UCSF
UC San Francisco Electronic Theses and Dissertations

Title

Precision Medicine Approach to Elucidate APOE Genotype and Sex-Specific Transcriptomic Changes in Alzheimer's Disease

Permalink

https://escholarship.org/uc/item/62q0f2cc

Author

Belonwu, Stella

Publication Date

2021

Supplemental Material

https://escholarship.org/uc/item/62q0f2cc#supplemental

Peer reviewed|Thesis/dissertation

Precision Medicine Approach to Elucidate APOE Genotype and Sex-Specific
Transcriptomic Changes in Alzheimer's Disease

by
Stella Belonwu

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Pharmaceutical Sciences and Pharmacogenomics

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

Martin Kampmann
_____
Chair

Marina Sirota
_____

Yadong Huang
_____

_____

_____
Committee Members

DEDICATION

This work is dedicated to Ndị nke Ajuluchuku and Belọnwụ

Dalu

Igbo Kwenu!

ACKNOWLEDGEMENTS

Whilst becoming a scientist, there have been several challenges to overcome. Science is slow, and there are many pressures, unforeseen variables, and failures along the way. Being a first-generation American and PhD-trained scientist, I was convinced that I could tackle these on my own. Fortunately, that was temporary, and I sought out a community of peers, mentors, and resources that helped guide me on my UCSF journey.

First and foremost, I would like to thank my dissertation advisor Dr. Marina Sirota for her immeasurable support. When I first met Marina, I was seeking advice on a project. With just that conversation, I was moved by how creative she was, and how she talked about her trainees with high regard. Since joining her lab, I have felt more respected as a scientist. She is very open-minded, understanding, and available. She always made me feel like I was part of the team, and no matter what I did, even in moments of doubt or confusion, I knew that I had her support. She's never short of wisdom, she is one of the most hardworking people I know, and I am always in awe of everything that she does. Her curiosity and drive have been very infectious and part of what makes her a great leader. I couldn't have chosen a better PhD mentor, and I am very grateful to have been a part of her lab.

Next, I would like to thank Dr. Martin Kampmann and Dr. Yadong Huang, members of my dissertation committee for all their guidance and support on my PhD project and career exploration. I would also like to thank Dr. D'Anne Duncan, Dr. Allan Balmain, Naledi Saul, Dr. Anum Azam, Dr. Gabriela Monsalve, Dr. Aparna Lakkaraju, Dr. Trever Bivona, Dr. Liz Silva, Dr. Katie Pollard, and Dr. Rada Savic for their roles in helping me navigate grad school.

Their support was unconditional, and I am honored to have them in my network. I would like to thank Manish Paranjape, Dr. Alice Taubes, Dr. Arjun Arkal Rao, Dr. Brian Grone, Dr. Kelly Zalocusky, Dr. Dena Dubal, Dr. Gabriela Fragiadakis, and the Carry the One Radio podcast team for their time, helpful insights, and collaboration.

My time in the Sirota lab presented a lot of opportunities that helped me grow personally and professionally, in ways I could not previously imagine. I reviewed a manuscript for publication, I mentored three junior scientists, I presented at and got awards from a conference, I submitted three papers for publication, and I met an amazing group of people. I would like to thank Dr. Idit Kosti, Dr. Dan Bunis, Dr. Katharine Yu, Dr. Tomiko Oskotsky, Dr. Dmitry Rychkov, Caroline Warly Solsberg, Alice Tang, Roman Kouznetsov, Yaqiao Li, and the rest of the Sirota lab and affiliates for their endless support and open collaboration.

I would also like to extend a special thank you to my UCSF and non-UCSF friends, especially Dr. Adam Melgoza, Dina Buitrago Silva, Jhia Jackson, Dr. Aude Bouagnon, Dr. Cesar Marquez, Dr. Emma Hughes, Rosie Williams, and Kwabena Andoh-Baidoo for always believing in me, supporting me, correcting me, and tolerating my shenanigans ;). They have played a big part in my PhD and provided an irreplaceable, sustainable community that helped me grow into who I am today. I cherish all our memories thus far and I look forward to what the future brings.

Most importantly, I would like to thank my family: Ifu, Chioma, Aunty Odera, Uzo, Uzua, Ife, the Uzomas, Mummy, and many others around the world. I am thankful for you all and I really appreciate all the faith you have in me. I hope I continue making you proud and that I have started on a path that many of us will follow.

v

Finally, I would like to thank everyone I have positively interacted with during my time at

UCSF for all the parts they played in helping me to thrive and actualize my full potential.

CONTRIBUTIONS

Some chapters in this dissertation contain material that is currently under consideration/review. They do not necessarily represent the final published form.

Chapter 3 of this dissertation is a reprint of a pre-print for work that is submitted and under review: Belonwu SA, Li Y, Bunis D, Rao AA, Warly Solsberg C, Oskotsky T, Taubes A, Grone B, Zalocusky K, Fragiadakis GK, Huang Y, Sirota M. Single-cell transcriptomic analysis elucidates APOE genotype-specific changes across cell types in two brain regions in Alzheimer's disease. Preprint available at: https://www.researchsquare.com/article/rs-291648/v1

Author Contributions

SB and MS conceived the study. SB designed the study and performed data analysis and interpretation of results (sn-RNAseq differential gene expression analysis and pathway functional enrichment), generated figures, and drafted the manuscript. DB assisted in developing methods and figures for the analysis and drafted the manuscript. AAR assisted in developing methods for data analysis. YL performed network analysis, generated figures for the network analysis, and drafted the manuscript. CWS, TO, AT, BG, KZ, GKF, and YH contributed to the discussion of methods and results as well as the implications of the findings. MS oversaw the study. All authors read and contributed to the final manuscript. The funders had no role in the design, implementation, or preparation of this manuscript.

Graduate Research Fellowship Program under Grant No. 1650113. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Chapter 4 of this dissertation is a reprint of a pre-print for work that is submitted and under review: Belonwu S*, Li Y*, Bunis D, Rao AA, Warly Solsberg C, Tang A, Fragiadakis GK, Dubal D, Oskotsky T, Sirota M. Sex-stratified single-cell RNA-Seq analysis identifies sex-specific and cell type-specific transcriptional responses in Alzheimer's disease across two brain regions. *equal contributors. Preprint available at: https://www.researchsquare.com/article/rs-418653/v1

Author Contributions

SB, YL, and MS conceived the study. SB performed data analysis and interpretation of results (data wrangling and pre-processing, cell type-identification, batch correction, snRNAseq differential gene expression analysis, and pathway functional enrichment), generated and compiled figures, and drafted the manuscript. YL generated figures and interpreted results for the enriched pathway network analysis and drafted the manuscript. DB assisted in developing methods and figures for the analysis and drafted the manuscript. AAR assisted in developing

methods for data analysis. CWS, AT, GF, TO, and DD contributed to the discussion of methods and results as well as the implications of the findings. MS oversaw the study. All authors read and contributed to the final manuscript. The funders had no role in the design, implementation, or preparation of this manuscript.

Competing Interests

MS is on the advisory board of twoXAR. Other authors declare no competing financial interests.

Hippos don't forget.
-Shannon L. Giguere

ABSTRACT

Precision Medicine Approach to Elucidate *APOE* Genotype and Sex-Specific Transcriptomic Changes in Alzheimer's Disease

Stella Belonwu

Alzheimer's Disease (AD) is a complex neurodegenerative disease resulting from both environmental and genetic risk factors that accounts for most dementia cases. AD is one of the top causes of death in the United States, and it is increasing in prevalence along with the growing aging population. AD has no cure, and there are limited therapeutics available to reverse it. Thus, it is of high priority to understand the underlying molecular mechanisms associated with AD progression to gain insight of ways to target it. Among the risk factors of AD are the apolipoprotein E (*APOE*) ε4 allele and sex. APOE4 has been identified as the largest genetic risk factor for AD, yet its molecular underpinnings are obscure. Additionally, sex differences have also been clinically documented in AD, but the molecular mechanisms explaining these differences remain elusive. To understand how *APOE* genotype and sex contribute to differing vulnerabilities in AD, we leverage publicly available transcriptomic datasets to explore *APOE* genotype-specific disease-related changes on a bulk and single-cell level, and sex-specific disease-related changes on a single-cell level.

In Chapter 1, we leverage human AD bulk RNA-sequencing (RNA-Seq) datasets spanning 7 brain regions (temporal cortex, cerebellum, dorsolateral prefrontal cortex, anterior prefrontal cortex, posterior superior temporal gyrus or Wernicke's area, perirhinal cortex, and inferior frontal gyrus or Broca's area) containing 494 AD and 262 non-demented controls. We performed a case versus control APOE4-stratified analysis in each brain region separately. We split each

dataset into APOE4-negative ("E4NEG": APOE3/3 (homozygous for allele ε3)) and APOE4-positive ("E4POS": APOE3/4 (heterozygous ε3/ε4) and APOE4/4 (homozygous for allele ε4)) samples, compared AD to control samples within each subgroup, and examined disease-related gene expression and pathway changes in each subgroup. We identified new and previously studied transcriptomic changes based on the presence of APOE4 with some overlap across brain regions. While we observed neuroinflammatory pathways in all samples, we also observed an emphasis on stress response, hormone and receptor signaling, and epigenetic regulation in E4NEG samples, and an emphasis on metabolic changes, lipid metabolism, clearance and recovery from deleterious events, and ion, iron, and vitamin homeostasis in E4POS samples.

In the last decade, decreased costs and advances in high-throughput sequencing (HTS) technology and analytics tools have transformed the scientific field. With HTS, scientists have been able to generate bigger genomic datasets with unbiased insights in a timely fashion. In chapter 1, we took advantage of this by leveraging publicly available human AD bulk RNA-Seq datasets. However, these were traditional or "bulk" RNA-Seq datasets, which, although they provided enormous insights into the biological changes in AD, these insights were of the average biological changes, and as a result could not reveal more complex changes that explain the heterogeneity of AD. Fortunately, single-cell RNA-Seq emerged to identify and assess subpopulations of cells otherwise considered to be homogeneous. In chapters 2 and 3, we leverage publicly available human AD single-cell RNA-Seq datasets to increase our resolution and determine brain cell type vulnerabilities in AD with regards to *APOE* genotype and sex.

Following our bulk analysis, to elucidate more complex *APOE* genotype-specific disease-relevant changes masked by the bulk analysis, we leverage the first two single-nucleus RNA-Seq (snRNA-Seq) AD datasets from human brain samples, including nearly 55,000 cells from the prefrontal and entorhinal cortices (Chapter 2). We performed a case vs control *APOE* genotype-stratified differential gene expression (DGE) analysis and pathway network enrichment in astrocytes, microglia, neurons, oligodendrocytes, and oligodendrocyte progenitor cells of APOE3/3 and APOE3/4 samples. We observed more global transcriptomic changes in E4POS AD cells and identified differences across *APOE* genotypes primarily in glial cell types.

Next, to explore sex differences in AD, we also leveraged the same two snRNA-Seq datasets and utilized nearly 74,000 cells from human prefrontal and entorhinal cortex samples (Chapter 3). We performed a case vs control sex-stratified DGE analysis and pathway network enrichment in a cell type-specific manner like Chapter 2. In the prefrontal cortex, we observed sex-specific gene and pathway differences in AD most prominently in glial cells, and in the entorhinal cortex, we observed shared genes and pathways to be perturbed in opposing directions between sexes in AD relative to healthy state.

Ultimately, this dissertation identifies disease-relevant transcriptomic perturbations on a bulk and single-cell level that suggest differing mechanisms of neurodegeneration based on *APOE* genotype and sex. The findings here highlight the importance of incorporating *APOE* genotype and sex in future multiomic exploration of AD pathogenesis and progression, and will have implications for precision medicine approaches in the diagnosis and treatment of AD.

TABLE OF CONTENTS

CHAPTER 1: Identification of *APOE* genotype-specific gene signatures and enriched pathways in Alzheimer's Disease across seven brain regions

CHAPTER 2: Stratified analysis of single cell transcriptomic Alzheimer's disease data elucidates APOE genotype mediated changes across cell types in two brain regions

CHAPTER 3: Sex-stratified single-cell RNA-Seq analysis identifies sex-specific and cell type-specific transcriptional responses in Alzheimer's disease across two brain regions.

LIST OF FIGURES

LIST OF TABLES

SUPPLEMENTARY MATERIALS

Supplementary Table 1.1: MayoRNAseq Temporal Cortex Differentially Expressed Genes

Supplementary Table 1.2: MayoRNAseq Cerebellum Differentially Expressed Genes

Supplementary Table 1.3: ROSMAP Dorsolateral Prefrontal Cortex Differentially Expressed
Genes

Supplementary Table 1.4: MSBB Anterior prefrontal cortex (Brodmann Area 10) Differentially
Expressed Genes

Supplementary Table 1.5: MSBB Posterior superior temporal gyrus/
Wernicke's area (Brodmann Area 22) Differentially Expressed Genes

Supplementary Table 1.6: MSBB Perirhinal cortex (Brodmann Area 36) Differentially Expressed
Genes

Supplementary Table 1.7: MSBB Inferior frontal gyrus/ Broca's area (Brodmann Area 44)
Differentially Expressed Genes

Supplementary Table 1.8: MayoRNAseq Temporal Cortex Gene Ontology

Supplementary Table 1.9: MayoRNAseq Cerebellum Gene Ontology

Supplementary Table 1.10: ROSMAP Dorsolateral Prefrontal Cortex Gene Ontology

Supplementary Table 1.11: MSBB Anterior prefrontal cortex (Brodmann Area 10) Gene
Ontology

Supplementary Table 1.12: MSBB Posterior superior temporal gyrus/
Wernicke's area (Brodmann Area 22) Gene Ontology

Supplementary Table 1.13: MSBB Perirhinal cortex (Brodmann Area 36) Gene Ontology

Supplementary Table 1.14: MSBB Inferior frontal gyrus/ Broca's area (Brodmann Area 44)
Gene Ontology

CHAPTER 1

Identification of *APOE* genotype-specific gene signatures and enriched pathways in Alzheimer's Disease across seven brain regions

1.1 BACKGROUND AND SIGNIFICANCE

Alzheimer's Disease (AD) is a growing neurodegenerative disease with no cure, which results from a combination of genetic and environmental factors. AD is one of the top leading causes of death in the United States, and it imposes a high burden on patients and caregivers. As AD is set to increase in prevalence along with the growing aging population, it is of high priority to study and understand the mechanisms associated with its progression to give insight into ways to target it. Apolipoprotein E (APOE) is a lipid-binding protein involved in the metabolism and transport of lipids that is highly expressed in the brain, primarily in astrocytes[1,2]. The ε4 subtype of *APOE* is the greatest known genetic risk factor for developing late-onset AD[1]. Thus far, scientists have observed the dose-dependent way APOE4 increases risk for AD[3]; however, it is still unclear how it may affect pathogenesis on a transcriptomic level. To understand how *APOE* genotype influences transcriptional profiles in AD pathogenesis, we analyzed publicly available bulk human RNA-sequencing datasets across 7 brain regions from 494 AD and 262 non-demented controls. We performed a case-control *APOE* genotype-stratified analysis focusing on the individuals with and without the ε4 allele of *APOE* in each brain region. In our study, we identified shared and distinct disease-relevant differentially expressed genes and pathways in each brain region, implying different mechanisms of AD pathogenesis related to the presence of the ε4 allele of *APOE*. Our findings highlight the importance of studying disease-related changes in AD based on *APOE* genotype and present approaches to studying heterogeneity and disease progression in AD.

1.2 RESULTS AND DISCUSSION

AD is a pervasive neurodegenerative disease with no cure, which is caused by a combination of genetic and environmental factors. Among its risk factors is the ε4 allele of the apolipoprotein E (*APOE*) gene[1,3]. APOE4 is the largest known genetic risk factor for AD; however, the exact mechanism at which it facilitates AD pathogenesis remains elusive. The increasing availability of AD transcriptomic datasets and analytical tools provide worthwhile means to explore these potential mechanisms. In this study, we leveraged bulk RNA-sequencing (RNA-Seq) datasets across 7 brain regions (temporal cortex, cerebellum, dorsolateral prefrontal cortex, anterior prefrontal cortex, posterior superior temporal gyrus or Wernicke's area, perirhinal cortex, and inferior frontal gyrus or Broca's area) containing 494 AD and 262 non-demented controls (**Table 1.1, Figure 1.1**). We performed an APOE4-stratified analysis, where we split each dataset into APOE4-negative ("E4NEG": APOE3/3) and APOE4-positive ("E4POS": APOE3/4 and APOE4/4) samples. We examined disease-related changes in each of these groups by comparing cases to controls and used DEGs we identified to perform gene ontology (GO) and pathway enrichment.

1.2.1 APOE4-stratified DGE analysis identifies distinct AD-related transcriptomic changes in each brain region

In our gene-level analysis, we visualized the correlation of samples used in each analysis to each other using hierarchical clustering. From this, we observed some clustering by *APOE* genotype in addition to diagnosis and sex (**Fig. 1.2-1.8a**). Additionally, while we observed shared DEGs by E4NEG and E4POS samples, we observed some DEGs to be uniquely differentially expressed (**Fig. 1.2-1.8b**) and an overall difference in disease-related gene expression changes (**Fig. 1.2-**

1.8c). To consolidate novel and previously identified AD genes in our DGE analysis, we performed functional enrichment for lists of DEGs that increased and decreased in expression for each APOE4 group and brain region. We identified several shared and unique enriched GO and pathway frameworks in all brain regions (**Fig. 1.2-1.8d, Fig. 1.2-1.7e**).

1.2.2 Temporal Cortex

The temporal cortex (TCX) is a part of the temporal lobe, which is associated with auditory and visual processing, recognition, and memory acquisition[4–6]. In the TCX, some DEGs we identified specific to E4NEG samples included AD candidate gene *MS4A3*[7], a member of the *MS4A* gene cluster whose expression has been demonstrated to correlate with *TREM2* expression and suggested to modulate AD risk[8], and *CRH*, a mediator of neuroinflammation[9], both of which decreased in expression in AD (**Fig. 1.2d, Supplementary Table 1.1**). Additionally, *MMP10*, part of the matrix metalloproteinase family, which are expressed across brain cell types and have been studied as AD biomarkers, was uniquely increased in expression in AD for E4NEG samples[10]. In the E4POS samples, we observed *UPK1B*, which encodes a potential regulator of normal bladder epithelium structure and function[11] and has not been previously associated with AD, and *CXCL8*, which encodes a chemokine that has been associated with AD[12], to be uniquely increased in expression. In our functional enrichment analysis, we uniquely observed GO frameworks from DEGs with decreased expression related to receptor ligand activity, neuropeptide hormone activity and myelin sheath structure maintenance, and DEGs with increased expression related to transmembrane receptor protein kinase activity, immune receptor activity, and transporter activity in the E4NEG population (**Fig. 1.2e (left), Supplementary Table 1.8**). In the E4POS population, we uniquely observed GO frameworks from DEGs with increased expression related to protease

binding, platelet derived growth factor binding, phospholipase inhibitor activity, and cell adhesion mediator activity. Additionally, through pathway analysis, from DEGs with decreased expression, we observed neuroactive ligand-receptor interaction in E4NEG samples in addition to cytokine receptor interactions and TGF-beta signaling from DEGs with increased expression (**Fig. 1.2e (right), Supplementary Table 1.15**). Interestingly, we found protein digestion and absorption to be enriched from DEGs with decreased expression in E4NEG samples but enriched from DEGs with increased expression in E4POS samples. In E4POS samples, from DEGs with increased expression, we observed pathways such as focal adhesion, NF-kappa-B signaling pathway, and AGE-RAGE signaling. Overall, while we observed inflammatory pathways in both APOE4 groups, E4NEG samples demonstrated an emphasis on hormone and receptor activity compared to cell adhesion, oxidative stress, and phospholipase regulation in E4POS samples.

1.2.3 Cerebellum

The cerebellum (CBE) is a portion of the hindbrain involved in motor control and supporting cognitive functions not limited to learning, emotional control, and short term memory[13]. In the CBE, some DEGs we identified with increased expression and specific to E4NEG samples included heat shock chaperones encoded by *HSPA6*, *HSPA1A* and *HSPA1B*, and several genes encoding histones, which may be indicative of epigenetic alterations observed in AD[14] (**Fig. 1.3d, Supplementary Table 1.2**). In E4POS samples, some DEGs we uniquely identified with increased expression include *DES*, which is known for maintaining muscle cell structural integrity, has not been previously studied in AD, but has been extensively in heart failure[15], and *SRPX2*, which also has not been previously studied in the context of AD, but has been in developmental disorders and speech[16]. From GO enrichment, we did not identify any shared GO terms across APOE4 groups

(**Fig. 1.3e (left), Supplementary Table 1.9**). We observed terms enriched from DEGs with increased expression in E4NEG samples involving unfolded protein binding and chaperone binding, and a term enriched from DEGs with decreased expression involving receptor regulator activity. In E4POS samples, we observed extracellular matrix structural organization to be enriched from DEGs with increased expression, and activities enriched from DEGs with decreased expression involving transmembrane transporter activity, hydrolase, kinase, and phosphatase activity. From pathway enrichment, we observed shared pathways (e.g., viral carcinogenesis, neutrophil extracellular trap formation) across APOE4 groups which were enriched from DEGs with increased expression in E4NEG samples and enriched from DEGs with decreased expression in E4POS samples (**Fig. 1.3e (right), Supplementary Table 1.16**). Overall, in the CBE, we observed disease-related changes to involve stress response and potential epigenetic influence in E4NEG samples, and dysregulation in transporter, kinase, and fatty acid breakdown in E4POS samples.

