. *Nature* **555**, 538-542 (2018).
33. Thrupp, N.*, et al.* Single-Nucleus RNA-Seq Is Not Suitable for Detection of Microglial Activation Genes in Humans. *Cell reports* **32**, 108189 (2020).
34. Stuart, T.*, et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e1821 (2019).
35. Benaglia, T., Chauveau, D., Hunter, D.R. & Young, D.S. mixtools: An R Package for Analyzing Mixture Models. *2009* **32**, 29 (2009).
36. R Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing* (2017).
37. Yu, W., Uzun, Y., Zhu, Q., Chen, C. & Tan, K. scATAC-pro: a comprehensive workbench for single-cell chromatin accessibility sequencing data. *Genome Biology* **21**, 94 (2020).
38. Kriaucionis, S. & Heintz, N. The Nuclear DNA Base 5-Hydroxymethylcytosine Is Present in Purkinje Neurons and the Brain. *Science* **324**, 929-930 (2009).
39. McCarthy, D.J., Campbell, K.R., Lun, A.T.L. & Wills, Q.F. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics (Oxford, England)* **33**, 1179-1186 (2017).
40. Kiselev, V.Y.*, et al.* SC3: consensus clustering of single-cell RNA-seq data. *Nature Methods* **14**, 483-486 (2017).
41. Cao, J.*, et al.* The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496-502 (2019).
42. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature Biotechnology* **36**, 411 (2018).
43. Darmanis, S.*, et al.* A survey of human brain transcriptome diversity at the single cell level. *Proceedings of the National Academy of Sciences* **112**, 7285-7290 (2015).
44. Zhang, Y.*, et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biology* **9**, R137 (2008).
45. 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. *The American journal of psychiatry* **174**, 1185-1194 (2017).

Appendix 2: Permissions to Reprint

Nature Portfolio Journals

Permission to reprint for author reuse of Nature Portfolio journals is described at the link below and the relevant information is copied below:

https://www.nature.com/nature-portfolio/reprints-and-permissions/permissions-requests#Author%20reuse

"Authors have the right to reuse their article's Version of Record, in whole or in part, in their own thesis. Additionally, they may reproduce and make available their thesis, including Springer Nature content, as required by their awarding academic institution.

Authors must properly cite the published article in their thesis according to current citation standards."

The relevant citations are:

Chapter 2: Nagy, C., Maitra, M., Tanti, A. et al. Single-nucleus transcriptomics of the prefrontal cortex in major depressive disorder implicates oligodendrocyte precursor cells and excitatory neurons. Nat Neurosci 23, 771–781 (2020). https://doi.org/10.1038/s41593-020-0621-y

Chapter 3: Maitra, M., Mitsuhashi, H., Rahimian, R. et al. Cell type specific transcriptomic differences in depression show similar patterns between males and females but implicate distinct cell types and genes. Nat Commun 14, 2912 (2023). https://doi.org/10.1038/s41467-023-38530-5

Appendix 1 : Maitra, M., Nagy, C., Chawla, A. et al. Extraction of nuclei from archived postmortem tissues for single-nucleus sequencing applications. Nat Protoc 16, 2788–2801 (2021).

https://doi.org/10.1038/s41596-021-00514-4

Current Behavioral Neuroscience Reports

Permission to reprint for author reuse of Springer Nature content is described at the link below and the relevant information is copied below:

https://www.springer.com/gp/rights-permissions/obtaining-permissions/882

"Authors have the right to reuse their article's Version of Record, in whole or in part, in their own thesis. Additionally, they may reproduce and make available their thesis, including Springer Nature content, as required by their awarding academic institution. Authors must properly cite the published article in their thesis according to current citation standards."

The relevant citation is:

Chapter 1 review of single-cell technologies: Maitra, M., Nagy, C. & Turecki, G. Sequencing the Human Brain at Single-Cell Resolution. Curr Behav Neurosci Rep 6, 197–208 (2019). https://doi.org/10.1007/s40473-019-00192-3

Co-first author permission

June 7, 2023

I, Corina Nagy, give full permission for the co-authored publications, Nagy, Maitra et al. Nature Neuroscience (2020) PMID: 32341540 and Maitra, Nagy et al. Nature Protocol (2021) PMID: 33972803 to be used as a part of Malosree Maitra's PhD thesis, submitted to the IPN at McGill University.

Corina Nagy, PhD.