1.2.4 Dorsolateral prefrontal cortex

The dorsolateral prefrontal cortex (DLPFC) is a part of the prefrontal cortex involved in higher order cognitive and sensory processes like pain, hunger, reward-related decision making, and maintenance of attention and working memory[17,18]. In the DLPFC, some DEGs we identified specific to E4NEG included *FREM3*, a gene important for endocytic recycling, which is associated with aging, major depressive disorder, and related symptoms such as insomnia and slower perception[19,20], and similar to the TCX, the neuroinflammation mediator *CRH*[9] (**Fig. 1.4d, Supplementary Table 1.3**). Both DEGs were decreased in expression in AD samples. In E4NEG samples, some DEGs with increased expression included *PF4*, which encodes a chemokine and

platelet factor involved in platelet aggregation and inflammatory processes that has been associated with heparin-induced thrombocytopenia and briefly studied in the context of vascular pathology in AD[21,22] , and *MORC1*, which encodes an epigenetic regulator studied in relation to stress and major depressive disorder[23]. In the E4POS samples, we observed increased expression of *MT1H*, a member of the metallothionein family and tumor suppressor suggested to be involved in vulnerability to AD pathogenesis[24]. From our GO enrichment, we did not identify any shared GO terms across APOE4 groups (**Fig. 1.4e (left), Supplementary Table 1.10**). In E4NEG samples, like the TCX, we identified GO frameworks from DEGs with decreased expression involving receptor ligand activity and neuropeptide hormone activity. In the E4POS samples, we observed GO frameworks from DEGs with decreased expression in AD related to vitamin and sulfate binding, and lyase activity. Additionally, we observed GO frameworks from DEGs with increased expression in AD related to transmembrane transporter activity. From pathway enrichment, like the TCX, we observed neuroactive ligand receptor interaction to be uniquely enriched from DEGs with decreased expression in E4NEG samples in AD (**Fig. 1.4e (right), Supplementary Table 1.17**). For E4POS samples, we observed pathways from DEGs with decreased expression in AD related to amino acid biosynthesis as well as carbon and amino acid metabolism, and pathways from DEGs with increased expression in AD related to glycosaminoglycan biosynthesis of keratan sulfate and the synaptic vesicle cycle. Overall, we observed a few novel genes not previously studied in AD to be unique in each APOE4 group. In E4NEG samples, like the CBE, we observed disease-related changes to involve stress response and epigenetic regulation, and like the TCX, we observed decreased receptor ligand and hormone activity. In E4POS samples, we observed disease-related changes involving decreased vitamin binding and metabolic activities, and increased neuronal regeneration mechanisms[25].

1.2.5 Anterior prefrontal cortex

The anterior prefrontal cortex (Brodmann Area (BA)10) is a region of the frontal lobe involved in risk-reward decision making, perceiving odors, working memory, and pain[26,27]. In BA 10, some DEGs we identified specific to E4NEG samples included *HAMP*, which encodes iron regulator hepcidin, that has been associated with AD progression and other neurodegenerative diseases though iron-induced oxidative damage[28,29], *APOL4*, which has not been extensively studied, but so far has been associated with schizophrenia[30], and *CSF3*, which encodes a cytokine and growth factor that facilitates granulocyte production in the blood and is important for maintaining neuronal plasticity[31] (**Fig. 1.5d , Supplementary Table 1.4**). These were all decreased in expression in E4NEG samples in AD. In E4POS samples, we observed increased expression of *TCEA3*, which encodes an apoptosis promoter primarily studied in cancer, and bitter taste receptor gene *TAS2R31*, which may be indicative of olfactory fluctuations reported in old age and the onset of neurologic disorders[32]. We did not identify any shared terms across APOE4 groups in our GO (**Fig. 1.5e (left), Supplementary Table 1.11**) and pathway enrichment (**Fig. 1.5e (right), Supplementary Table 1.18**). For E4NEG samples, GO and pathway terms were enriched from DEGs with decreased expression in AD. GO terms involved receptor ligand and cytokine activity, and pathways involved vascular smooth muscle contraction, TNF, NF kappa B and IL-17 signaling pathways. For E4POS samples, GO and pathway terms enriched from DEGs with increased expression in AD were related to taste transduction and GO and pathway terms enriched from DEGs with decreased expression in AD were related to steroid, estradiol and retinol biosynthesis and metabolism. Overall, in E4NEG samples, we observed disease-related changes from signaling and inflammatory activities, and in E4POS samples from alterations in taste perception and metabolic activity.

1.2.6 Posterior superior temporal gyrus (Wernicke's area)

The Wernicke's area (BA 22) is a part of the temporal lobe specialized in cognitive functions such as speech and language comprehension[33]. In BA 22, some DEGs we identified specific to E4NEG included *RASL11B*, a member of the *RAS* family suggested to be involved in inflammation and conditions like cancer and arteriosclerosis[34,35], which was decreased in expression in AD, as well as *TGFB1* and *ILIR2*, cytokines and inflammatory regulators that have been implicated numerous conditions including AD[36–38], both of which were increased in expression in AD (**Fig. 1.6d, Supplementary Table 1.5**). In E4POS samples, we uniquely observed decreased expression of *PVALB*, which encodes a subset of calcium-binding interneurons whose metabolic deficiencies have been explored in relation to autism spectrum disorder, schizophrenia and attention deficit disorder[39], and *OSTN*, which is associated with protection against inflammation, cardiotoxicity, apoptosis, neuronal dysfunction, and oxidative stress[40,41]. In our functional enrichment analysis, we did not identify any shared terms across APOE4 groups in our GO enrichment (**Fig. 1.6e (left), Supplementary Table 1.12**); however, we observed one shared pathway term (insulin secretion) from DEGs with decreased expression (**Fig. 1.6e (right), Supplementary Table 1.19**). In E4NEG samples, we identified GO terms from DEGs with increased expression in AD involving GTPase activity, growth factor binding, intracellular calcium activated chloride channel activity, and calmodulin activity. We also identified unique pathway terms from DEGs with increased expression in AD involving PI3K-Akt signaling pathway, Hippo signaling pathway and complement and coagulation cascades, and from DEGs with decreased expression in AD involving axon guidance. In E4POS samples, we identified unique GO and pathways from DEGs with decreased expression in AD. GO terms primarily included ribonucleoside binding, ion channel activity and transmembrane transporter activity, and pathways included thyroid hormone synthesis

and signaling, calcium signaling, oxidative phosphorylation, and cardiac muscle contraction. Overall, in E4NEG samples, we observed disease-related changes encompassing increased inflammation, cellular proliferation and phosphorylation, and decreased axon guidance. In E4POS samples, in addition to inflammation, we also observed disruptions in calcium signaling metabolism, cardiovascular physiology, ion channel activity, and transmembrane activity.

1.2.7 Perirhinal cortex

The Perirhinal cortex (BA 36) is a section of the medial temporal lobe (MTL) involved in memory formation and retention[42]. As part of the MTL, it undergoes atrophy during early stages of AD progression[43]. In BA 36, some DEGs we identified specific to E4NEG samples included *GRB7*, an oncogenomic driver demonstrated to mediate EGFR/ErbB signaling, which has not been previously studied in AD[44], and *OR2M5*, which encodes an olfactory receptor gene, may imply altered olfactory abilities reported in old age and the onset of neurologic disorders, and may also provide additional evidence of impaired olfactory abilities that are associated with decreased volume of the hippocampus and parahippocampus, which are also part of the MTL[45,46] (**Fig. 1.7d, Supplementary Table 1.6**). These were both decreased in expression in E4NEG samples in AD. In E4POS samples, we observed a unique decrease in expression of *CALY*, a gene that encodes an endosome important for regulating homeostasis of neurons and maintaining synaptic plasticity[47]. While not extensively studied in AD, it may have implications for neurodegeneration. We did not observe any enriched GO (**Fig. 1.7e, Supplementary Table 1.13**) or pathway terms (**Fig. 1.7e, Supplementary Table 1.20**) for E4POS samples. In E4NEG samples enriched terms from DEGs with increased expression involved kinase and immune receptor activity, and Wnt, TNF, TGF, PI3K-Akt, MAPK and NF-kappa-B signaling pathways. Additionally, in E4NEG samples

enriched terms from DEGs with decreased expression involved GABAergic synaptic activity, axon guidance, oxidative phosphorylation, calcium signaling, ion channel activity, and neurotransmitter receptor signaling. Overall, in E4NEG samples, disease-related changes involve increased immune, kinase, and inflammatory processes as well as decreased metabolism, synaptic activity and receptor signaling. E4POS samples, on the other hand involved processes not limited to reduced synaptic plasticity and clearance and recovery from deleterious events.

1.2.8 Inferior frontal gyrus (Broca's area)

The Broca's area (BA 44) is a part of the frontal lobe specialized in cognitive processes related to speech production[48]. In BA 44, in addition to a few taste receptor genes, one of the DEGs we identified specific to E4NEG DEGs meeting our fold change and significance cutoffs with increased expression was *HSPB7*, which encodes chaperones expressed in cardiomyocytes that are important for heart development and maintenance of cardiac muscle integrity[49,50] (**Fig. 1.8d, Supplementary Table 1.7**). In E4POS samples, one of the DEG we uniquely identified with increased expression included *STEAP4*, which encodes a protein important for maintaining iron and copper homeostasis, metabolic functions, and inflammatory responses[51,52]. Additionally, some DEGs we uniquely identified in E4POS samples with decreased expression included transmembrane protein encoding gene *TMEM160*, and *PDXP*, a key regulator of vitamin B6 synthesis and the PI3K/AKT/ERK1/2 pathway[53,54]. We only identified functional enrichment terms meeting our significance thresholds for GO in E4NEG samples (**Fig. 1.8e, Supplementary Table 1.21**). These included GO terms from DEGs with increased expression for bitter taste receptor activity, and from DEGs with decreased expression for copper ion binding and fibroblast growth factor binding. Overall, in BA 44, we observed the least overlaps of DEGs across APOE4

groups (**Fig. 1.8b**) and only observed enriched functions in E4NEG samples. We observed disease-related changes related to stress responses, cardiac muscle physiology and taste transduction in E4NEG samples, and in E4POS samples, we observed disease-related changes related to metal ion homeostasis, transmembrane transport, metabolism, and inflammation.

1.2.9 Comparative analysis across brain regions demonstrates APOE4 influence on AD-related transcriptomic changes

Thus far, we observed shared and distinct transcriptional changes that demonstrate potential differences in AD pathophysiology in populations with and without APOE4. We identified novel as well as previously explored genes in AD. When we performed hierarchical clustering of pseudobulk expression using fold changes of genes present in our analyses by APOE4 group and brain region, we observed some clustering by the presence of APOE4 (**Fig. 1.9a**). Additionally, through clustering enriched GO (**Fig. 1.9b**) and pathway terms (**Fig. 1.9c**) using the corresponding directions of DEG expression changes, we also observed some clustering by APOE4 presence and brain region.

1.2.10 Limitations and Future Directions

With examining APOE4-specific disease related changes across 7 brain regions, there are some limitations to note. First, we encountered batch effects in five out of our seven datasets. To mitigate this limitation, we performed batch correction (**Fig. 1.10, Fig. 1.11, Fig. 1.12**) using the limma package[55], and we included batch as the first covariate in all differential gene expression analyses. Next, we were also limited by the distribution of samples across *APOE* genotypes and diagnosis. There were more samples in the E4NEG group, and there were fewer controls compared to cases

especially in the E4POS group. All datasets also contained samples primarily from individuals of Caucasian descent, which limits translating findings to other populations. Additionally, with using publicly available datasets, samples were processed and annotated differently. To include some uniformity across datasets we sought to identify cases and controls based on similar criteria. We also did not stratify our analysis by age or disease severity, so we could not infer disease related transcriptomic changes based on them. Finally, bulk tissue transcriptomics masks more complex disease-related transcriptomic changes in AD.

We hope that more types of transcriptomic datasets become available from diverse populations across multiple brain regions with more *APOE* genotypes. With the availability of more robust datasets some potential avenues to explore include: 1) a meta-analysis of multiple datasets from the same brain regions to provide more power to assess brain region-specific changes and study the spread of pathology across the brain, 2) a validation of key genes and pathways with other transcriptomic studies and eventually *in vitro* and *in vivo* experimentation, and 3) data integration with other multiomic approaches. Additionally, to go beyond the resolution of bulk datasets, single-cell AD datasets would also present a worthwhile means to explore heterogeneity, to identify cell type-specific changes in AD across *APOE* genotypes, and to help elucidate potential mechanisms of AD progression based on *APOE* genotype.

1.3 METHODS

1.3.1 Materials Availability

This study did not generate new unique reagents.

1.3.2 Data and Code Availability

We obtained publicly available bulk RNA-Seq datasets from the Accelerating Medicines Partnership-Alzheimer's Disease Portal (AMP-AD) under The RNAseq Harmonization Study (rnaSeqReprocessing) found at https://www.synapse.org/#!Synapse:syn9702085. The rnaSeqReprocessing study created uniformly processed RNAseq datasets from donors from three studies across 7 brain regions (**Table 1.1**). We downloaded all datasets on August 12, 2019. The MayoRNAseq study comprised samples from the temporal cortex and cerebellum (counts: https://www.synapse.org/#!Synapse:syn8690799; covariates: https://www.synapse.org/#!Synapse:syn3817650). The Religious Orders Study and Memory and Aging Project (ROSMAP) Study comprised samples from the dorsolateral prefrontal cortex (counts: https://www.synapse.org/#!Synapse:syn8691134 ; covariates: https://www.synapse.org/#!Synapse:syn3157322). The Mount Sinai Brain Bank (MSBB) study comprised data from Brodmann areas 10 (BA 10), BA 22, BA 36, and BA 44, which are the anterior prefrontal cortex, posterior superior temporal gyrus or Wernicke's area, perirhinal cortex, and inferior frontal gyrus or Broca's area, respectively (counts: https://www.synapse.org/#!Synapse:syn8691099; covariates: https://www.synapse.org/#!Synapse:syn7392158). Access to the dorsolateral prefrontal cortex dataset requires a formal request to ROSMAP. To enable other researchers to explore these datasets, all code necessary for recreating the reported analyses and figures within R from our analysis are available on Github at https://github.com/stebel5/AD_APOE_bulkRNAseq.

1.3.3 Identification of Study Cohorts

MayoRNAseq

We acquired publicly available datasets from the temporal cortex (TCX) and cerebellum (CBE) from their respective repositories. These datasets included transcriptome data for 275 CBE and 276 TCX samples retrieved from 312 Caucasian individuals with AD, progressive supranuclear palsy (PSP), pathologic aging (PA) or elderly controls (CON), who had no neurodegenerative diseases. Subjects were diagnosed based on NINCDS-ADRDA criteria[56], which measure tau neurofibrillary tangles using Braak stage[57] and neuritic plaque density using Consortium to Establish a Registry for Alzheimer's Disease (CERAD) scores[58]. All AD subjects had confirmed diagnoses based on NINCDS-ADRDA criteria and had Braak stages $\geq$ IV, and control subjects had Braak stages $\leq$ III, CERAD scores of 0 (none) or 1 (sparse) and lacked pathological diagnoses of neurodegenerative diseases. We selected AD and control samples, removed samples with missing data and low RNA integrity, and selected samples with APOE3/3 (homozygous for allele ε3), APOE3/4 (heterozygous ε3/ε4), and APOE4/4 (homozygous for allele ε4) genotypes. After filtering, there were 78 AD and 65 control TCX samples, and 78AD and 66 control CBE samples (**Table 1.1**).

ROSMAP

We acquired publicly available datasets from the dorsolateral prefrontal cortex (DLPFC) from its repository. We selected cases and controls using similar criteria as Wan et al[59]. Referring to the Rush Alzheimer's Disease Center clinical codebook, cases were defined as samples with Braak stages $\geq$ IV, CERAD scores $\leq$ 2, and final consensus cognitive diagnosis (cogdx) of 4 (AD: Alzheimer's dementia and NO other cause of CI (NINCDS PROB AD)). Controls were defined

as samples with Braak stages ≤ III, CERAD scores ≥ 3, and cogdx of 1 (NCI: No cognitive impairment (No impaired domains)). We removed duplicated samples, samples with missing data and samples with low RNA integrity. We selected samples with APOE3/3 (homozygous for allele ε3), APOE3/4 (heterozygous ε3/ε4), and APOE4/4 (homozygous for allele ε4) genotypes. After filtering, there were 144 AD and 70 control DLPFC samples (**Table 1.1**).

MSBB

We acquired publicly available datasets from Brodmann Area 10 (BA10): Anterior prefrontal cortex, BA 22: Posterior Superior Temporal Gyrus/ Wernicke's area, BA 36: Perirhinal cortex, and BA 44: Inferior Frontal Gyrus/ Broca's area from their respective repositories. We selected cases and controls using similar criteria as Wan et al[59] . Note, the definitions for CERAD score differ between ROSMAP and MSBB studies. Cases were defined as samples with Braak stages ≥ IV, NP.1(Neuropathology Category as measured by CERAD (1=Normal, 2=Definite AD, 3=probable AD, 4=possible AD)) ≥ 2, and Clinical Dementia Rating (CDR) ≥ 1. Controls were defined as samples with Braak stages ≤ III, CERAD ≤ 1, and CDR ≤ 0.5. We removed samples with missing data, samples with relatively large reads mapped to rRNA regions over total reads measured as rRNA rate >5%, samples with QC actions listed as "Remap" or "Exclude," and samples with low RNA integrity. We selected samples with APOE3/3 (homozygous for allele ε3), APOE3/4 (heterozygous ε3/ε4), and APOE4/4 (homozygous for allele ε4) genotypes. Overall, we had a total of 255 samples from 93 individuals encompassing 4 brain regions. After splitting samples by their respective Brodmann areas, we removed duplicate samples by using the latest or resequenced samples when possible. After filtering, we had 59 AD and 19 control

BA 10 samples, 52 AD and 13 control BA 22 samples, 36 AD and 19 control BA 36 samples, and 47 AD and 10 control BA 44 samples (**Table 1.1**).

1.3.4 Quality Control: Batch Correction and Outlier Removal

All data processing was conducted separately for each dataset with R[60] version 4.0.3 (2020-10-10) using RStudio[61]. We performed differential gene expression (DGE) analysis using the DESeq2 (v1.30.0) package[62] and generated visualizations using ggplot2[63], pheatmap[64], and UpsetR[65]. We followed the DESeq2 RNA-Seq workflow steps as outlined by Love et al[66]. We created a DESeq2 *dds* object using the *DESeqDataSetFromMatrix* function from matching count matrices and sample covariate datasets, and a design formula containing with batch, diagnosis, sex, age, and APOE4 status (E4 negative ("E4NEG": APOE3/3) and E4 positive ("E4POS: APOE3/4, APOE4/4)). We filtered out non-protein-coding genes and genes with low counts by removing rows with less than a total of 10 reads. Next, we set our factor levels to have controls as a reference level, ran the *DESeq*[62] function on our model, and selected rows that converged.

To visualize our data and assess potential confounding variables, we transformed the *dds* object using a variance stabilizing transformation blinded by sample information. We extracted transformed values using the assay function, applied the *prcomp* function to the extracted matrices to perform principal components analysis (PCA), and visualized PC1 and PC2 for all variables in our design formula in a scatter plot (**Fig. 1.10, Fig. 1.11, Fig. 1.12**). We also computed pairwise correlations of the extracted transformed values using the *cor* base R function and generated clustering heatmaps of samples by samples, clustered by the variables in our design formula. We observed and corrected batch effects in the DLPFC, BA 10, BA 22, BA 36, and BA 44 datasets using the *removeBatchEffect* function from the limma package[55]. We confirmed successful batch correction using PCA and hierarchical clustering and included batch as the first covariate in the

design formula to account for batch in the differential gene expression (DGE) analysis (**Fig. 1.10c-d, Fig. 1.11, Fig. 1.12**). We calculated outliers as samples with PC1 or PC2 z-scores more than 3 standard deviations from the mean of their respective PC1 or PC2 scores. We removed outliers from the DLPFC dataset: one male and one female APOE3/3 AD sample, and one female APOE3/4 AD sample.

1.3.5 APOE4-stratified DGE and Pathway Enrichment Analysis

After *DESeq*[62] was run on the *dds* object and samples and genes were filtered based on QC criteria, we split each *dds* object by APOE4 categories "E4NEG" and "E4POS". We ran a Wald test for each APOE4-split *dds* object using the results function with AD and control as contrast arguments and a Benjamini Hochberg (BH) p-value correction method. To shrink effect size and improve log2 fold change (LFC) estimates, we passed the *dds* objects into the *lfcShrink* function using the *apeglm* method[67]. We selected differentially expressed genes (DEGs) using LFCs > 0.4 and false discovery rates (FDR) < 0.05 (**Table 1.2, Supplementary Tables 1.1-1.7**).

Next, we performed an overrepresentation analysis of DEGs from the APOE4-stratified analysis of samples in 7 brain regions using the *clusterprofiler* package. We queried DEGs split by upregulated and downregulated expression, to perform gene ontology (**Supplementary Tables 1.8-1.14**) and pathway analysis (**Supplementary Tables 1.15-1.21**) using *enrichGO* and *enrichKEGG* functions, respectively. We used all protein coding genes surveyed in the DGE analysis as the universe and selected enriched biological terms and pathways with an FDR < 0.05.

1.4 REFERENCES

1.	Long, J. M. & Holtzman, D. M. Alzheimer Disease: An Update on Pathobiology and Treatment Strategies. Cell 179, 312–339 (2019).

2.	Yu, J.-T., Tan, L. & Hardy, J. Apolipoprotein E in Alzheimer's Disease: An Update. Annual Review of Neuroscience 37, 79–100 (2014).

3.	Karch, C. M. & Goate, A. M. Alzheimer's disease risk genes and mechanisms of disease pathogenesis. Biological Psychiatry 77, 43–51 (2015).

4.	Martin, A. & Chao, L. L. Semantic memory and the brain: structure and processes. Current Opinion in Neurobiology 11, 194–201 (2001).

5.	Friederici, A. D. The Brain Basis of Language Processing: From Structure to Function. Physiological Reviews 91, 1357–1392 (2011).

6.	Haxby, J. V. et al. Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. Science 293, 2425–2430 (2001).

7.	Nettiksimmons, J., Tranah, G., Evans, D. S., Yokoyama, J. S. & Yaffe, K. Gene-based aggregate SNP associations between candidate AD genes and cognitive decline. Age (Dordr) 38, 41 (2016).

8.	Deming, Y. et al. The MS4A gene cluster is a key modulator of soluble TREM2 and Alzheimer's disease risk. Science Translational Medicine 11, (2019).

9.	Kempuraj, D. et al. Brain and Peripheral Atypical Inflammatory Mediators Potentiate Neuroinflammation and Neurodegeneration. Front. Cell. Neurosci. 11, (2017).

10.	Abe, K. et al. Influence of plasma matrix metalloproteinase levels on longitudinal changes in Alzheimer's disease (AD) biomarkers and cognitive function in patients with mild

cognitive impairment due to AD registered in the Alzheimer's Disease Neuroimaging Initiative database. Journal of the Neurological Sciences 416, 116989 (2020).

11.   UPK1B uroplakin 1B - Gene - GTR - NCBI.

https://www.ncbi.nlm.nih.gov/gtr/genes/7348/.

12.   Stuart, M. J. & Baune, B. T. Chemokines and chemokine receptors in mood disorders, schizophrenia, and cognitive impairment: A systematic review of biomarker studies. Neuroscience & Biobehavioral Reviews 42, 93–115 (2014).

13.   Koziol, L. F. et al. Consensus Paper: The Cerebellum's Role in Movement and Cognition. Cerebellum 13, 151–177 (2014).

14.   Nativio, R. et al. An integrated multi-omics approach identifies epigenetic alterations associated with Alzheimer's disease. Nature Genetics 52, 1024–1035 (2020).

15.   Hnia, K., Ramspacher, C., Vermot, J. & Laporte, J. Desmin in muscle and associated diseases: beyond the structural function. Cell Tissue Res 360, 591–608 (2015).

16.   Schirwani, S., McConnell, V., Willoughby, J. & Balasubramanian, M. Exploring the association between SRPX2 variants and neurodevelopment: How causal is it? Gene 685, 50–54 (2019).

17.   Graham, A. L., Gluck, M. E., Votruba, S. B., Krakoff, J. & Thearle, M. S. Perseveration augments the effects of cognitive restraint on ad libitum food intake in adults seeking weight loss. Appetite 82, 78–84 (2014).

18.   Miller, E. K. & Cohen, J. D. An Integrative Theory of Prefrontal Cortex Function. Annual Review of Neuroscience 24, 167–202 (2001).

19.	Nikolova, Y. et al. FRAS1-related extracellular matrix 3 (FREM3) single-nucleotide polymorphism effects on gene expression, amygdala reactivity and perceptual processing speed: An accelerated aging pathway of depression risk. Front. Psychol. 6, (2015).

20.	Shi, C., Zhang, K., Wang, X., Shen, Y. & Xu, Q. A study of the combined effects of the EHD3 and FREM3 genes in patients with major depressive disorder. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics 159B, 336–342 (2012).

21.	Patriarcheas, V., Pikoulas, A., Kostis, M., Charpidou, A. & Dimakakos, E. Heparin-induced Thrombocytopenia: Pathophysiology, Diagnosis and Management. Cureus 12, e7385 (2020).

22.	Zhang, W., Huang, W. & Jing, F. Contribution of blood platelets to vascular pathology in Alzheimer's disease. J Blood Med 4, 141–147 (2013).

23.	Thomas, M. et al. Investigation of MORC1 DNA methylation as biomarker of early life stress and depressive symptoms. Journal of Psychiatric Research 120, 154–162 (2020).

24.	Miller, J. A., Woltjer, R. L., Goodenbour, J. M., Horvath, S. & Geschwind, D. H. Genes and pathways underlying regional and cell type changes in Alzheimer's disease. Genome Med 5, 48 (2013).

25.	Keratan Sulfate: Biosynthesis, Structures, and Biological Functions | SpringerLink. https://link.springer.com/protocol/10.1007%2F978-1-4939-1714-3_30.

26.	Peng, K., Steele, S. C., Becerra, L. & Borsook, D. Brodmann area 10: Collating, integrating and high level processing of nociception and pain. Progress in Neurobiology 161, 1–22 (2018).

27.	Gilbert, S. J. et al. Functional Specialization within Rostral Prefrontal Cortex (Area 10): A Meta-analysis. Journal of Cognitive Neuroscience 18, 932–948 (2006).

28.   Gemmati, D. et al. Polymorphisms in the genes coding for iron binding and transporting proteins are associated with disability, severity, and early progression in multiple sclerosis. BMC Med Genet 13, 70 (2012).

29.   Urrutia, P. et al. Inflammation alters the expression of DMT1, FPN1 and hepcidin, and it causes iron accumulation in central nervous system cells. Journal of Neurochemistry 126, 541–549 (2013).

30.   Takahashi, S. et al. Association of SNPs and haplotypes in APOL1, 2 and 4 with schizophrenia. Schizophrenia Research 104, 153–164 (2008).

31.   Rajkumar, A. P. et al. Postmortem Cortical Transcriptomics of Lewy Body Dementia Reveal Mitochondrial Dysfunction and Lack of Neuroinflammation. The American Journal of Geriatric Psychiatry 28, 75–86 (2020).

32.   Doty, R. L. Age-Related Deficits in Taste and Smell. Otolaryngologic Clinics of North America 51, 815–825 (2018).

33.   Javed, K., Reddy, V., M Das, J. & Wroten, M. Neuroanatomy, Wernicke Area. in StatPearls (StatPearls Publishing, 2021).

34.   Liu, J. & Li, Y. Upregulation of MAPK10, TUBB2B and RASL11B may contribute to the development of neuroblastoma. Mol Med Rep 20, 3475–3486 (2019).

35.   Stolle, K. et al. Cloning, genomic organization, and tissue-specific expression of the RASL11B gene. Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression 1769, 514–524 (2007).

36.   Niculescu, A. B. et al. Blood biomarkers for memory: toward early detection of risk for Alzheimer disease, pharmacogenomics, and repurposed drugs. Mol Psychiatry 25, 1651–1672 (2020).

37.   Palomo, J., Dietrich, D., Martin, P., Palmer, G. & Gabay, C. The interleukin (IL)-1 cytokine family – Balance between agonists and antagonists in inflammatory diseases. Cytokine 76, 25–37 (2015).

38.   Peters, V. A., Joesting, J. J. & Freund, G. G. IL-1 receptor 2 (IL-1R2) and its role in immune regulation. Brain Behav Immun 32, 1–8 (2013).

39.   Janickova, L. & Schwaller, B. Parvalbumin-Deficiency Accelerates the Age-Dependent ROS Production in Pvalb Neurons in vivo: Link to Neurodevelopmental Disorders. Front. Cell. Neurosci. 14, (2020).

40.   Hu, C. et al. Osteocrin attenuates inflammation, oxidative stress, apoptosis, and cardiac dysfunction in doxorubicin-induced cardiotoxicity. Clinical and Translational Medicine 10, e124 (2020).

41.   Ataman, B. et al. Evolution of Osteocrin as an activity-regulated factor in the primate brain. Nature 539, 242–247 (2016).

42.   Suzuki, W. A. & Naya, Y. The Perirhinal Cortex. Annu. Rev. Neurosci. 37, 39–53 (2014).

43.   Buckner, R. L. et al. Molecular, structural, and functional characterization of Alzheimer's disease: Evidence for a relationship between default activity, amyloid, and memory. Journal of Neuroscience 25, 7709–7717 (2005).

44.   Chu, P.-Y., Tai, Y.-L. & Shen, T.-L. Grb7, a Critical Mediator of EGFR/ErbB Signaling, in Cancer Development and as a Potential Therapeutic Target. Cells 8, 435 (2019).

45.   Kubota, S. et al. Hippocampus and Parahippocampus Volume Reduction Associated With Impaired Olfactory Abilities in Subjects Without Evidence of Cognitive Decline. Front. Hum. Neurosci. 14, (2020).

46.   Buckley, M. J. The Role of the Perirhinal Cortex and Hippocampus in Learning,

Memory, and Perception. The Quarterly Journal of Experimental Psychology Section B 58, 246–

268 (2005).

47.   Muthusamy, N., Chen, Y.-J., Yin, D.-M., Mei, L. & Bergson, C. Complementary roles of

the neuron-enriched endosomal proteins NEEP21 and calcyon in neuronal vesicle trafficking.

Journal of Neurochemistry 132, 20–31 (2015).

48.   Flinker, A. et al. Redefining the role of Broca's area in speech. PNAS 112, 2871–2875

(2015).

49.   Mercer, E. J., Lin, Y.-F., Cohen-Gould, L. & Evans, T. Hspb7 is a cardioprotective

chaperone facilitating sarcomeric proteostasis. Developmental Biology 435, 41–55 (2018).

50.   Liao, W.-C., Juo, L.-Y., Shih, Y.-L., Chen, Y.-H. & Yan, Y.-T. HSPB7 prevents cardiac

conduction system defect through maintaining intercalated disc integrity. PLoS Genet 13, (2017).

51.   Gordon, H. M. et al. STEAP4 expression in human islets is associated with differences in

body mass index, sex, HbA1c, and inflammation. Endocrine 56, 528–537 (2017).

52.   Scarl, R. T., Lawrence, C. M., Gordon, H. M. & Nunemaker, C. S. STEAP4: its emerging

role in metabolism and homeostasis of cellular iron and copper. J Endocrinol 234, R123–R134

(2017).

53.   Li, H. et al. Lactoferrin Induces the Synthesis of Vitamin B6 and Protects HUVEC

Functions by Activating PDXP and the PI3K/AKT/ERK1/2 Pathway. Int J Mol Sci 20, (2019).

54.   Jeanclos, E. et al. Improved cognition, mild anxiety-like behavior and decreased motor

performance in pyridoxal phosphatase-deficient mice. Biochimica et Biophysica Acta (BBA) -

Molecular Basis of Disease 1865, 193–205 (2019).

55. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 43, e47 (2015).

56. McKhann, G. M. et al. The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimer's and Dementia 7, 263–269 (2011).

57. Braak, H. & Braak, E. Staging of alzheimer's disease-related neurofibrillary changes. Neurobiology of Aging 16, 271–278 (1995).

58. Mirra, S. S. et al. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part II. Standardization of the neuropathologic assessment of Alzheimer's disease. Neurology 41, 479–486 (1991).

59. Wan, Y.-W. et al. Meta-Analysis of the Alzheimer's Disease Human Brain Transcriptome and Functional Dissection in Mouse Models. Cell Rep 32, 107908 (2020).

60. R Core Team. R: A Language and Environment for Statistical Computing. https://www.r-project.org/ (2020).

61. R Studio Team. RStudio: Integrated Development Environment for R. https://rstudio.com/ (2020).

62. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15, 550 (2014).

63. Wickham, H. ggplot2: Elegant Graphics for Data Analysis. (Springer-Verlag, 2009). doi:10.1007/978-0-387-98141-3.

64. Kolde, R. pheatmap: Pretty Heatmaps. https://CRAN.R-project.org/package=pheatmap.

65. Conway, J. R., Lex, A. & Gehlenborg, N. UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics 33, 2938–2940 (2017).

66. Love, M. I., Anders, S., Kim, V. & Huber, W. RNA-Seq workflow: gene-level

exploratory analysis and differential expression. F1000Res 4, (2015).

67. Zhu, A., Ibrahim, J. G. & Love, M. I. Heavy-tailed prior distributions for sequence count

data: removing the noise and preserving large differences. Bioinformatics 35, 2084–2092 (2019).

1.5 TABLES

Table 1.1: Bulk RNA-sequencing study cohort characteristics

Study	Brain Region	Total	AD				Control			
			Female		Male		Female		Male	
			E4-	E4+	E4-	E4+	E4-	E4+	E4-	E4+
Mayo Clinic Brain Bank	Temporal cortex	143	19	28	16	15	27	4	30	4
	Cerebellum	144	16	29	19	14	26	5	31	4
Mount Sinai Brain Bank	Anterior prefrontal cortex (BM10)	78	24	19	10	6	5	2	9	3
	Posterior superior temporal gyrus/ Wernicke's area (BM22)	65	22	14	10	6	3	2	7	1
	Perirhinal cortex (BM36)	55	18	8	7	3	5	2	9	3
	Inferior frontal gyrus/ Broca's area (BM44)	57	22	10	9	6	2	1	6	1
Religious Orders Study and Memory and Aging Project	Dorsolateral Prefrontal Cortex	214	62	40	21	21	33	4	29	4

Table 1.2: Number of differentially expressed genes from APOE4 negative and positive samples across brain regions

Brain Region	E4 Negative Upregulated	E4 Negative Downregulated	E4 Positive Upregulated	E4 Positive Downregulated
Temporal cortex	698	527	251	46
Cerebellum	498	443	33	14
Dorsolateral Prefrontal Cortex	66	20	4	2
Anterior prefrontal cortex (BM10)	81	77	2	2
Posterior superior temporal gyrus/ Wernicke's area (BM22)	588	74	23	102
Perirhinal cortex (BM36)	1161	519	12	7
Inferior frontal gyrus/ Broca's area (BM44)	60	20	84	122

1.6 FIGURES
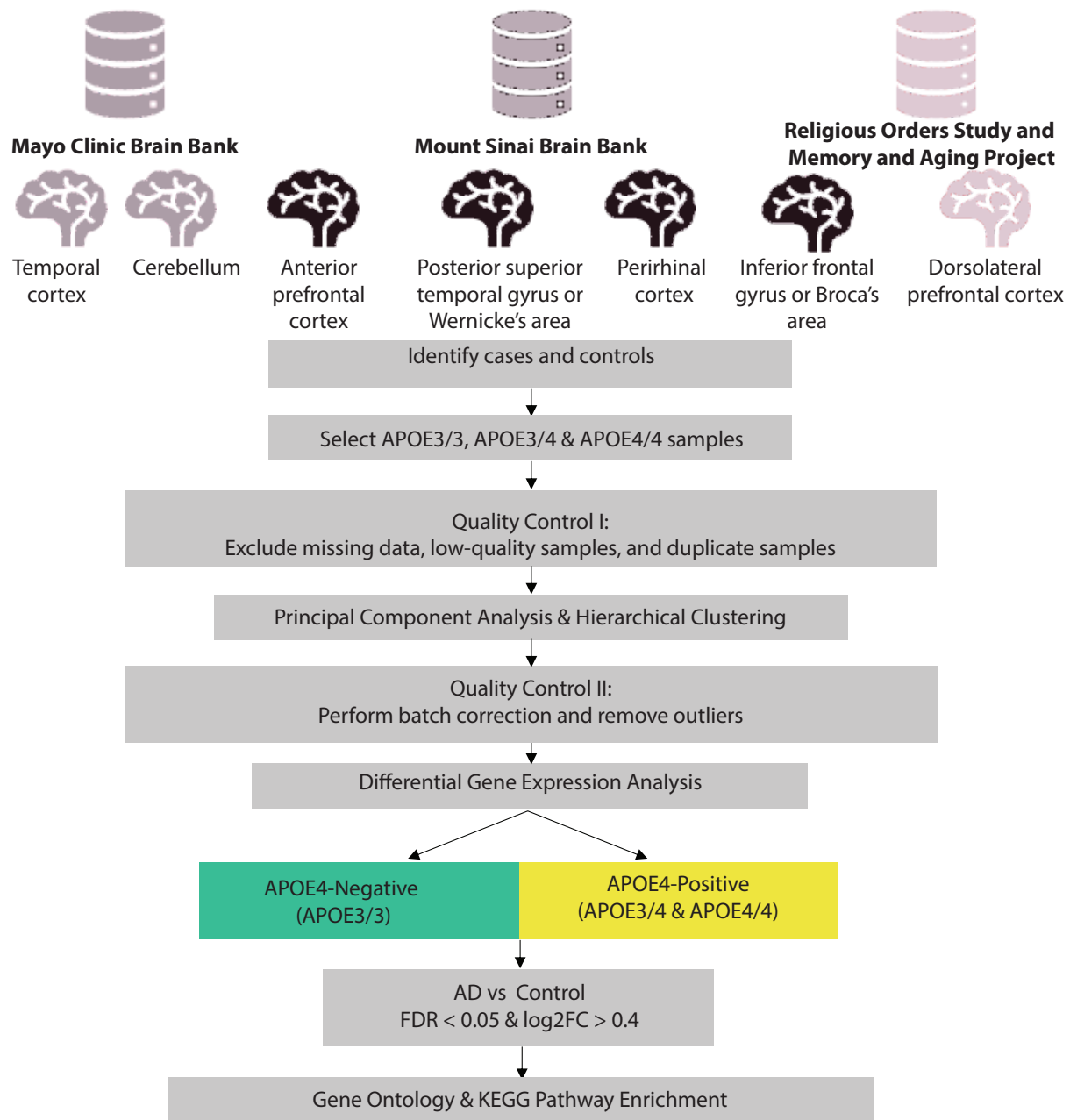
Figure 1.1: Overview of cohort sample definition and workflow for *APOE* genotype-stratified differential gene expression and functional enrichment analysis. Bulk RNA-sequencing data was acquired from 3 databases spanning 7 brain regions. Cases and controls were determined based on tau (Braak) and amyloid β plaque (CERAD) burdens, as well as cognitive diagnosis scores and clinical dementia ratings where relevant. Samples with APOE3/3 (homozygous for allele ε3), APOE3/4 (heterozygous ε3/ε4), and APOE4/4 (homozygous for allele ε4) genotypes were selected. Quality control was performed to remove missing, low quality or duplicate samples. Principal component analysis and hierarchical clustering were performed to assess confounding variables. Batch correction and outlier removal were performed where necessary. Differential gene expression analysis was performed separately for each brain region stratified by APOE4 groupings. Differentially expressed genes (DEGs) were selected using a false discovery rate (FDR) < 0.05 and log2fold change > 0.4. DEGs were passed as inputs for gene ontology and pathway enrichment.

Figure 1.2: Differential gene expression and pathway signatures in the temporal cortex (TCX) . a. Hierarchical clustering of samples by samples, b. Pairwise DEG plots and corresponding upset plots indicating intersections of AD versus non-AD differentially expressed genes (DEGs) (Benjamini-Hochberg (BH) adjusted p-value < 0.05 and absolute log2 fold change (FC) > 0.4) across APOE4 groups. Rows correspond to APOE4 groups. The bar chart shows the number of single and common sets of DEGs across APOE4 groups. Single filled dots represent a unique set of DEGs for the corresponding APOE4 group. Multiple filled black dots connected by vertical lines represent common sets of DEGs across cell types, APOE4 groups, c. log2 FC scores of all genes in the DE analysis, d. Volcano plots of DEGs for both APOE4 groups, e. Gene ontology (left) and KEGG pathway (right) frameworks (BH adjusted p-value < 0.05) for both APOE4 groups with colors representing whether DEGs used in analysis were increased in expression, (brown), or decreased in expression (teal) for cases compared to controls. White represents a term not being enriched.

29

Figure 1.3: Differential gene expression and pathway signatures in the cerebellum (CBE). a. Hierarchical clustering of samples by samples, b. Pairwise DEG plots and corresponding upset plots indicating intersections of AD versus non-AD differentially expressed genes (DEGs) (Benjamini-Hochberg (BH) adjusted p-value < 0.05 and absolute log2 fold change (FC) > 0.4) across APOE4 groups. Rows correspond to APOE4 groups. The bar chart shows the number of single and common sets of DEGs across APOE4 groups. Single filled dots represent a unique set of DEGs for the corresponding APOE4 group. Multiple filled black dots connected by vertical lines represent common sets of DEGs across cell types, APOE4 groups, c. log2 FC scores of all genes in the DE analysis, d. Volcano plots of DEGs for both APOE4 groups, e. Gene ontology (left) and KEGG pathway (right) frameworks (BH adjusted p-value < 0.05) for both APOE4 groups with colors representing whether DEGs used in analysis were increased in expression, (brown), or decreased in expression (teal) for cases compared to controls. White represents a term not being enriched.

30

Figure 1.4: Differential gene expression and pathway signatures in the dorsolateral prefrontal cortex (DLPFC). a. Hierarchical clustering of samples by samples, b. Pairwise DEG plots and corresponding upset plots indicating intersections of AD versus non-AD differentially expressed genes (DEGs) (Benjamini-Hochberg (BH) adjusted p-value < 0.05 and absolute log2 fold change (FC) > 0.4) across APOE4 groups. Rows correspond to APOE4 groups. The bar chart shows the number of single and common sets of DEGs across APOE4 groups. Single filled dots represent a unique set of DEGs for the corresponding APOE4 group. Multiple filled black dots connected by vertical lines represent common sets of DEGs across cell types, APOE4 groups, c. log2 FC scores of all genes in the DE analysis, d. Volcano plots of DEGs for both APOE4 groups, e. Gene ontology (left) and KEGG pathway (right) frameworks (BH adjusted p-value < 0.05) for both APOE4 groups with colors representing whether DEGs used in analysis were increased in expression, (brown), or decreased in expression (teal) for cases compared to controls. White represents a term not being enriched.

31

Figure 1.5: Differential gene expression and pathway signatures in the anterior prefrontal cortex (BA 10).
a. Hierarchical clustering of samples by samples, b. Pairwise DEG plots and corresponding upset plots indicating intersections of AD versus non-AD differentially expressed genes (DEGs) (Benjamini-Hochberg (BH) adjusted p-value < 0.05 and absolute log2 fold change (FC) > 0.4) across APOE4 groups. Rows correspond to APOE4 groups. The bar chart shows the number of single and common sets of DEGs across APOE4 groups. Single filled dots represent a unique set of DEGs for the corresponding APOE4 group. Multiple filled black dots connected by vertical lines represent common sets of DEGs across cell types, APOE4 groups, c. log2 FC scores of all genes in the DE analysis, d. Volcano plots of DEGs for both APOE4 groups, e. Gene ontology (left) and KEGG pathway (right) frameworks (BH adjusted p-value < 0.05) for both APOE4 groups with colors representing whether DEGs used in analysis were increased in expression, (brown), or decreased in expression (teal) for cases compared to controls. White represents a term not being enriched.

Figure 1.6: Differential gene expression and pathway signatures in the posterior superior temporal gyrus/ Wernicke's area (BA 22). a. Hierarchical clustering of samples by samples, b. Pairwise DEG plots and corresponding upset plots indicating intersections of AD versus non-AD differentially expressed genes (DEGs) (Benjamini-Hochberg (BH) adjusted p-value < 0.05 and absolute log2 fold change (FC) > 0.4) across APOE4 groups. Rows correspond to APOE4 groups. The bar chart shows the number of single and common sets of DEGs across APOE4 groups. Single filled dots represent a unique set of DEGs for the corresponding APOE4 group. Multiple filled black dots connected by vertical lines represent common sets of DEGs across cell types, APOE4 groups, c. log2 FC scores of all genes in the DE analysis, d. Volcano plots of DEGs for both APOE4 groups, e. Gene ontology (left) and KEGG pathway (right) frameworks (BH adjusted p-value < 0.05) for both APOE4 groups with with colors representing whether DEGs used in analysis were increased in expression, (brown), or decreased in expression (teal) for cases compared to controls. White represents a term not being enriched.
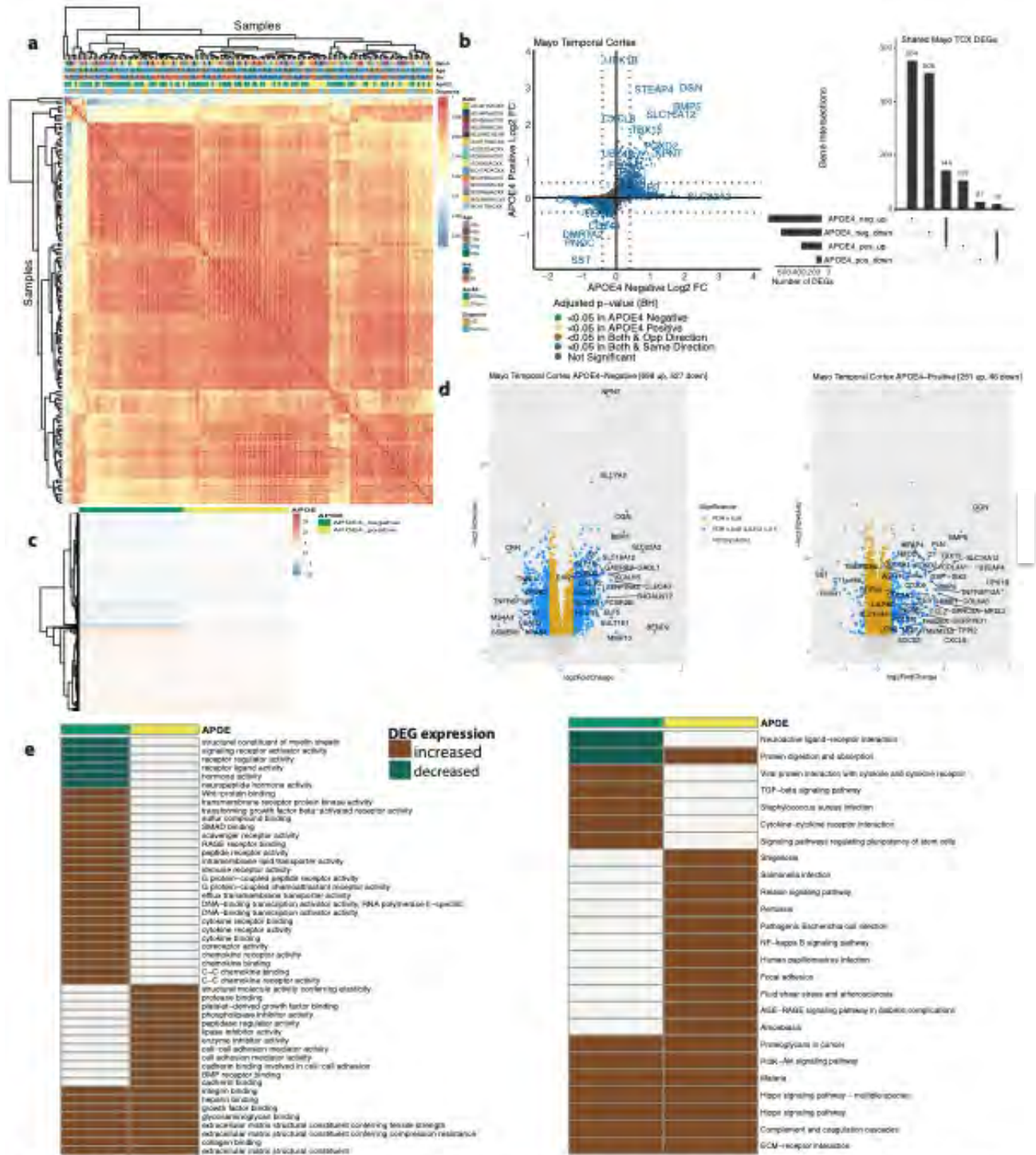
Figure 1.7: Differential gene expression and pathway signatures in the perirhinal cortex (BA 36). a. Hierarchical clustering of samples by samples, b. Pairwise DEG plots and corresponding upset plots indicating intersections of AD versus non-AD differentially expressed genes (DEGs) (Benjamini-Hochberg (BH) adjusted p-value < 0.05 and absolute log2 fold change (FC) > 0.4) across APOE4 groups. Rows correspond to APOE4 groups. The bar chart shows the number of single and common sets of DEGs across APOE4 groups. Single filled dots represent a unique set of DEGs for the corresponding APOE4 group. Multiple filled black dots connected by vertical lines represent common sets of DEGs across cell types, APOE4 groups, c. log2 FC scores of all genes in the DE analysis, d. Volcano plots of DEGs for both APOE4 groups, e. Gene ontology (left) and KEGG pathway (right) frameworks (BH adjusted p-value < 0.05) for both APOE4 groups with with colors representing whether DEGs used in analysis were increased in expression, (brown), or decreased in expression (teal) for cases compared to controls. White represents a term not being enriched.

Figure 1.8: Differential gene expression and pathway signatures in the inferior frontal gyrus/ Broca's area (BA 44). a. Hierarchical clustering of samples by samples, b. Pairwise DEG plots and corresponding upset plots indicating intersections of AD versus non-AD differentially expressed genes (DEGs) (Benjamini-Hochberg (BH) adjusted p-value < 0.05 and absolute log2 fold change (FC) > 0.4) across APOE4 groups. Rows correspond to APOE4 groups. The bar chart shows the number of single and common sets of DEGs across APOE4 groups. Single filled dots represent a unique set of DEGs for the corresponding APOE4 group. Multiple filled black dots connected by vertical lines represent common sets of DEGs across cell types, APOE4 groups, c. log2 FC scores of all genes in the DE analysis, d. Volcano plots of DEGs for both APOE4 groups, e. Gene ontology frameworks (BH adjusted p-value < 0.05) for both APOE4 groups with with colors representing whether DEGs used in analysis were increased in expression, (brown), or decreased in expression (teal) for cases compared to controls. White represents a term not being enriched.
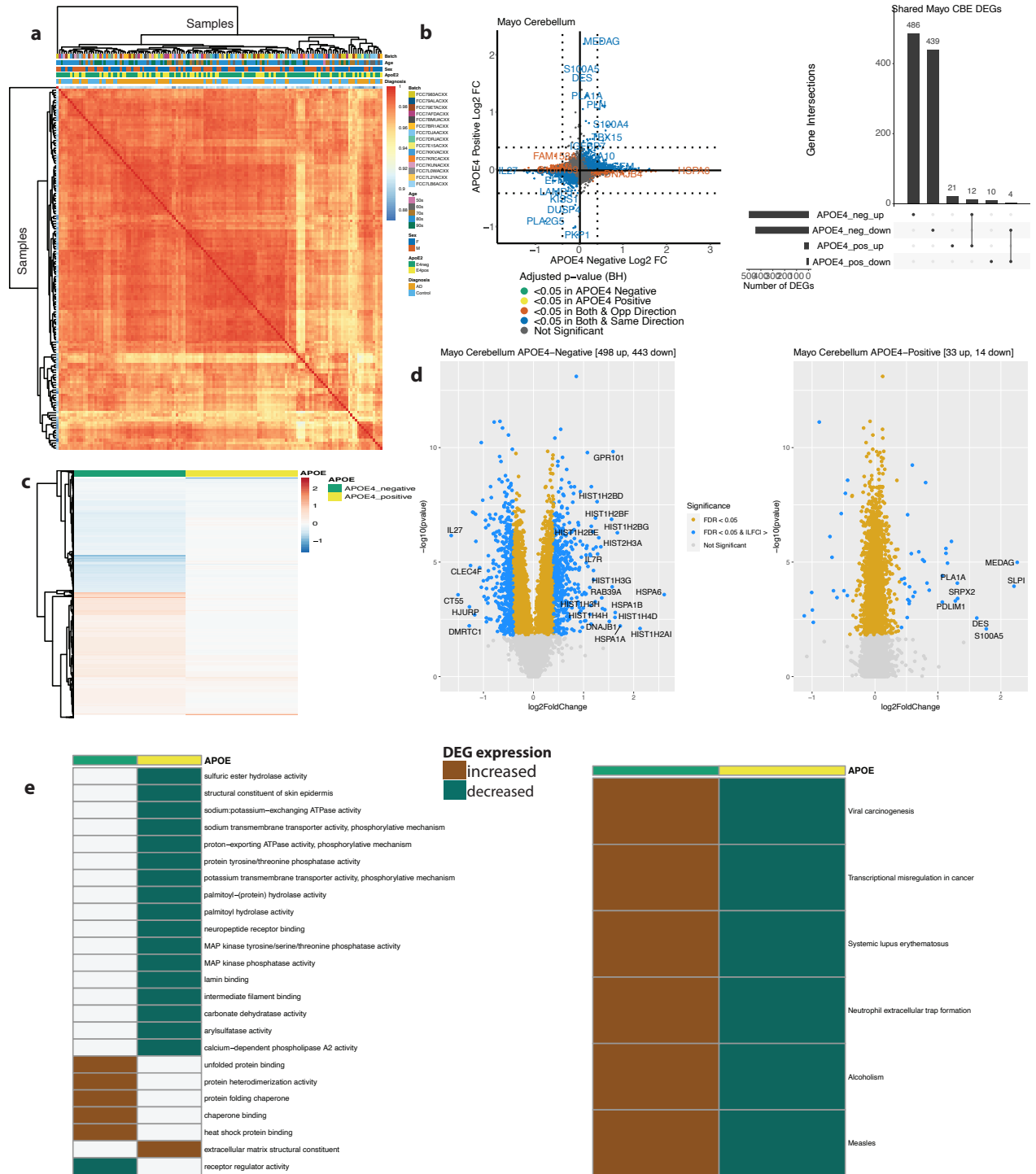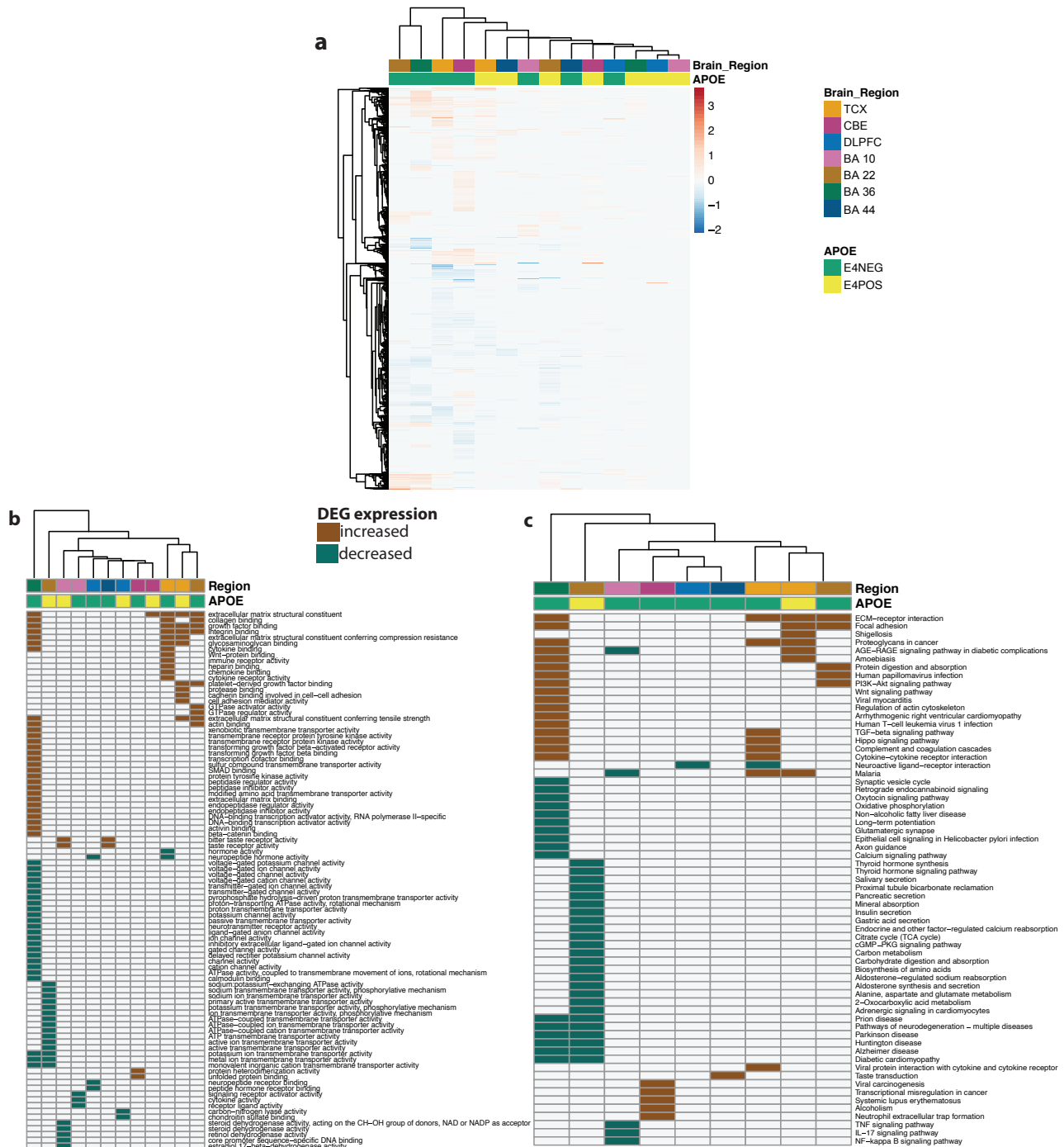
Figure 1.9: Comparison disease-related transcriptomic changes across brain regions and *APOE* genotype
a. log2 FC scores of all genes in the DE analysis clustered by brain region and *APOE* genotype,
b-c. Gene ontology (b) and KEGG pathway (c) frameworks (BH adjusted p-value < 0.01) for all brain regions and APOE4 groups, clustered by brain region and APOE4 group. Colors represent whether DEGs used in analysis were upregulated (brown), or down regulated (teal) for cases compared to controls. White represents a term not being enriched. Brain regions include temporal cortex (TCX) and cerebellum (CBE), dorsolateral prefrontal cortex (DLPFC), anterior prefrontal cortex BA 10), superior temporal gyrus (BA 22), perirhinal cortex (BA 36), and inferior frontal gyrus (BA 44).

36
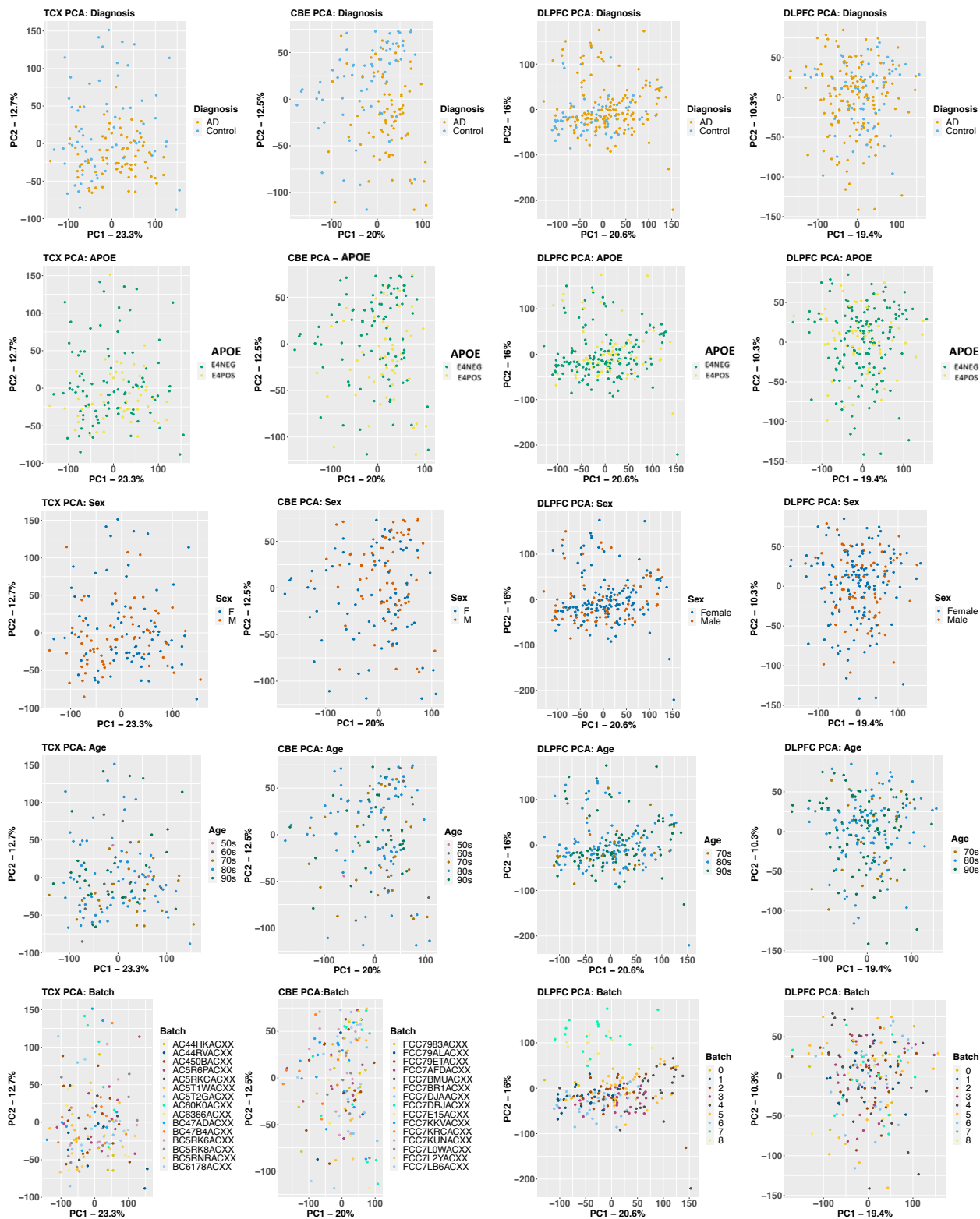
Figure 1.10: Dimensionality reduction of samples from the temporal cortex (TCX), cerebellum (CBE), and dorsolateral prefrontal cortex (DLPFC) by diagnosis, APOE genotype, sex, age, and batch. PCA plots for a. TCX, b. CBE, c. DLPFC pre batch correction and outlier removal, d. DLPFC post batch correction and outlier removal
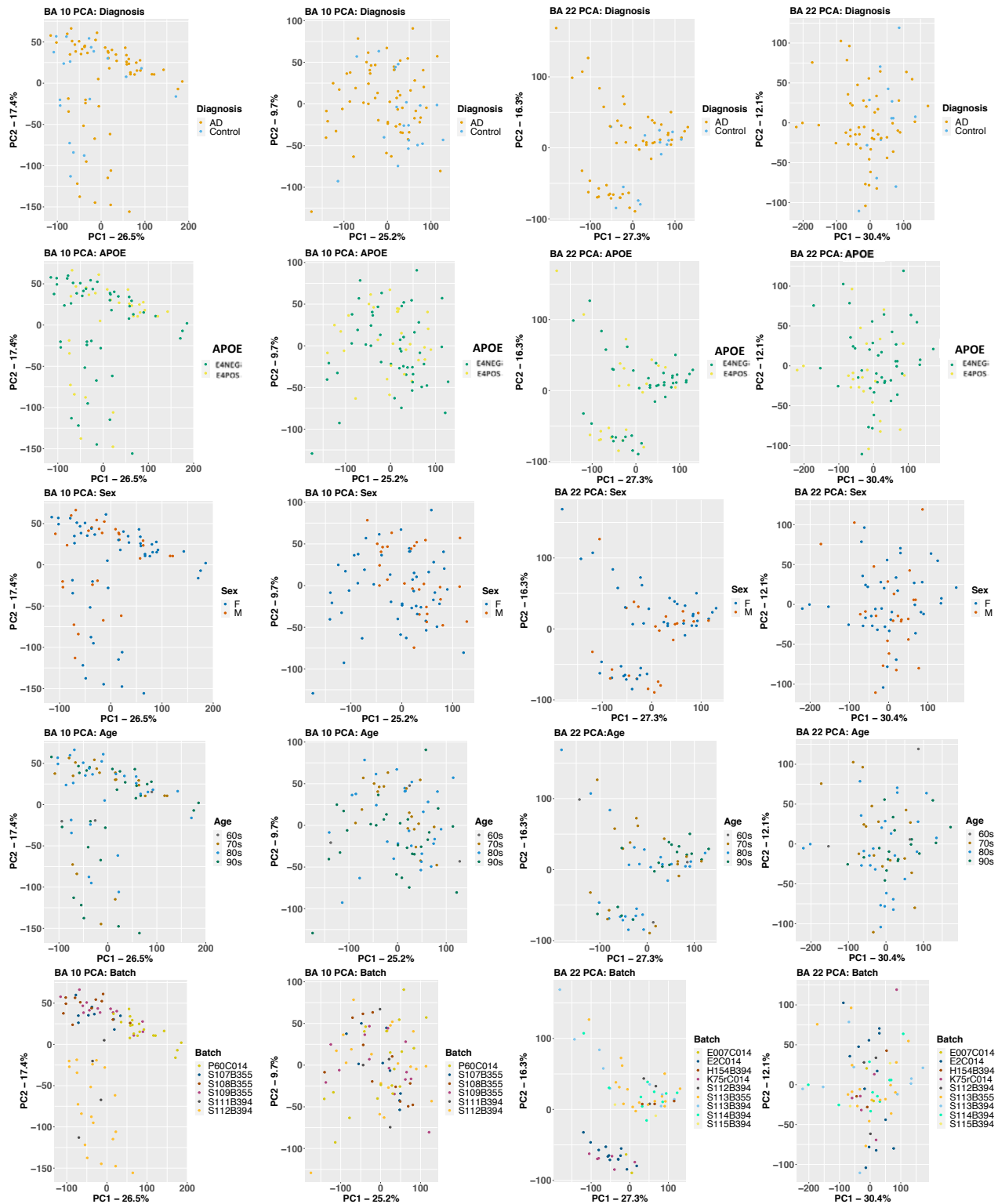
Figure 1.11: Dimensionality reduction of samples from the anterior prefrontal cortex (BA 10) and superior temporal gyrus (BA 22) by diagnosis, APOE genotype, sex, age, and batch. a. BA 10 before batch correction, b. BA 10, c. BA 22 before batch correction, d. BA 22 after batch correction
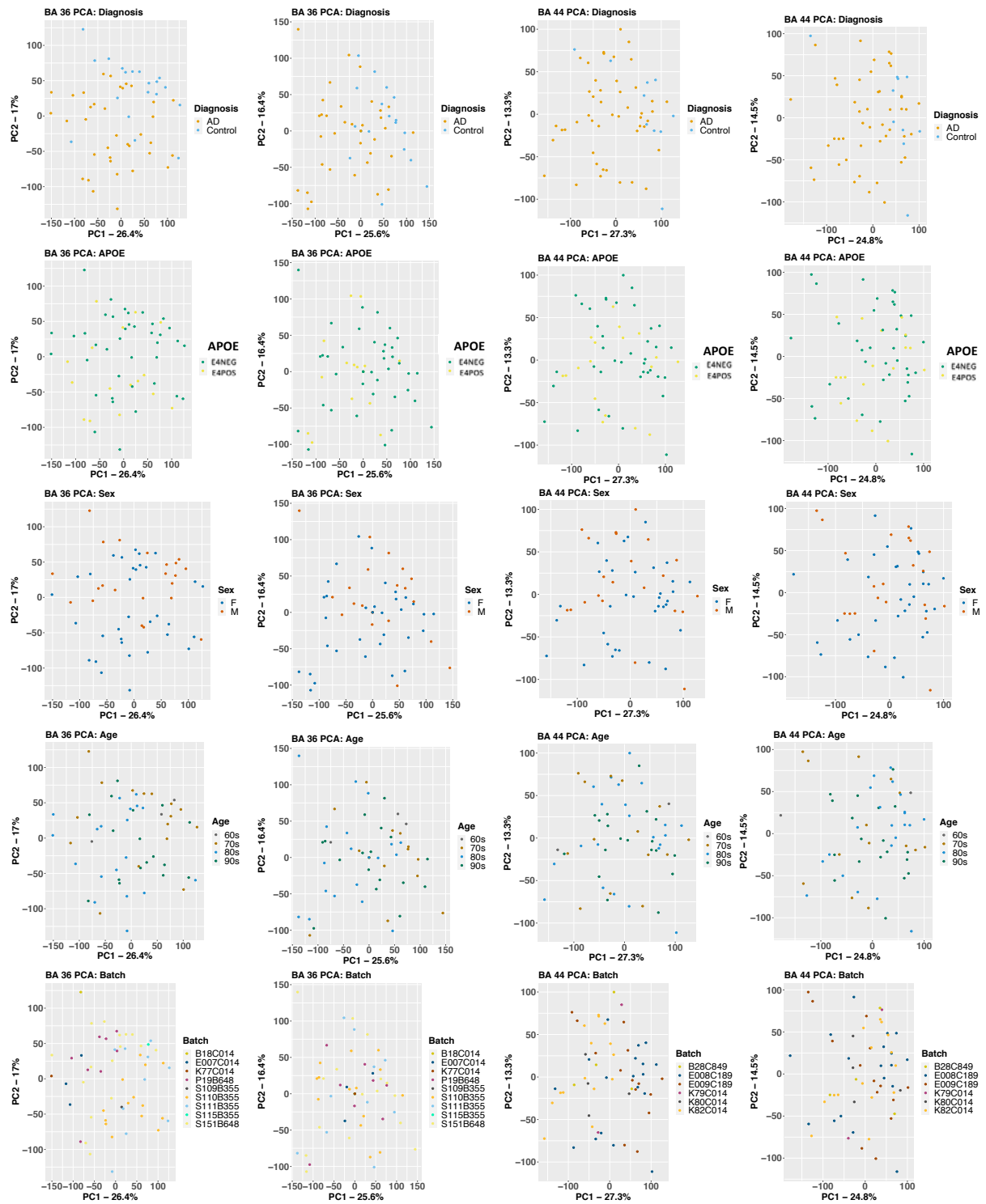
Figure 1.12: Dimensionality reduction of samples from the perirhinal cortex (BA 36) and inferior frontal gyrus (BA 44) by diagnosis, APOE genotype, sex, age, and batch batch. a. BA 36 before batch correction, b. BA 36, c. BA 44 before batch correction, d. BA 44 after batch correction

CHAPTER 2

Single-cell transcriptomic analysis elucidates *APOE* genotype-specific changes across cell types in two brain regions in Alzheimer's disease.

2.1 ABSTRACT

Alzheimer's Disease (AD) is a complex neurodegenerative disease that gravely affects patients

and imposes an immense burden on caregivers. Apolipoprotein E4 (APOE4) has been identified

as the most common genetic risk factor for AD, yet the molecular mechanisms connecting

APOE4 to AD are not well understood. Past transcriptomic analyses in AD have revealed *APOE*

genotype-specific transcriptomic differences; however, these differences have not been explored

at a single-cell level. Here, we leverage the first two single-nucleus RNA sequencing AD

datasets from human brain samples, including nearly 55,000 cells from the prefrontal and

entorhinal cortices. We observed more global transcriptomic changes in APOE4 positive AD

cells and identified differences across *APOE* genotypes primarily in glial cell types. Our findings

highlight the differential transcriptomic perturbations of APOE isoforms at a single-cell level in

AD pathogenesis and have implications for precision medicine development in the diagnosis and

treatment of AD.

2.2 INTRODUCTION

Alzheimer's disease (AD) is a heterogeneous neurodegenerative disorder, which accounts for at least 60% of dementia cases[1]. Further underscoring the importance of AD research, cases of AD are projected to increase by more than 3-fold by 2050, yet there are currently no disease altering treatments[2,3]. AD is defined by pathological hallmarks of aggregated extracellular amyloid-β (Aβ) plaques, and intracellular tau neurofibrillary tangles[1,4]. As a complex disease, AD has several environmental risk factors. Demographic risk factors include advanced age, low education level, and female sex. AD genetic risk factors such as Aβ precursor protein (*APP*), presenilin 1 (*PSEN1*), and presenilin 2 (*PSEN2*) lead to dominantly inherited early-onset AD and account for less than 1% of AD cases[1,4,5].

The strongest genetic risk factor for late-onset or sporadic AD is the ε4 allele of the apolipoprotein E (*APOE*) gene. In humans, there are three common *APOE* allelic variants: ε2, ε3, and ε4, which differ based on single substitutions at amino acid residues 112 and 158. The ε3 allele is the most common variant, and is generally considered as a neutral form[4,6,7]. The ε2 allele is considered protective, and the ε4 allele is associated with increasing the risk of developing AD in a gene dose dependent manner[4,8]. Specifically, one copy of the ε4 allele of *APOE* increases the risk of developing AD by 3- to 4-fold, and two copies increases this risk by 12- to 15-fold[4,5].

APOE is a lipid-binding protein, which serves a central role in regulating lipid transport and metabolism. It is highly expressed in the liver and brain, where in the latter, it is primarily expressed in astrocytes[4,6]. APOE's functionality in the central nervous system has implications for AD in both Aβ-dependent and Aβ-independent pathways. For instance, in addition to

regulating Aβ clearance, APOE regulates lipoprotein metabolism, supports cell proliferation, repairs membranes, supports myelination, and maintains blood brain barrier (BBB) integrity[4,6,9]. With regards to APOE isoforms, APOE4 has been linked to promoting Aβ retention by blocking its LRP1-mediated clearance[9,10], insulin resistance through impaired insulin signaling[11], BBB dysfunction and increased permeability[8,12], and regulating glycogen synthase kinase 3 (*GSK3*), a kinase highly involved in phosphorylation of tau[6,13]. Our study aims to identify transcriptomic differences associated with APOE isoforms at a single-cell level to better understand the underlying mechanisms contributing to AD pathophysiology and their specificity to each isoform. Transcriptomics represent a valuable means of understanding molecular underpinnings in disease conditions [10,14–18]; however, to our knowledge, in AD, APOE isoforms are yet to be investigated at a single-cell level, which can depict molecular profiles that would be otherwise masked in a bulk analysis.

In recent years, single-cell transcriptomic datasets were generated from the prefrontal[19] and entorhinal[20] cortices of human AD patients. First, Mathys and colleagues performed single-cell RNA sequencing (RNA-Seq) using prefrontal cortex samples from 24 individuals with high Aβ burden and related AD pathology, and 24 individuals with little to no Aβ burden or other pathologies. Second, Grubman and colleagues surveyed single-nucleus transcriptomes from the entorhinal cortices of 6 AD individuals and 6 cognitively normal controls. While both studies provided valuable human transcriptomic profiles at single-cell resolution and consistently reported cell type-variable *APOE* expression in AD, upregulated in microglia and downregulated in astrocytes, the authors did not examine cell type-specific gene expression differences in disease based on *APOE* genetic variants. In this study, we leverage these two publicly available

datasets to study the cell type-specific transcriptomic effects of *APOE* genotype in AD across two brain regions: the prefrontal and entorhinal cortices. We aim to answer the following questions: 1) Which cell types are most affected at the transcriptomic level by *APOE* genotype in the context of AD? 2) What are the global and cell type-specific transcriptomic changes with respect to *APOE* genotype in the context of AD? and 3) Are there any transcriptomic changes that are specific to APOE4 that better explain AD pathophysiology?

2.3 RESULTS

2.3.1 Sample classification and analytic workflow

We classified samples into AD and control groups based on tau tangle and Aβ plaque burdens, using Braak clinical staging and Consortium to Establish a Registry for Alzheimer's Disease (CERAD) scores[21], respectively (AD: Braak stage ≥ 4, CERAD score ≤ 2; Control: Braak stage ≤ 3, CERAD score ≥ 3) (**Fig. 2.1**). Next, from the prefrontal cortex cohort (**Table 2.1**), we analyzed single nucleus RNA-Seq (snRNA-seq) data containing 43,831 cells (**Supplementary Table 2.1**) and 17,593 genes, and from the entorhinal cortex cohort (**Table 2.2**), we analyzed snRNA-seq data containing 9,587 cells (**Supplementary Table 2.2**) and 10,850 genes. Both datasets were acquired from different sets of individuals.

Due to the limited number of samples for relatively rare *APOE* genotypes, we focused our analysis on comparisons between AD and non-AD groups with APOE3/3 (homozygous for allele ε3) and APOE3/4 (heterozygous ε3/ε4) genotypes. We performed an *APOE* genotype-stratified differential gene expression (DGE) analysis comparing age-matched AD cases to controls, with sex as a covariate, in excitatory (Ex), and inhibitory (In) neurons for the prefrontal cortex specifically, undistinguished neurons (Neu) for the entorhinal cortex, and astrocytes (Ast), microglia (Mic), oligodendrocytes (Oli), and oligodendrocyte progenitor cells (OPCs) for both cohorts (**Supplementary Fig. 2.1**). Differentially expressed genes (DEGs) were selected using cutoffs of a Benjamini-Hochberg (BH) adjusted p-value < 0.05 and log2 fold change > 0.25. DEGs were further passed as inputs to identify pathways for subsequent network analysis. We examined gene expression and network changes in AD compared to non-AD samples to identify cell type-specific and shared changes based on *APOE* genotype (**Fig. 2.1**).

2.3.2 APOE genotype-stratified DGE analysis in the prefrontal cortex identifies distinct AD-related changes in astrocytes, oligodendrocytes, and OPCs

Leveraging data from Mathys et al., we identified DEGs in all cell type and *APOE* genotype pairings when comparing AD to control tissue from 43,831 cells covering 17,593 genes. Interestingly, DEGs were primarily downregulated in APOE3/4 astrocytes, oligodendrocytes and OPCs, while they were primarily upregulated in both APOE3/3 and APOE3/4 neurons (**Fig. 2.2a**). Altogether, across all cell types we identified 278 unique DEGs (**Supplementary Table 2.3**). Of the 278 DEGs, 8 were specific to APOE3/3 and 135 were specific to APOE3/4. We observed DEGs previously linked to AD (*CLU*[5,22], *CCK*[23–25], *NRGN*[26,27], *DHFR*[28,29], *ERBB4*[30,31], *NRXN1*[32]), which were shared by APOE3/3 and APOE3/4 cells. In most cases, expression differences in these genes were in the same direction across genotypes, but with greater fold changes in APOE3/4 as compared to APOE3/3 cells (**Fig. 2.2b**). Across cell types, while the majority of DEGs were shared and in consistent direction across APOE3/3 and APOE3/4 cells (**Fig. 2.2c, yellow color; Supplementary Fig. 2.2**), there were a few shared DEGs with opposite directionality of expression changes, such as *DOCK4* in microglia, *SPARCL1* in neurons, and *FRYL* in oligodendrocytes (**Fig. 2.2c, pink color; Supplementary Fig. 2.2**).

Notably, some DEGs in AD patients relative to controls were shared across multiple cell types (**Fig. 2.3a**). Examples of some DEGs in AD patients relative to controls that overlap most across cell types within or across *APOE* genotypes include *APP* binding family B member 1 interacting protein (*APBB1IP*), and *DOCK8*, a protein highly involved in brain development and immune response[33], were differentially expressed in most APOE3/4 cell types and in APOE3/3 neurons. Interestingly, for both *APBB1IP* and *DOCK8*, we observed cell type-specific effects. Both genes

45

were downregulated in astrocytes, oligodendrocytes and OPCs and upregulated in microglia and neurons from APOE3/4 AD patients versus APOE3/4 controls. In APOE3/3 individuals, both genes were only significantly upregulated in neurons in AD patients versus controls. *APOE* itself was also differentially expressed in AD patients versus controls, with an increase in both APOE3/4 and APOE3/3 neurons and in APOE3/3 microglia as well as a decrease in APOE3/4 astrocytes, oligodendrocytes, and OPCs. *MTRNR2L12* expression, which encodes a humanin isoform necessary for neuroprotection and anti-apoptotic function suggested to have utility as a blood marker for cognitive disability and early dementia for adults with Down Syndrome[34,35], was very similar to *APOE* expression. Hierarchical clustering of samples using AD compared to control pseudobulk cell type gene expression (**Fig. 2.2d**) showed samples to cluster by *APOE* genotype before cell type identity for all cell types except neurons. Generally, through our *APOE* genotype-stratified analysis we observed more similarities in AD versus control DEGs across *APOE* genotypes in neuronal populations (both excitatory and inhibitory neurons), and differences primarily in non-neuronal cells (astrocytes, oligodendrocytes, and OPCs). In addition to identifying shared DEGs across cell types and *APOE* genotypes, we also observed a larger range of log2 fold change in the analysis of APOE3/4 AD versus control (-0.834, 1.032; median = -0.273) compared to the analysis of APOE3/3 AD versus control (-0.503,1.115; median = 0.342), which we visualized in several shared DEGs such as *LINGO1*, *NRXN1*, *RASGEF1B*, and *CLU* (**Fig. 2.3b**).

As the prefrontal cohort contained a sole non-AD sample with the APOE3/4 genotype from a male donor, we performed a sensitivity analysis in male samples to determine whether similar gene signatures remain. We identified 300 unique DEGs across all cell types (**Supplementary**

Fig. 2.3; Supplementary Table 2.4). Of these DEGs, 18 were specific to APOE3/3 cells and 128 to APOE3/4 cells. Like the previous analysis, we observed more differences in perturbed gene profiles across *APOE* genotypes in astrocytes, oligodendrocytes, and OPCs, where DEGs were primarily downregulated in APOE3/4 cells. Additionally, clustering samples using AD compared to control pseudobulk cell type gene expression also showed a stronger clustering by *APOE* genotype than cell type identity (**Supplementary Fig. 2.3**).

2.3.3 APOE genotype-stratified DGE analysis in the entorhinal cortex identifies distinct AD-related changes in microglia and oligodendrocytes

Leveraging data from Grubman et al., we identified DEGs in all cell type and *APOE* genotype pairings when comparing AD to control tissue from 9,587 cells and 10,850 genes. We found DEGs to be primarily downregulated in APOE3/3 AD versus control and upregulated in APOE3/4 AD versus control (**Fig. 2.4a**). Altogether, across all cell types we identified 232 unique DEGs (**Supplementary Table 2.5**). Of the DEGs, 29 were specific to the APOE3/4 AD, and none were specific to the APOE3/3 AD. In each cell type, we observed more DEGs in the APOE3/4 comparison, some of which were shared with the APOE3/3 analysis, though often with consistent opposite directionality (**Fig. 2.4b; yellow (same) and pink (opposite) colors; Supplementary Fig. 2.2**). We observed a higher proportion of common DEGs across *APOE* groups in microglia and oligodendrocytes than in other cell types, and in most cases, there was opposite directionality of gene expression changes between the APOE3/3 AD versus control comparison and APOE3/4 AD versus control comparison. Overall, clustering samples using AD compared to control pseudobulk cell type gene expression (**Fig. 2.4c**) showed consistent clustering of samples by *APOE* genotype.

When surveying DEG overlaps across cell types in the entorhinal cortex, consistent with the prefrontal cortex analysis, we observed more DEGs in AD patients relative to controls shared across APOE3/4 cell types than across APOE3/3 cell types (**Supplementary Fig. 2.4**). To highlight some of these DEGs that overlap most across cell types, in the APOE3/3 case-control comparisons, six DEGs –*ATP1B1* [36], a sodium and potassium ATPase necessary for regulating ionic gradients; *CST3*[37], an AD risk factor; *GPC5*[38], a neurotrophic factor; *MEG3*[39], a long non-coding RNA and apoptosis regulator; *NRXN1*; and *LINC00486*, a relatively uncharacterized long non-coding transcript – were shared by all cell types. *LINC00486* was upregulated in all APOE3/3 cell types in AD, *ATP1B1*, *GPC5*, *MEG3*, and *NRXN1* were downregulated in all APOE3/3 cell types in AD, and *CST3* was downregulated in all APOE3/3 cell types in AD, except OPCs where it was upregulated. These DEG's were also reflected in APOE3/4 cells, with *LINC00486* upregulated in all cell types, *ATP1B1* and *MEG3* upregulated in non-neuronal cell types, *NRXN1* upregulated in oligodendrocytes and downregulated in all other cell types, *GPC5* downregulated in astrocytes and upregulated in all other cell types, and *CST3* downregulated in astrocytes and upregulated in neurons and oligodendrocytes in case-control comparisons.

Overall, in the APOE3/4 case-control comparisons, 87 DEGs were shared in all cell types, with 64 consistently upregulated in AD tissue and 23 with mixed directionality across cell types when comparing AD to control tissue (**Supplementary Fig. 2.4**). Of these shared DEGs, a few with higher absolute log2 fold changes between AD and controls include *MBP*, a gene important for myelination[40,41] that was upregulated in all APOE3/4 cell types in AD except oligodendrocytes, and *LINGO1*, which was upregulated in all APOE3/4 cell types as well as APOE3/3 astrocytes and OPCs in AD. Interestingly the average log2 fold change for *LINGO1* in APOE3/4 AD

samples (3.52) was much higher than that of the APOE3/3 AD samples (0.451). Additionally, protein folding *HSPA1A*, the neuroprotective chaperone and apoptosis regulator *CRYAB*[42], and quinoid dihydropteridine reductase (*QDPR*) were upregulated in all APOE3/4 cell types in AD. However, *HSPA1A* was downregulated in APOE3/3 microglia, oligodendrocytes, and OPCs, *CRYAB* was downregulated in APOE3/3 oligodendrocytes, and *QDPR* was downregulated in APOE3/3 microglia in AD. The latter two genes have previously been observed to be upregulated in oligodendrocytes and OPCs of pathologically confirmed AD individuals[19], most of them are usually APOE4 carriers. We also observed a larger range of case-control log2 fold change in APOE3/4 cells (-2.918, 3.839; median= 0.688) compared to APOE3/3 cells (-2.385, 2.227; median= -0.436), which we visualized in a few shared DEGs such as *LINGO1*, *NRXN1*, FTL, and *ADGRL3* (**Supplementary Fig. 2.4**). Largely, when comparing AD to non-AD cells in the entorhinal cortex, while we observed changes relevant to AD pathophysiology across APOE3/3 and APOE3/4 genotypes, we also observed flipped DEG expression profiles across both *APOE* genotypes primarily in non-neuronal cells, and more universal transcriptional changes and changes of higher amplitude in the APOE3/4 AD versus control comparison as compared to APOE3/3 AD versus control comparison.

2.3.4 Comparative analysis across brain regions shows more AD-related transcriptomic changes in the entorhinal cortex compared to the prefrontal cortex, with consistent APOE genotype-specific disease signatures

We observed a higher number of DEGs and larger log2 fold change magnitudes across cell types in the entorhinal cortex than in the prefrontal cortex in AD. The number of shared DEGs within cell types across *APOE* genotype groups was highest in the entorhinal cortex in AD, while the

number of shared DEGs within cell types across brain regions was highest in APOE3/4 cells in AD (**Fig. 2.5a**). With hierarchical clustering of per-cell and genotype group pseudobulk expression, while we do not observe total clustering by any of the variables examined, we see some clustering by brain region, and within these brain regions, by *APOE* genotype (**Fig. 2.5b**).

2.3.5 Pathway and network analysis reveal APOE genotype-specific perturbed biological processes primarily in glial cells across brain regions

Pathway enrichment was performed using g:Profiler[43], a web tool that performs functional enrichment using an input of gene lists. Separate lists of upregulated and downregulated DEGs in AD relative to control, with a BH corrected adjusted p-value < 0.05 and a relaxed absolute log2 fold change cutoff of above 0.1, in each cell type and *APOE* genotype were used as inputs for g:Profiler (**Supplementary Tables 2.6-2.9**). A network analysis was performed to cluster the disease enriched pathways into biologically relevant groups using pathways with an adjusted p-value < 0.01 as inputs. Modules of biological themes were generated for each cell type based on the *APOE* genotype (**Fig. 2.6, Supplementary Fig. 2.5 and 2.6**).

In astrocytes from the prefrontal cortex, we identified six enriched functional modules in both APOE3/3 and APOE3/4 AD relative to controls (**Fig. 2.6a**). Five out of six were downregulated in AD, and one, the *LINGO1-TROY-NgR* complex, which was previously suggested to be important for modulating glial-neuronal interactions in demyelinating lesions, was upregulated in AD[44]. In APOE3/3 astrocytes, ion and acid transport, glutamate receptor activity (*mGLUR2, mGLUR3, mGLUR4, mGLUR7, mGLUR8*), metabolic (aspartate uptake, astrocytic metabolism) as well as autolysosome activities (scavenging class receptors, secondary lysosome,

autolysosome) were downregulated in AD, and myelin maintenance (*PRNP*, *ASAH1*), cell adhesion (*FLRT3*, *LPHN3*, *UNC5B*, *UNC5D*), and Vascular endothelial growth factor (*VEGF*) induced heat shock protein 90 (hsp90) complex were upregulated in AD, indicating perturbation in processes important for autophagy and stress response which are known to accompany disease progression[4,5]. APOE3/4 astrocytes uniquely showed upregulation in pathways related to post-synaptic scaffold proteins (e.g., *DLGAP1*, *DLG4*, *DLC1* and *SHANK3*) and actin assembly at cell junctions, but downregulation of synaptic membrane and neurotransmitter pathways, neurogenesis, and nervous system development in AD.

In APOE3/3 astrocytes of the entorhinal cortex, we observed a downregulation of ion and neurotransmitter transport related pathways (intracellular ion and ferritin iron sequestering) in AD. APOE3/4 astrocytes in the entorhinal cortex had mostly upregulated pathway enrichment modules in AD, in contrast to what was observed in prefrontal cortex (**Supplementary Fig. 2.6**). Many of these pathways governing cellular homeostasis, such as ATP synthesis, transmembrane cation transport, amyloid fibril formation and exosome regulation, and macromolecule and protein plasma membrane localization.

Microglia, the resident brain macrophage, contributes to neuroinflammation in AD and produces APOE upon activation in the brain[1,4]. Differentially enriched pathways were predominantly upregulated in APOE3/4 microglia in AD patients in both prefrontal (**Supplementary Fig. 2.5**) and entorhinal cortices (**Fig. 2.6b**), while downregulated in APOE3/3 microglia. Within the entorhinal cortex, changes in gliogenesis, myelination, cation transmembrane transport, cellular projection, synaptic spine development, and synaptic junction assembly pathway network

51

modules were shared in APOE3/3 AD and APOE3/4 AD but perturbed in opposite directions, downregulated in APOE3/3 and upregulated in APOE3/4 microglia (**Fig. 2.6b**). The *ITGAV-ITGB-SPP1* complex, not previously linked to AD to our knowledge, was significantly upregulated in both brain regions in APOE3/3 microglia in AD, but only in the prefrontal cortex in APOE3/4 microglia in AD (**Fig. 2.6b and Supplementary Fig. 2.5**). The downregulation of iron homeostasis and ferritin complex, a protein that binds to iron and reflects the level of iron storage in the body, was observed in APOE3/3 microglia and astrocytes of both prefrontal and entorhinal cortex in AD (**Fig. 2.6a, Fig. 2.6b, and Supplementary Fig. 2.5**).

Overall, network analysis comparing neurons from two brain regions yielded many similar perturbed biological processes within each *APOE* genotype in AD (**Fig. 2.6c**). In APOE3/3 neurons, shared differentially perturbed processes between brain regions were mostly related to regulation of membrane homeostasis, neuron projection, and synaptic development. Pathway networks in APOE3/3 neurons specific to the prefrontal cortex pertain to cell structure development (actomyosin actin-based structure, extension growth development, anchoring junction, cell adherens), while the entorhinal cortex showed unique modules relevant to cellular energy production (oxidative respirasome synthesis, metabolic ATP nucleotide process). From APOE3/4 neurons, we observed a more diverse population of shared network modules between the two brain regions, including functional processes related to protein trafficking vesicles, myelination, membrane assembly, and voltage gated channel and neurotransmitter receptor regulation. Amyloid fibril formation was uniquely differentially regulated in APOE3/4 neurons and observed in both brain regions in AD, while an amyloid beta precursor formation module was only observed APOE3/4 neurons in prefrontal cortex in AD.

In oligodendrocytes, which provide myelination, APOE3/3 carriers in the prefrontal cortex showed an upregulation of the *ITGAV-ITGB-SPP1* complex and downregulation of pathways related to myelin organization (e.g., juxtaparanode region of axon), ion transport activity, protein refolding, and regulation of MAP kinase signaling activity (e.g., positive regulation of Erk1 and Erk2) in AD. APOE3/4 oligodendrocytes, on the other hand, showed upregulation of stress responses through chaperone mediated protein folding, and downregulation of axon guidance and nervous system development processes in AD. In the entorhinal cortex, we observed modules of processes including neurogenesis, gliogenesis, amyloidosis, aerobic metabolic processes, and exocytosis to be upregulated in APOE3/4 cells and downregulated in APOE3/3 cells in AD (**Supplementary Fig. 2.6**). Lastly, we observed postsynaptic structural specialization to be uniquely downregulated in APOE3/4 oligodendrocytes.

For OPCs in the prefrontal cortex, there were no common network modules across *APOE* genotypes. In APOE3/3 AD, we identified downregulation for brain cell development processes (*AHI1-NPHP1-HAP1*) (**Supplementary Fig. 2.5**). In APOE3/4 OPCs, we observed upregulated modules for the ferritin, GAIT and *LINGO1-TROY-NgR* complexes, and downregulation for glutamatergic synaptic activity, plasma membrane and cell organization, and lipoprotein density in AD, which may have implications for neuronal integrity and lipid transport and metabolism. In the entorhinal cortex of AD, we also observed upregulation of the *LINGO1-TROY-NgR*, and downregulation of glutamatergic signaling in APOE3/3 OPCs in AD. Specific to APOE3/4 OPCs in AD, we identified upregulation of processes related to aerobic metabolic processes, stress response, autophagy, amyloid fibril regulation, demyelination, and immune response.

2.4 DISCUSSION

APOE4 is the greatest known genetic risk factor for AD; however, along with other APOE isoforms, its molecular profiles are yet to be investigated at a single-cell level. Here, we analyzed recently available single-cell transcriptomic datasets from two brain regions to better understand how *APOE* genotype plays into transcriptional profiles of AD in a cell type-specific manner. We aimed to understand whether transcriptional differences exist, and if so, how they might be represented in different cell types across brain regions; which cell types were most affected by *APOE* genotype; what changes were shared or dissimilar across cell types; and whether such findings are consistent across brain regions. We performed an *APOE* genotype-stratified differential gene expression analysis comparing AD to control samples within each cell type. Due to the limited number of samples for relatively rare *APOE* genotypes, we restricted our analysis to individuals with APOE3/3 and APOE3/4 genotypes.

In both the prefrontal and entorhinal cortices, we observed shared and unique gene signatures across these *APOE* genotypes that were often cell type-specific, but sometimes spanned many cell types (**Fig. 2.2, Fig. 2.3, Fig. 2.4, and Supplementary Fig. 2.4**). In both brain regions, we observed differing molecular profiles between *APOE* genotypes primarily in glial cells. Interestingly, in both brain regions, we observed a subset of shared DEGs and enriched pathway networks to be perturbed in opposite directions between *APOE* genotypes in AD relative to healthy state, which may indicate potential compensatory or deleterious mechanisms in disease progression in each genotype. Additionally, we observed more DEGs unique to APOE3/4 cells in AD versus control when compared to DEGs for APOE3/3 cells in AD versus control and more DEG overlaps across cell types in APOE3/4 AD, suggesting distinct disease-relevant molecular

54

profiles between *APOE* genotypes and more global AD-related molecular responses when one copy of the APOE4 allele is present.

In the prefrontal cortex, most DEGs that are common across cell types tend to be more strongly differentially expressed in APOE3/4 AD as compared to those in APOE3/3 AD. Additionally, we observed most of the *APOE* genotype-specific changes in APOE3/4 astrocytes, oligodendrocytes and OPCs, where these genes are predominantly downregulated in AD as compared to controls. Neurons, on the other hand, tended to exhibit DEGs of AD versus control that were common across *APOE* genotypes (**Fig. 2.2a and Supplementary Fig. 2.2**). Through hierarchical clustering of samples using AD compared to control pseudobulk cell type gene expression (**Fig. 2.2c**), we observed clustering by *APOE* genotype in all cell types except neurons.

In the entorhinal cortex, microglia and oligodendrocytes had the highest proportion of DEGs of AD versus control that were shared across *APOE* genotypes. Interestingly, these DEGs frequently exhibited opposite log fold-change direction between APOE3/3 AD cells and APOE3/4 AD cells, implying differing mechanisms of neurodegeneration based on the presence of the APOE4 isoform. Additionally, through hierarchical clustering of samples using AD compared to control pseudobulk cell type gene expression, we observed some influence of brain region and *APOE* genotype on gene expression (**Fig. 2.5b**). Compared to the prefrontal cortex, the entorhinal cortex, which is implicated in early stages of AD where tau begins to accumulate and the occurrence of synaptic and neuronal loss is associated with the onset of cognitive

impairment[1,4,45], had a larger log2 fold change range for DEGs overall, implying a greater magnitude of molecular changes in this region in AD.

Through pathway and network analysis, we identified biological processes potentially involved in AD pathogenesis that were uniquely modified by *APOE* genotype (**Fig. 2.6, Supplementary Fig. 2.5, and Supplementary Fig. 2.6**). While many essential cellular processes were differentially regulated in APOE3/3 neurons in AD, most were related to energy production, membrane regulation, and cellular signaling through synapse. APOE3/4 neurons in AD, on the other hand, demonstrated a perturbation of enriched pathways linked to myelination and protein trafficking vesicle regulation (both endocytosis and exosome), which are important cellular processes that protect the integrity of neurons by providing insulation and filtering toxic elements from these cells. This evidence suggests that APOE, a known lipid metabolizing protein, may play differential roles in maintaining essential metabolic processes for neuronal myelination and vesicle trafficking based on its isoform. Glial cells from APOE3/3 and APOE3/4 AD had many uniquely versus common altered biological processes, identified by the *APOE* genotype-specific pathway modules. This suggests that *APOE* genotype modifies glial cell biology in different ways compared to its effects on neuronal cell biology during AD progression. Further study on AD pathogenesis focusing on glial cell modification by the *APOE* genotype might facilitate personalized therapeutic development for AD patients with different *APOE* genotypes.

While we were able to examine *APOE* genotype-specific changes across cell types in both brain regions, some limitations exist. First, due to limited *APOE* genotypes that restricted our analysis

to APOE3/3 and APOE3/4 samples, we could not explore more transcriptional profiles such as that of higher AD risk genotypes like APOE4/4. Each dataset contained only one APOE3/4 control, which was a male sample in both cases. We performed a sensitivity analysis in males of the prefrontal cortex cohort (**Supplementary Fig. 2.3**), where we also observed more differences in perturbed gene profiles across *APOE* genotypes in astrocytes, oligodendrocytes, and OPCs, and a stronger clustering by *APOE* genotype than cell type identity.

The entorhinal cortex dataset also presents several constraints. It has a small sample size of four cases and five controls, which are also imbalanced with regards to *APOE* genotypes and sex. Specifically, all APOE3/3 samples are from female subjects, all APOE3/4 samples are from male subjects, one of the four cases is an APOE3/3 sample, and one of the five controls is an APOE3/4 sample. Additionally, we observed a batch effect, where cases were sequenced in separate batches from controls, and each batch contained only one sex. To mitigate these limitations, we used Seurat's integration workflow to integrate the batches and used dimensionality reduction to confirm appropriate batch correction (**Supplementary Fig. 2.1**). We also included sex as a covariate in our model for differential expression to account for batch while avoiding the collinearity observed with including batch.

Furthermore, we recognize some limits to our findings. Interpretation at the DEG level (**Fig. 2.3a, Fig. 2.5a, and Supplementary Fig. 2.4a**) was limiting without cell type-specific associations and AD-related mechanistic insights. To consolidate the novel and previously explored DEGs in AD into meaningful insights, we performed a pathway and network enrichment analysis. Comparing disease-relevant signatures across brain regions, we recognize

that our observations are influenced by the entorhinal cortex dataset's constraints and the variability in acquiring each cohort, which is sourced from different sets of individuals and studies. With this limitation, we could not explore further molecular profiles unique to each brain region and their implications for the spread of AD pathology. Overall, the nature of our analysis only allows for association of transcriptomic changes with *APOE* genotype, so links to causality might be hypothesized, but additional followup are needed to prove any such potential links.

Despite the limitations in our study, we present disease-relevant biology with regards to *APOE* genotype, which we hope spurs further investigation as more single-cell AD datasets become available. We hope that more single-cell AD datasets become available from more brain regions and from diverse sets of individuals, across different ages, racial and ethnic backgrounds, with a greater diversity of *APOE* genotypes and disease severity, thus allowing for more extensive insights. With more diverse genomic data, researchers may 1) integrate datasets from multiple sources and brain regions, 2) examine disease-relevant molecular changes based on *APOE* genotype across brain regions and covariates like age, sex, and severity of pathology, 3) further investigate cell type-and brain region-specificity to uncover *APOE* genotype related molecular profiles associated with the spread pathology, and 4) computationally validate findings with relevant multiomic studies, and subsequently conduct follow-up studies in vitro and in vivo. Ultimately, we identified key AD-related genes and pathways that are specific to *APOE* genotypes and cell types, especially glial cells, as well as certain consistently affected pathways. These results will inform how glial cells are potentially primary sites of AD-related transcriptional differences based on *APOE* genotype, suggesting possible mechanisms and

vulnerable cell subpopulations relevant to AD pathogenesis, and thus can help to facilitate precision medicine diagnostic and drug discovery efforts.

2.5 METHODS

2.5.1 Materials Availability

This study did not generate new unique reagents.

2.5.2 Data and Code Availability

Single nuclei RNA-Seq data and metadata were accessed from their respective repositories: the prefrontal cortex from the Accelerating Medicines Partnership Alzheimer's Disease Project (AMP-AD) Knowledge Portal under the Religious Orders Study and Memory and Aging Project (ROSMAP) (https://www.synapse.org/#!Synapse:syn18485175; https://www.synapse.org/#!Synapse:syn3157322), and the entorhinal cortex from a data repository provided by Grubman et al. (http://adsn.ddnetbio.com/). Data from the entorhinal cortex may also be accessed from the Gene Expression Omnibus (GEO) under the accession number GSE138852. Access to the prefrontal cortex dataset requires a formal request to ROSMAP. To enable other researchers to explore these datasets, all code necessary for recreating the reported analyses and figures within R, are available on Github at https://github.com/stebel5/AD_APOE_snRNAseq.

2.5.3 Study Cohort Identification

We acquired publicly available single nuclei RNA datasets from repositories specified by the first two single-cell transcriptomic AD studies[19,20]. Samples were classified based on tau neurofibrillary tangles, and amyloid β (Aβ) plaque burden, using Braak clinical staging and Consortium to Establish a Registry for Alzheimer's Disease (CERAD) scores[21]. Cases were identified as individuals with severe tau deposition (Braak stage ≥ 4) and high Aβ load (CERAD score ≤ 2), while non-AD controls were identified as individuals with low tau deposition (Braak stage ≤ 3) and low Aβ load (CERAD score ≥ 3). We also restricted our analysis to include samples with APOE3/3 (homozygous for allele ε3) and APOE3/4 (heterozygous ε3/ε4) genotypes due to the limited number of samples for relatively rare *APOE* genotypes (**Fig. 2.1**).

The prefrontal cortex dataset initially consisted of age and sex matched samples from 48 individuals with varying levels of AD pathology. For the prefrontal cortex *APOE* genotype-stratified analysis, samples consisted of 14 APOE3/3 controls, 1 APOE3/4 control, 9 APOE3/3 cases and 8 APOE3/4 cases (**Table 2.1**). The entorhinal cortex dataset initially consisted of age and sex matched samples from 6 AD and 6 control subjects, as indicated by Grubman et al. Cases were noted to have a history of AD, while controls had no history of AD or cognitive impairment, as reported by treating general practitioners. Braak scores were provided only for cases, while clinical history and amyloid pathology records were provided for all subjects. Amyloid pathology information was provided using the categories: "Numerous diffuse and neuritic Aβ plaque," "Occasional diffuse plaque in cortex," and "None." Using criteria from the Rush Alzheimer's Disease Center clinical codebook provided with the prefrontal cortex dataset, we converted these measures of neuritic plaques into CERAD scores of 1 (Definite), 3

(Possible), and 4 (No AD), respectively. This allowed us to systematically identify cases and controls in both datasets using the same criteria. For the entorhinal cortex *APOE* genotype-stratified analysis, samples consisted of 4 cases, and 5 controls (**Table 2.2**). Three of the cases were from APOE3/4 individuals, while one was from an APOE3/3 individual, and of the controls, four were from APOE3/3 individuals and the one was from an APOE3/4 individual.

2.5.4 Single Cell Data Processing, Cell Type Identification and Batch Correction

All data processing was conducted separately for each dataset with R[46] version 4.0.0 (2020-04-24) using RStudio[47], using Seurat[48] (v3.1.5). We generated visualizations using BioRender (https://biorender.com/) (**Fig. 2.1**), dittoSeq[49] (v1.0.2), an R package for analysis and color blind friendly visualization of single-cell and bulk RNA-Seq data, ggplot2, and UpsetR[50].

Prefrontal Cortex

We downloaded a filtered raw expression matrix of 17,296 genes and 70,634 cells from the prefrontal cortex from the AMP-AD Knowledge Portal and used Seurat's Read10x function to generate a count data matrix using the raw count matrix, cell names, and barcodes files provided. A Seurat object was created with the count data matrix and metadata, keeping genes present in at least 3 cells, and cells meeting cohort selection criteria with at least 200 genes. Additionally, we selected samples from APOE3/3 and APOE3/4 individuals (**Table 2.1**), which resulted in a dataset with 43,831 cells (**Supplementary Table 2.1**) and 17,593 genes. Log normalization was performed with a scale.factor of 10,000, and FindVariableFeatures was run using 3,188 features, as specified in the original paper. The data matrix was then scaled with "nCount_RNA" regressed out, and dimensionality reduction was performed with the appropriate dimensions selected based on the corresponding Principal component analysis (PCA) elbow plot.

Dimensionality reduction confirmed that there were no batch effects present (**Supplementary Fig. 2.1**). As we found the original paper's cell type identification to be comprehensive, we kept the cell type labels for the further analysis (**Supplementary Table 2.1**). Due to low cell counts, we did not analyze pericytes and endothelial cells.

Entorhinal Cortex

A filtered raw expression matrix of 10,850 genes and 13,214 cells from the entorhinal cortex was downloaded from a data repository provided by Grubman et al. Originally composed of 33,694 genes and 14,876 cells, genes and cells were filtered as described by Grubman et al. Cells from APOE3/3 and APOE3/4 individuals were selected (**Table 2.2**), and a Seurat object was created to consist of genes in at least 3 cells, and cells with at least 200 genes. Normalization was performed using Seurat's SCTransform method, and Seurat's integration workflow was performed to correct the confounded batches introduced by the experimental design. In this dataset, as shown in **Table 2.2**, control samples were processed separately from cases, male samples were processed separately from female samples, and all but one batch contained one *APOE* genotype. Dimensionality reduction was performed using values from the integrated assay to assess successful batch correction (**Supplementary Fig. 2.1**).

To identify cell types, we adopted techniques from the original paper. Briefly, Grubman et al. used Seurat's *AddModuleScore* function to calculate association scores using lists of brain cell type markers of an unspecified number from the BRETIGEA[51] package. They labeled cells based on which set of markers they had the highest score for, identified hybrids as cells where the highest and second highest score were within 20% of each other, and relabeled unidentified cells

62

based on z-score transformation of the gene score distribution. In our case, we used lists of 200 genes for astrocytes, neurons, microglia, oligodendrocytes, OPCs, and endothelial cells to label cells and hybrids to exclude as defined by Grubman et al. We further confirmed successful cell type identification by visualizing scores in a feature plot and assessing homogeneity and separation of clusters in PCA, and Uniform Manifold Approximation and Projection (UMAP) plots based on principal components and expression of top marker genes across cell types. Due to limitations in the number of cells, we excluded endothelial cells from further analyses, which resulted in a dataset comprising 10,850 genes and 9,587 cells (**Supplementary Table 2.2**).

2.5.5 Cell type-specific APOE genotype-stratified Differential Expression Analysis

To generate transcriptomic disease signatures relative to *APOE* genotype in each cell type, we used Limma-Voom[52,53]. We included the risk factor sex as a covariate in our design formula for both datasets. In the entorhinal cortex dataset, sex, instead of batch, also accounted for the confounding relationships introduced by the original study design, allowed for an appropriate model fit, and avoided the collinearity limitation observed with including batch in the design. Additionally, as samples were age matched, we also did not include age in our design formula. A *dge* list object was then created from a matrix of counts extracted from the corresponding Seurat objects. To improve the accuracy of mean-variance trend modeling and lower the severity of multiple testing correction, lowly expressed genes were filtered out using edgeR's *FilterByExpr* with default parameters. Normalization was performed with Trimmed Mean of M-values with singleton pairing (TMMwsp), followed by *voom*, model fitting with a contrast matrix of each case-control comparison for each cell type-*APOE* group, and Empirical Bayes fitting of standard errors. We performed a cell type-specific AD versus control gene expression comparison in each

APOE variant group separately in our defined prefrontal cortex cohort, entorhinal cohort, and male-only prefrontal cortex cohort, in which we excluded sex as a covariate. Differentially expressed genes (DEGs) were selected using a Benjamini-Hochberg (BH) corrected p-value less than 0.05, and an absolute log base 2-fold change greater than 0.25. We visualized unique and shared disease related gene expression changes in cell types of each *APOE* genotype using pairwise and violin plots of gene expression, hierarchical clustering of samples using AD compared to control pseudobulk cell type gene expression, and Upset plots, where genes with more overlaps across the groups compared were prioritized for labeling.

2.5.6 Functional Enrichment Analysis and Network Visualization

We performed an overrepresentation analysis of DEGs from the cell type-specific *APOE* genotype-stratified analysis of cells from the prefrontal and entorhinal cortex using gprofiler[43], a web tool for functional enrichment using an input gene list. We queried DEGs comparison split by upregulated and downregulated expression to identify enriched pathways. In addition to Gene Ontology cellular components, biological processes, and molecular functions, our enrichment analysis also provided pathways from the Human Protein Atlas, Human Phenotype Ontology, KEGG, Reactome, and Wiki pathways. We followed a previously established protocol[54] for network enrichment analysis on pathway results derived from our cell type-specific DEGs. Briefly, pathway results were imported into the Cytoscape visualization application, EnrichmentMap. We collapsed redundant and related pathways into single biological themes and further filtered significant pathways using a BH adjusted p-value <0.01. Individual biological themes were defined and summarized using the AutoAnnotate Cytoscape application.

2.7 REFERENCES

1.   Lane, C. A., Hardy, J. & Schott, J. M. Alzheimer's disease. European Journal of Neurology 25, 59–70 (2018).

2.   Maresova, P., Hruska, J., Klimova, B., Barakovic, S. & Krejcar, O. Activities of Daily Living and Associated Costs in the Most Widespread Neurodegenerative Diseases: A Systematic Review. Clin Interv Aging 15, 1841–1862 (2020).

3.   Hebert, L. E., Weuve, J., Scherr, P. A. & Evans, D. A. Alzheimer disease in the United States (2010–2050) estimated using the 2010 census. Neurology 80, 1778–1783 (2013).

4.   Long, J. M. & Holtzman, D. M. Alzheimer Disease: An Update on Pathobiology and Treatment Strategies. Cell 179, 312–339 (2019).

5.   Karch, C. M. & Goate, A. M. Alzheimer's disease risk genes and mechanisms of disease pathogenesis. Biological Psychiatry 77, 43–51 (2015).

6.   Yu, J.-T., Tan, L. & Hardy, J. Apolipoprotein E in Alzheimer's Disease: An Update. Annual Review of Neuroscience 37, 79–100 (2014).

7.   Roses, M. D., Allen D. Apolipoprotein e alleles as risk factors in alzheimer's disease. Annu. Rev. Med. 47, 387–400 (1996).

8.   Montagne, A. et al. APOE4 leads to blood–brain barrier dysfunction predicting cognitive decline. Nature 1–6 (2020) doi:10.1038/s41586-020-2247-3.

9.   Mahley, R. W., Weisgraber, K. H. & Huang, Y. Apolipoprotein E4: A causative factor and therapeutic target in neuropathology, including Alzheimer's disease. Proceedings of the National Academy of Sciences of the United States of America 103, 5644–5651 (2006).

10.   Wan, Y.-W. et al. Meta-Analysis of the Alzheimer's Disease Human Brain Transcriptome and Functional Dissection in Mouse Models. Cell Rep 32, 107908 (2020).

11.   Zhao, N. et al. Apolipoprotein E4 impairs neuronal insulin signaling by trapping insulin receptor in the endosomes. Neuron 96, 115-129.e5 (2017).

12.   Zlokovic, B. V. Neurovascular pathways to neurodegeneration in Alzheimer's disease and other disorders. Nature Reviews Neuroscience 12, (2011).

13.   Hoe, H.-S., Freeman, J. & Rebeck, G. W. Apolipoprotein E decreases tau kinases and phospho-tau levels in primary neurons. Molecular Neurodegeneration 1, 18 (2006).

14.   Magistri, M., Velmeshev, D., Makhmutova, M. & Faghihi, M. A. Transcriptomics Profiling of Alzheimer's Disease Reveal Neurovascular Defects, Altered Amyloid-β Homeostasis, and Deregulated Expression of Long Noncoding RNAs. J Alzheimers Dis 48, 647–665.

15.   Wang, M. et al. Integrative network analysis of nineteen brain regions identifies molecular signatures and networks underlying selective regional vulnerability to Alzheimer's disease. Genome Medicine 8, 104 (2016).

16.   Patel, H., Dobson, R. J. B. & Newhouse, S. J. A Meta-Analysis of Alzheimer's Disease Brain Transcriptomic Data. Journal of Alzheimer's Disease 68, 1635–1656 (2019).

17.   Allen, M. et al. Human whole genome genotype and transcriptome data for Alzheimer's and other neurodegenerative diseases. Scientific Data 3, 1–10 (2016).

18.   Neff, R. A. et al. Molecular subtyping of Alzheimer's disease using RNA sequencing data reveals novel mechanisms and targets. Science Advances 7, eabb5398 (2021).

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. Nat Neurosci 22, 2087–2097 (2019).

21.   Mirra, S. S. et al. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part II. Standardization of the neuropathologic assessment of Alzheimer's disease. Neurology 41, 479–486 (1991).

22.   Kok, E. H. et al. CLU, CR1 and PICALM genes associate with Alzheimer's-related senile plaques. Alzheimer's Research & Therapy 3, 12 (2011).

23.   Plagman, A. et al. Cholecystokinin and Alzheimer's disease: a biomarker of metabolic function, neural integrity, and cognitive performance. Neurobiology of Aging 76, 201–207 (2019).

24.   Mazurek, M. F. & Beal, F. M. Cholecystokinin and somatostatin in Alzheimer's disease postmortem cerebral cortex. Neurology 41, 716–719 (1991).

25.   Chen, X. et al. Cholecystokinin release triggered by NMDA receptors produces LTP and sound–sound associative memory. PNAS 116, 6397–6406 (2019).

26.   Jin, Z. et al. Identification and Characterization of Citrulline-modified Brain Proteins by Combining HCD and CID Fragmentation. Proteomics 13, 2682–2691 (2013).

27.   Thorsell, A. et al. Neurogranin in cerebrospinal fluid as a marker of synaptic degeneration in Alzheimer's disease. Brain Research 1362, 13–22 (2010).

28.   Philip, D. et al. Dihydrofolate reductase 19-bp deletion polymorphism modifies the association of folate status with memory in a cross-sectional multi-ethnic study of adults123. Am J Clin Nutr 102, 1279–1288 (2015).

29.   Cario, H. et al. Dihydrofolate Reductase Deficiency Due to a Homozygous DHFR Mutation Causes Megaloblastic Anemia and Cerebral Folate Deficiency Leading to Severe Neurologic Disease. The American Journal of Human Genetics 88, 226–231 (2011).

30.   Mouton-Liger, F. et al. CSF levels of the BACE1 substrate NRG1 correlate with cognition in Alzheimer's disease. Alzheimer's Research & Therapy 12, 88 (2020).

31.   Mitchell, R. M. et al. ErbB4 reduces synaptic GABAA currents independent of its receptor tyrosine kinase activity. Proc Natl Acad Sci U S A 110, 19603–19608 (2013).

32.   Mozhui, K. et al. Genetic regulation of Nrnx1 expression: an integrative cross-species analysis of schizophrenia candidate genes. Transl Psychiatry 1, e25 (2011).

33.   Nishikimi, A., Kukimoto-Niino, M., Yokoyama, S. & Fukui, Y. Immune regulatory functions of DOCK family proteins in health and disease. Experimental Cell Research 319, 2343–2349 (2013).

34.   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).

35.   Bik-Multanowski, M., Pietrzyk, J. J. & Midro, A. MTRNR2L12: A Candidate Blood Marker of Early Alzheimer's Disease-Like Dementia in Adults with Down Syndrome. J Alzheimers Dis 46, 145–150 (2015).

36.   Wen, H. et al. Neuroglobin mediates neuroprotection of hypoxic postconditioning against transient global cerebral ischemia in rats through preserving the activity of Na + /K + ATPases. Cell Death & Disease 9, 1–18 (2018).

37.   Hua, Y., Zhao, H., Lu, X., Kong, Y. & Jin, H. Meta-Analysis of the Cystatin C(CST3) Gene G73A Polymorphism and Susceptibility to Alzheimer's Disease. International Journal of Neuroscience 122, 431–438 (2012).

38.   Shin, J.-G. et al. Putative association of GPC5 polymorphism with the risk of inflammatory demyelinating diseases. Journal of the Neurological Sciences 335, 82–88 (2013).

39.   Zhou, Y., Zhang, X. & Klibanski, A. MEG3 noncoding RNA: a tumor suppressor. J Mol Endocrinol 48, R45-53 (2012).

40.   Koenning, M. et al. Myelin gene regulatory factor is required for maintenance of myelin and mature oligodendrocyte identity in the adult CNS. J Neurosci 32, 12528–12542 (2012).

41.   Ferrer, I. & Andrés-Benito, P. White matter alterations in Alzheimer's disease without concomitant pathologies. Neuropathol Appl Neurobiol 46, 654–672 (2020).

42.   Ousman, S. S. et al. Protective and therapeutic role for αB-crystallin in autoimmune demyelination. Nature 448, 474–479 (2007).

43.   Raudvere, U. et al. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res 47, W191–W198 (2019).

44.   Satoh, J., Tabunoki, H., Yamamura, T., Arima, K. & Konno, H. TROY and LINGO-1 expression in astrocytes and macrophages/microglia in multiple sclerosis lesions. Neuropathol Appl Neurobiol 33, 99–107 (2007).

45.   Khan, U. A. et al. Molecular drivers and cortical spread of lateral entorhinal cortex dysfunction in preclinical Alzheimer's disease. Nature Neuroscience 17, 304–311 (2014).

46.   R Core Team. R: A Language and Environment for Statistical Computing. https://www.r-project.org/ (2020).

47.   R Studio Team. RStudio: Integrated Development Environment for R. https://rstudio.com/ (2020).

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

49.   Bunis, D. G., Andrews, J., Fragiadakis, G. K., Burt, T. D. & Sirota, M. dittoSeq: universal user-friendly single-cell and bulk RNA sequencing visualization toolkit. Bioinformatics (2020) doi:10.1093/bioinformatics/btaa1011.

50.   Conway, J. R., Lex, A. & Gehlenborg, N. UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics 33, 2938–2940 (2017).

51.   McKenzie, A. T. et al. Brain Cell Type Specific Gene Expression and Co-expression Network Architectures. Scientific Reports 8, 1–19 (2018).

52.   Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 43, e47 (2015).

53.   Phipson, B., Lee, S., Majewski, I. J., Alexander, W. S. & Smyth, G. K. Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression. Ann Appl Stat 10, 946–963 (2016).

54.   Reimand, J. et al. Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. Nature Protocols 14, 482–517 (2019).

2.8 TABLES

Table 2.1: Prefrontal cortex cohort

ID	Sex	APOE	Age	Diagnosis	Batch
ROS32	Female	3/3	90	AD	3
ROS27	Female	3/4	90	AD	11
ROS33	Female	3/3	90	AD	5
ROS36	Female	3/3	90	AD	8
ROS28	Female	3/3	87	AD	10
ROS29	Female	3/4	76	AD	3
ROS34	Female	3/4	74	AD	2
ROS39	Male	3/3	89	AD	5
ROS45	Male	3/4	89	AD	1
ROS42	Male	3/3	87	AD	10
ROS41	Male	3/4	85	AD	4
ROS48	Male	3/4	86	AD	9
ROS43	Male	3/3	83	AD	4
ROS37	Male	3/3	86	AD	2
ROS44	Male	3/3	80	AD	8
ROS10	Female	3/3	90	Control	11
ROS8	Female	3/3	87	Control	9
ROS9	Female	3/3	87	Control	10
ROS6	Female	3/3	83	Control	6
ROS12	Female	3/3	81	Control	7
ROS3	Female	3/3	79	Control	3
ROS18	Male	3/3	90	Control	5
ROS14	Male	3/4	88	Control	1
ROS23	Male	3/3	87	Control	12
ROS16	Male	3/3	84	Control	4
ROS19	Male	3/3	80	Control	8
ROS13	Male	3/3	80	Control	1
ROS20	Male	3/3	80	Control	9
ROS15	Male	3/3	79	Control	2
ROS17	Male	3/3	76	Control	4

Table 2.2: Entorhinal cortex cohort

ID	Sex	APOE	Age	Diagnosis	Batch
AD1	Male	3/4	91	AD	AD1_AD2
AD2	Male	3/4	83.8	AD	AD1_AD2
AD4	Female	3/3	83.0	AD	AD3_AD4
AD6	Male	3/4	74.6	AD	AD5_AD6
Ct1	Female	3/3	67.3	Control	Ct1_Ct2
Ct2	Female	3/3	82.7	Control	Ct1_Ct2
Ct3	Male	3/3	72.6	Control	Ct3_Ct4
Ct4	Male	3/4	75.6	Control	Ct3_Ct4
Ct5	Male	3/3	77.5	Control	Ct5_Ct6

Supplementary Table 2.1: Prefrontal cortex cohort cell type composition

Cell Type	AD		AD (n)	Control		Ct (n)	Total
	APOE3/3	APOE3/4		APOE3/3	APOE3/4		
Astrocytes	859	352	1211	901	92	993	2204
Excitatory neurons	8198	3276	11474	9878	415	10293	21767
Inhibitory neurons	1708	998	2706	2838	208	3046	5752
Microglia	311	238	549	501	102	603	1152
Oligodendrocytes	4275	1459	5734	5290	292	5582	11316
Oligodendrocyte progenitor cells	562	275	837	760	43	803	1640
Total	15913	6598	22511	20168	1152	21320	43831

Supplementary Table 2.2: Entorhinal cortex cohort cell type composition

Cell Type	AD		AD (n)	Control		Ct (n)	Total
	APOE3/3	APOE3/4		APOE3/3	APOE3/4		
Astrocytes	124	257	381	1246	262	1508	1889
Microglia	46	56	102	234	26	260	362
Neurons	54	194	248	285	18	303	551
Oligodendrocytes	1080	2570	3650	1783	511	2294	5944
Oligodendrocyte progenitor cells	35	90	125	534	178	712	837
Total	1339	3167	4506	4086	995	5081	9587

2.9 FIGURES
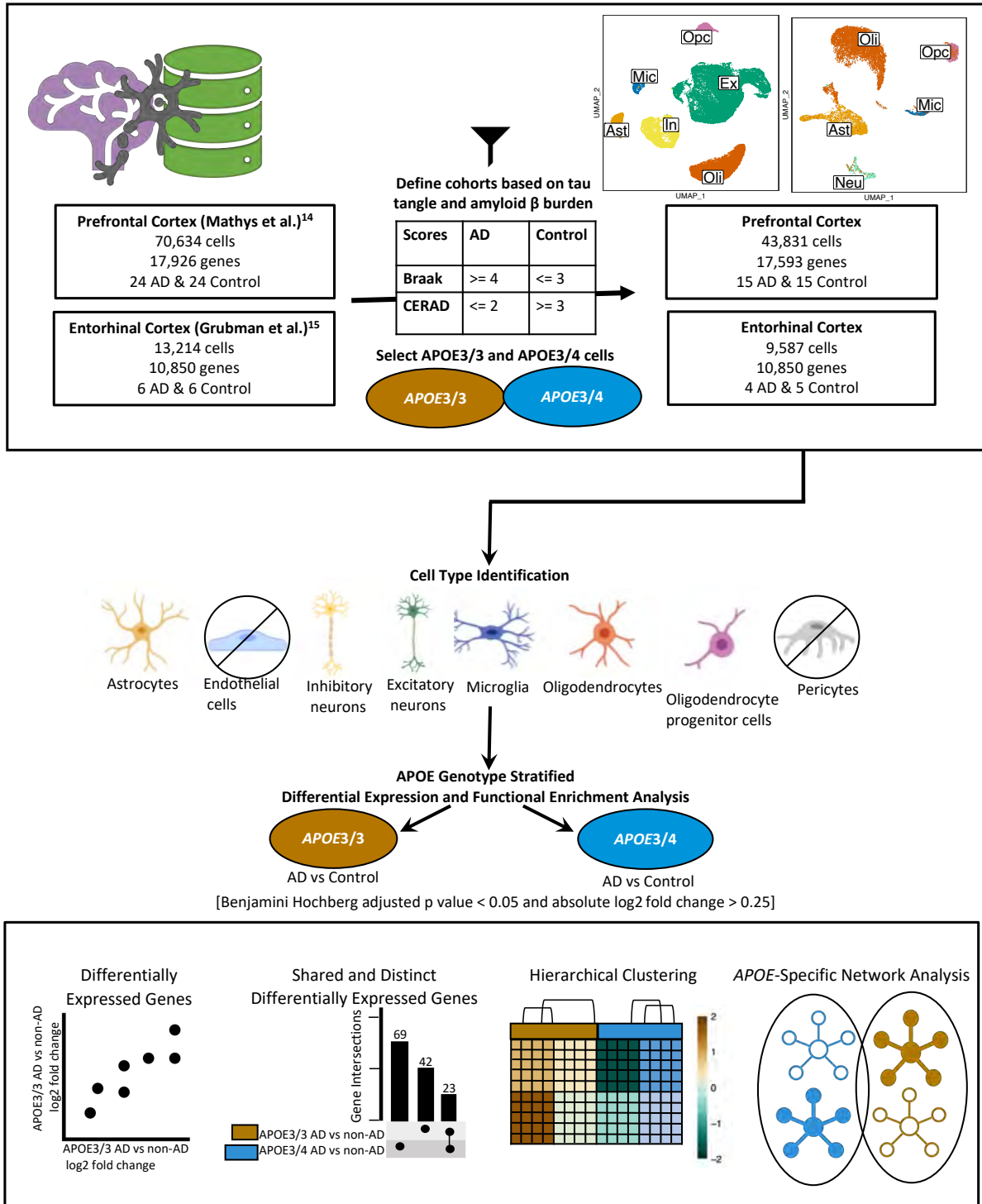
Figure 2.1: Overview of cohort sample definition and workflow for *APOE* genotype stratified cell type specific differential gene expression and functional enrichment. AD and non-AD cells were determined based on tau tangle (Braak) and amyloid β plaque (CERAD) burden. Cell types were identified, and AD versus non-AD differential expression and pathway network enrichment analyses were performed separately for APOE3/3 and APOE3/4 cells of each cell type.

a.

	APOE3/3 AD vs non-AD		APOE3/4 AD vs non-AD	
Ast	4	4	25	49
Ex	100	14	114	24
In	80	2	76	10
Mic	4	2	24	5
Oli	5	2	6	139
Opc	3	2	6	82
	Up	Down	Up	Down

Adjusted p-value (BH)
- <0.05 in APOE3/3 AD vs non-AD DEGs
- <0.05 in APOE3/4 AD vs non-AD DEGs
- <0.05 in Both APOE genotypes & Opposite Direction
- <0.05 in Both APOE genotypes & Same Direction
- Not Significant

b.

Gene	Cell type	*APOE*3/3 Log2 FC	*APOE*3/4 Log2 FC
CCK	Ex	-0.228	-0.664
CLU	Ast	0.268	1.032
	Ex	0.234	0.368
	In	0.255	0.578
NRGN	Ex	-0.036	-0.608
	In	0.309	0.460
	Oli	-0.086	-0.295
DHFR	Ex	1.063	0.415
	In	0.611	0.418
	Oli	-0.065	-0.473
ERBB4	Ex	0.466	0.627
NRXN1	Ast	-0.269	-0.305
	Oli	-0.091	-0.265

Figure 2.2: *APOE* genotype stratified cell type specific disease signatures in the prefrontal cortex. a. AD versus non-AD differentially expressed gene (DEG) counts for astrocytes (Ast), excitatory (Ex) and inhibitory (In) neurons, microglia (Mic), oligodendrocytes (Oli), and oligodendrocyte progenitor cells (Opc) in surveyed *APOE* genotypes. DEGs were selected using a Benjamini Hochberg adjusted p-value < 0.05 and absolute log2 fold change (FC) > 0.25, b. Subset of DEGs shared by both *APOE* genotypes and their corresponding log2 FC, c. Pairwise DEG plots of DEGs in APOE3/3 and APOE3/4 samples using log2 FC scores. Genes shown are significant and have a log2 FC > 0.25 in at least one *APOE* genotype. Colors indicate significance level of DEGs and whether DEGs are unique or shared by *APOE* genotypes, d. log2 FC scores of all genes in the DE analysis clustered by cell type and *APOE* genotype.

Figure 2.3: Shared and unique disease signatures across cell types in APOE3/3 and APOE3/4 prefrontal cortex samples. a. Upset plots indicating intersections of AD versus non-AD differentially expressed genes (DEGs) (Benjamini Hochberg (BH) adjusted p-value < 0.05 and absolute $\log_2$ fold change (FC) > 0.25) across cell types. Rows correspond to cell types. The bar chart shows the number of single and common sets of DEGs across cell types. Single filled dots represent a unique set of DEGs for the corresponding cell type. Multiple filled black dots connected by vertical lines represent common sets of DEGs across cell types, b. *LINGO1*, *RASGEF1B*, *NRXN1* and *CLU* expression. Asterisks represent meeting both significance (BH adjusted p-value < 0.05), and absolute $\log_2$ FC (> 0.25) thresholds. Colors correspond to *APOE* genotype and AD status.

75

a

	APOE3/3 AD vs non-AD		APOE3/4 AD vs non-AD	
	Up	Down	Up	Down
Ast	17	38	179	32
Mic	7	113	116	15
Neu	4	45	134	26
Oli	9	160	150	43
Opc	11	30	165	39

b

c

Samples · Genes

APOE: 3/3, 3/4

Celltype: Ast, Neu, Mic, Oli, Opc

Log2 FC: 3, 2, 1, 0, −1, −2, −3

Astrocytes · Microglia · Neurons · Oligodendrocytes · Oligodendrocyte progenitor cells

APOE3/4 AD vs non-AD Log2 FC

APOE3/3 AD vs non-AD Log2 FC

Adjusted p-value (BH)
- <0.05 in APOE3/3 AD vs non-AD DEGs
- <0.05 in APOE3/4 AD vs non-AD DEGs
- <0.05 in Both APOE genotypes & Opposite Direction
- <0.05 in Both APOE genotypes & Same Direction
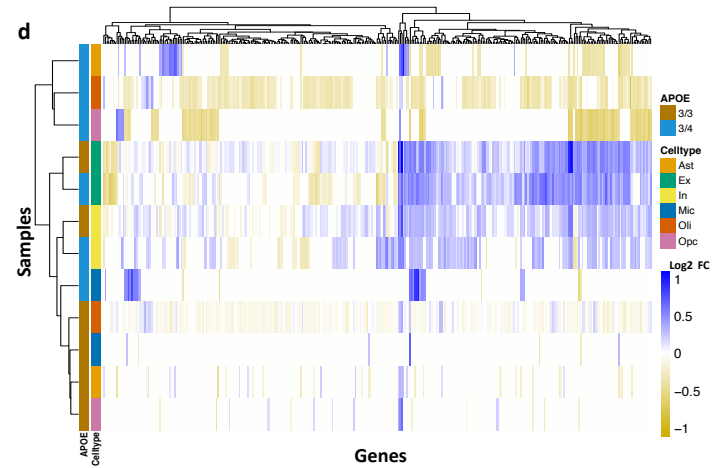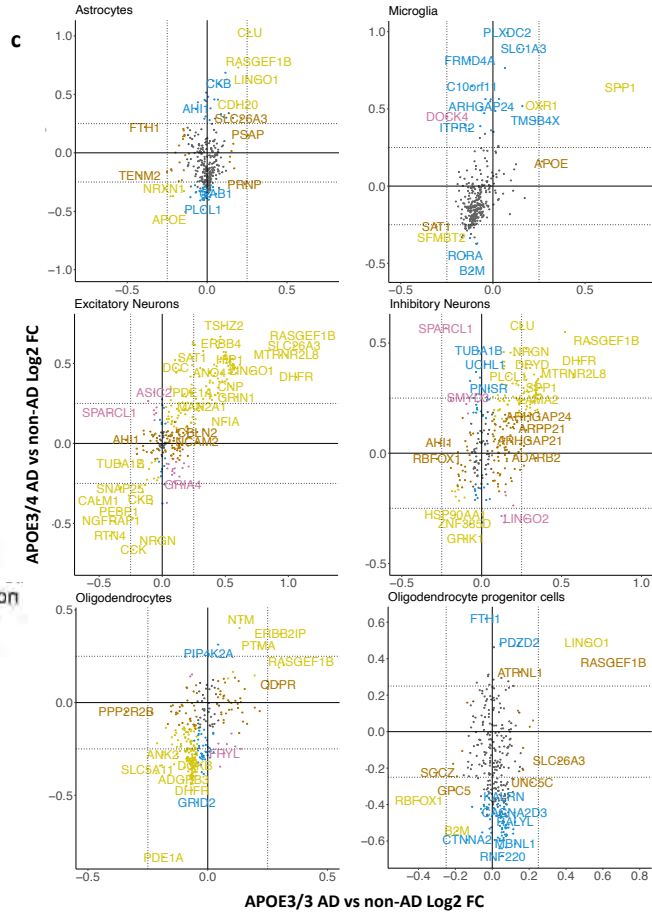- Not Significant

Figure 2.4: *APOE* genotype stratified cell type specific disease signatures in the entorhinal cortex. a. AD versus non-AD differentially expressed gene (DEG) counts for astrocytes (Ast, neurons (Neu), microglia (Mic), oligodendrocytes (Oli), and oligodendrocyte progenitor cells (Opc) in surveyed APOE genotypes. DEGs were selected using a Benjamini Hochberg adjusted p-value < 0.05 and absolute log2 fold change (FC) > 0.25, b. Pairwise DEG plots of DEGs in APOE3/3 and APOE3/4 samples using log2 FC scores. Genes shown are significant and have a log2 FC > 0.25 in at least one *APOE* genotype. Colors indicate significance level of DEGs and whether DEGs are unique or shared by *APOE* genotypes, c. log2 FC scores of all genes in the DE analysis clustered by cell type and *APOE* genotype
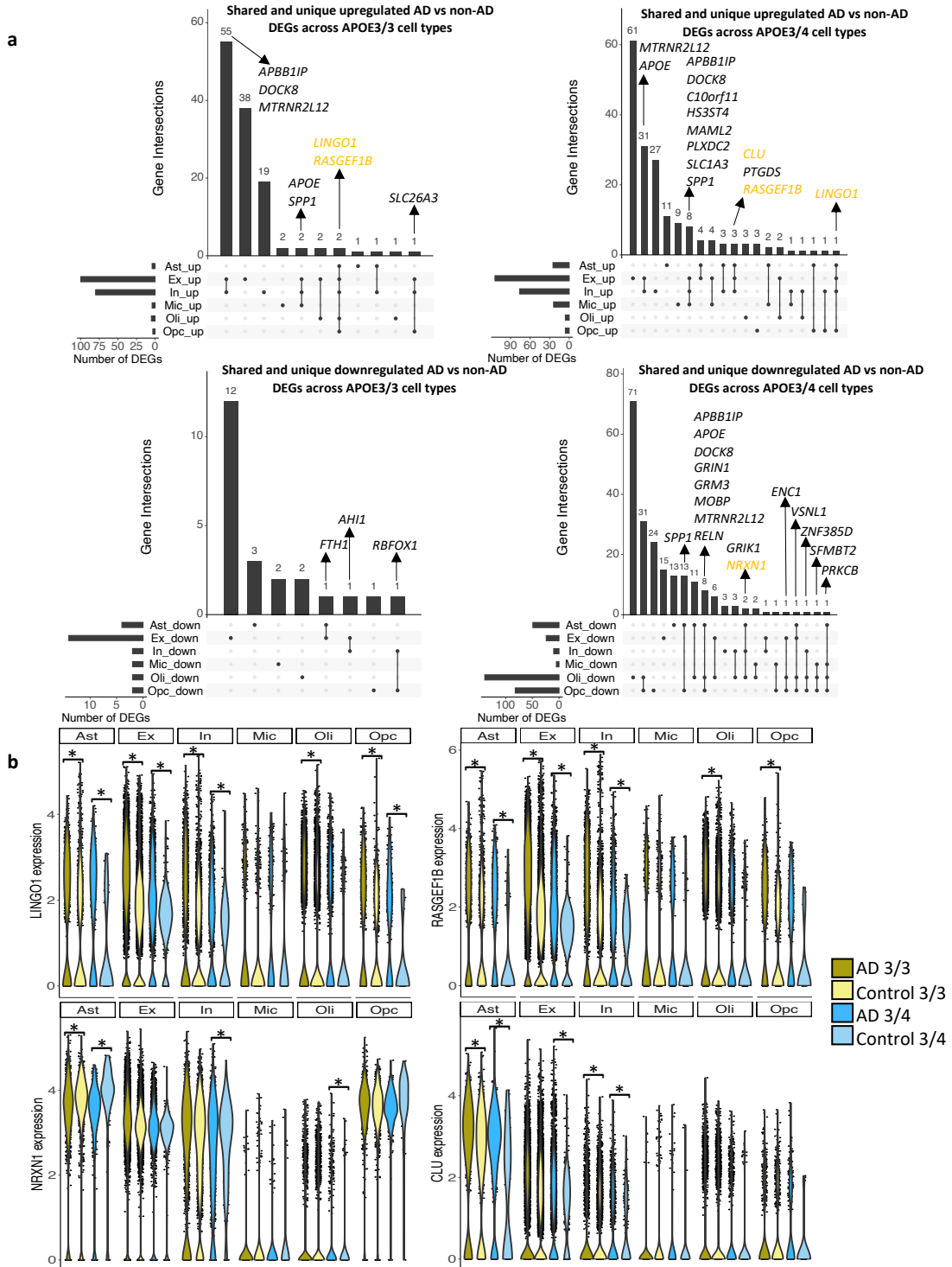
76
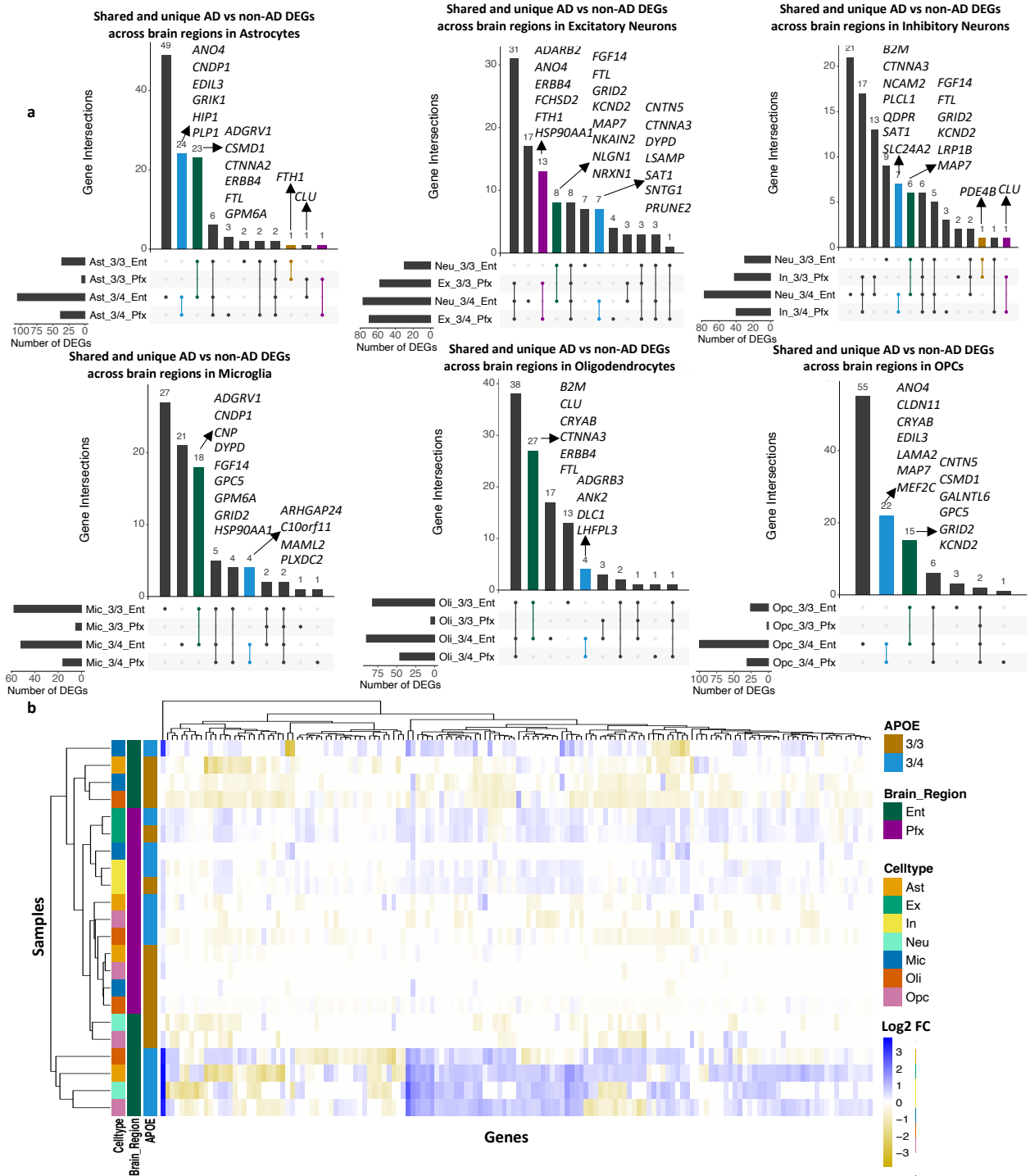
Figure 2.5: *APOE* **genotype stratified cell type specific disease signatures across brain regions.** a. Upset plots indicating intersections of AD versus non-AD differentially expressed genes (DEGs) (Benjamini Hochberg (BH) adjusted p-value < 0.05 and absolute log2 fold change (FC) > 0.25) within cell types across brain region and *APOE* genotype. Rows correspond to brain region and *APOE* genotype pairings. The bar chart shows the number of single and common sets of DEGs across brain regions and *APOE* genotype pairings. Single filled dots represent a unique set of DEGs for the corresponding brain region and *APOE* genotype pairing. Multiple filled black dots connected by vertical lines represent common sets of DEGs across brain region and *APOE* genotype pairings. Bar chart colors correspond to whether DEGs are shared between brain regions or APOE genotype using the bottom right key, b. log2 FC scores of all genes in the DE analysis of both brain regions clustered by cell type, brain region, and *APOE* genotype.
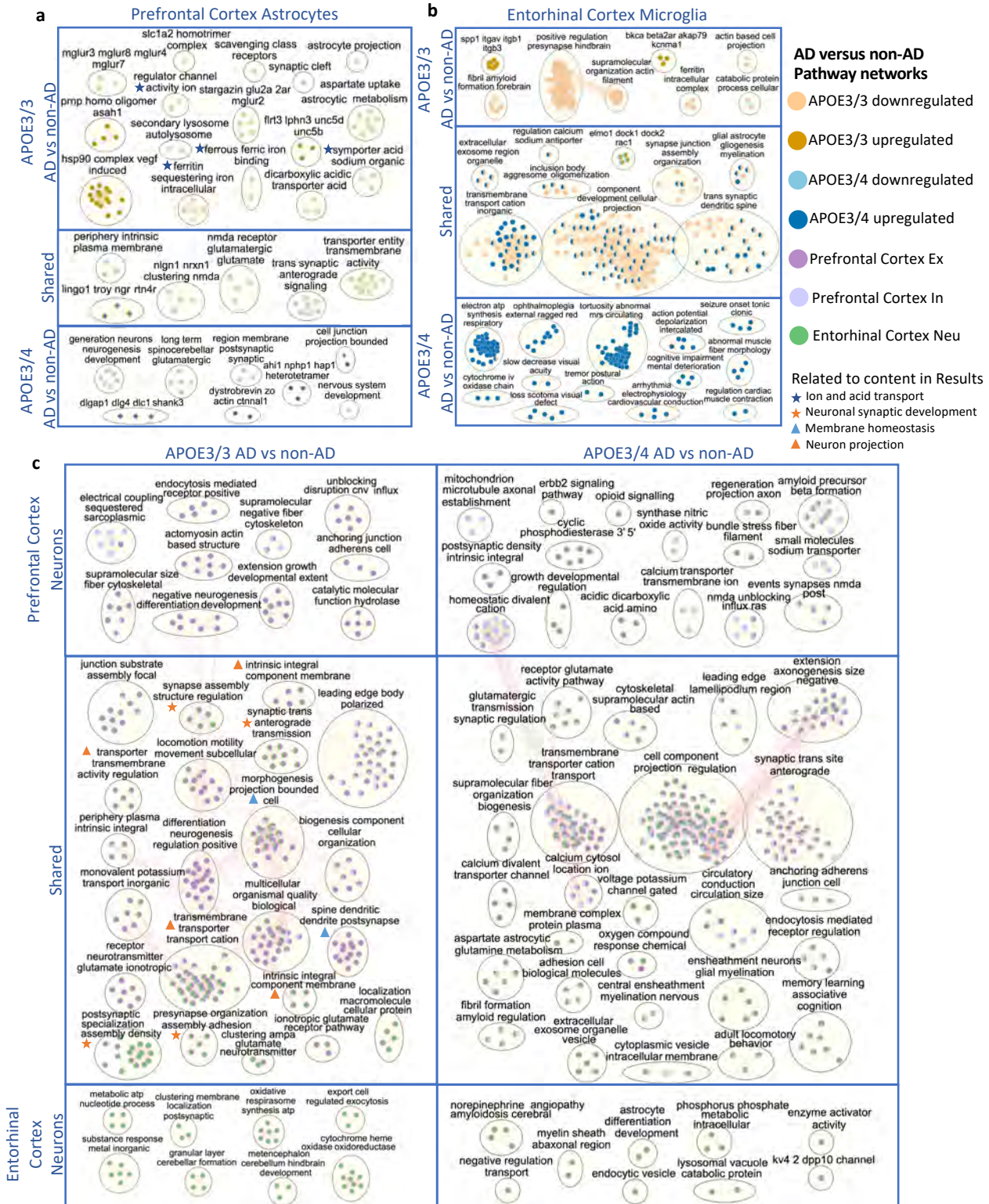
Figure 2.6: Enriched disease pathway networks in *APOE*3/3 and *APOE*3/4 cells. AD compared to non-AD functionally enriched pathways with a (Benjamini Hochberg (BH) adjusted p-value < 0.01 clustered into biological themes for: a. astrocytes of the prefrontal cortex, b. microglia of the entorhinal cortex, and c. prefrontal cortex excitatory (Ex) and inhibitory (In) neurons, and entorhinal cortex undistinguished neurons (Neu). Lines represent gene set overlaps with magnitude showed by thickness.
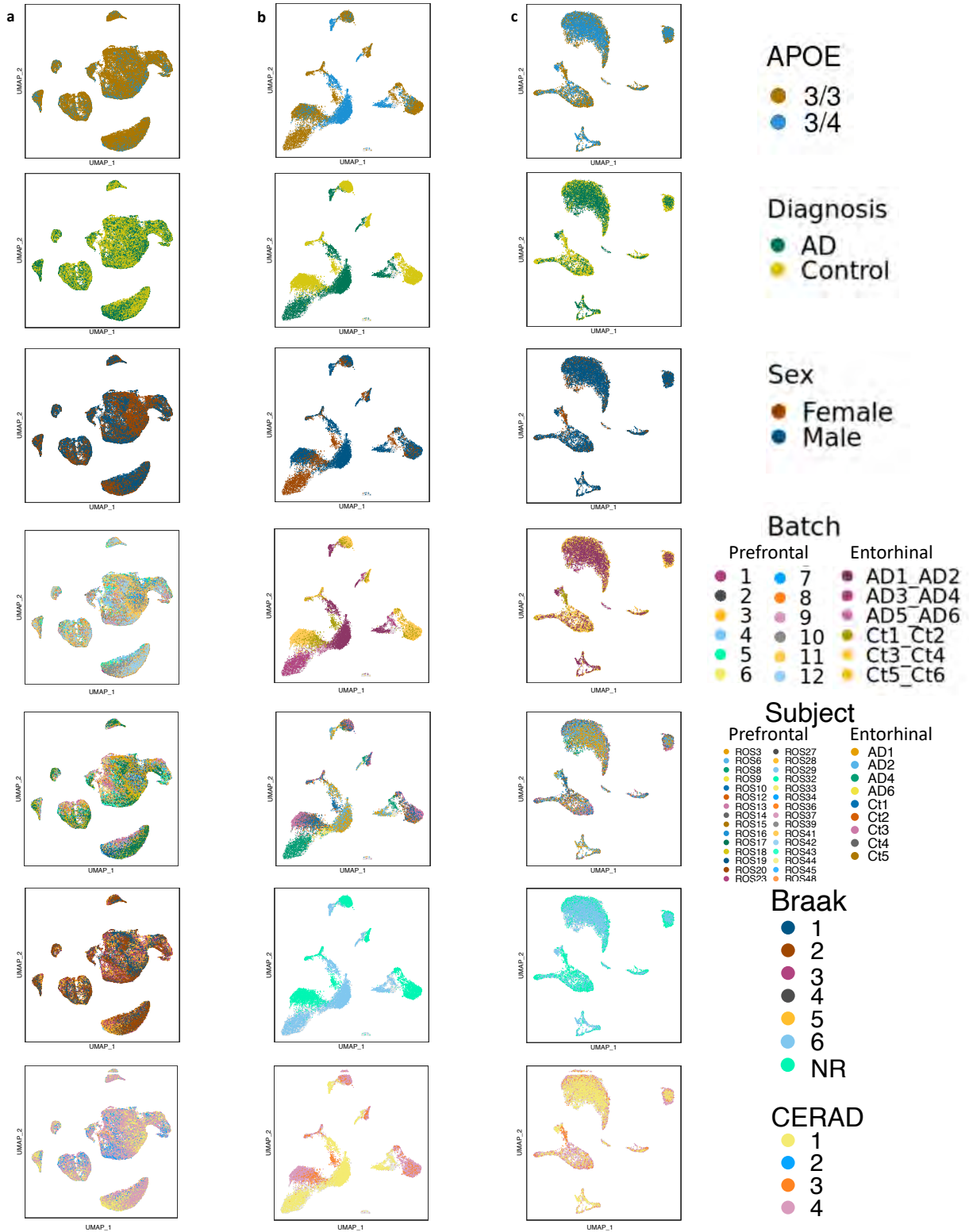
Supplementary Figure 2.1: Dimensionality reduction of prefrontal and entorhinal cortices cohort cells by *APOE* genotype, diagnosis, sex, batch, subject, Braak score, and CERAD score. Uniform Manifold Approximation and Projection (UMAP) of the a. prefrontal cortex cohort, b. entorhinal cortex cohort before batch correction, and c. entorhinal cortex cohort after batch correction. Each point represents a cell. NR: not reported.

Shared and unique *APOE*3/3 and *APOE*3/4 disease signatures within cell types in the prefrontal cortex
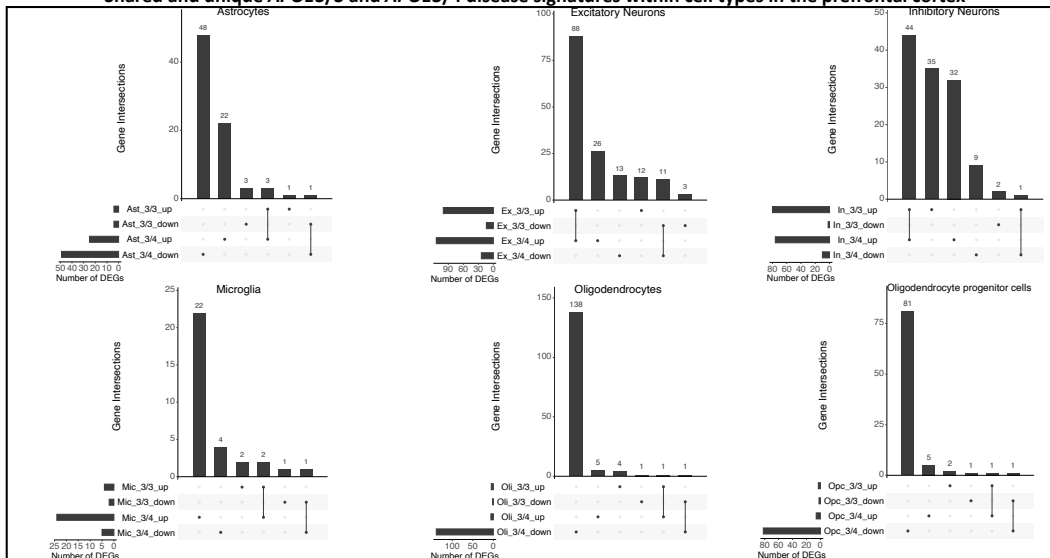
Shared and unique *APOE*3/3 and *APOE*3/4 disease signatures within cell types of male samples in the prefrontal cortex
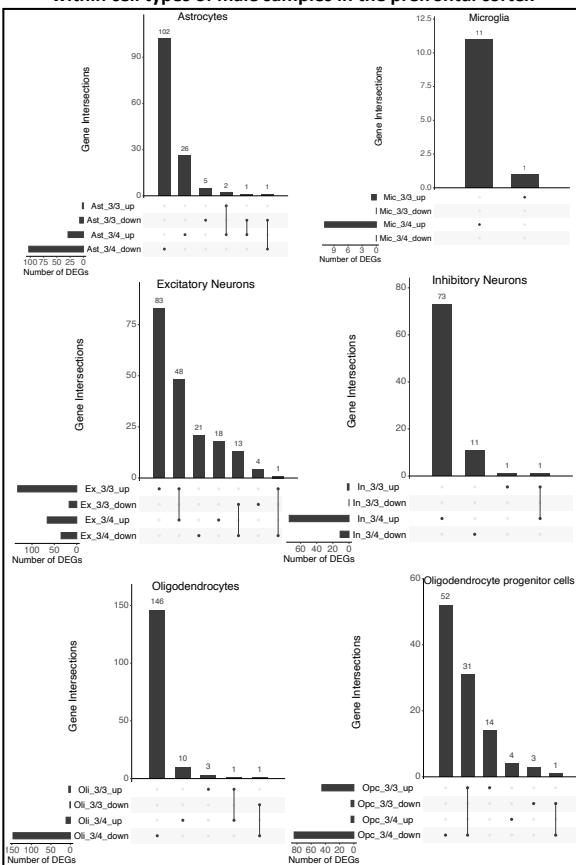
Shared and unique *APOE*3/3 and *APOE*3/4 disease signatures within cell types in the entorhinal cortex
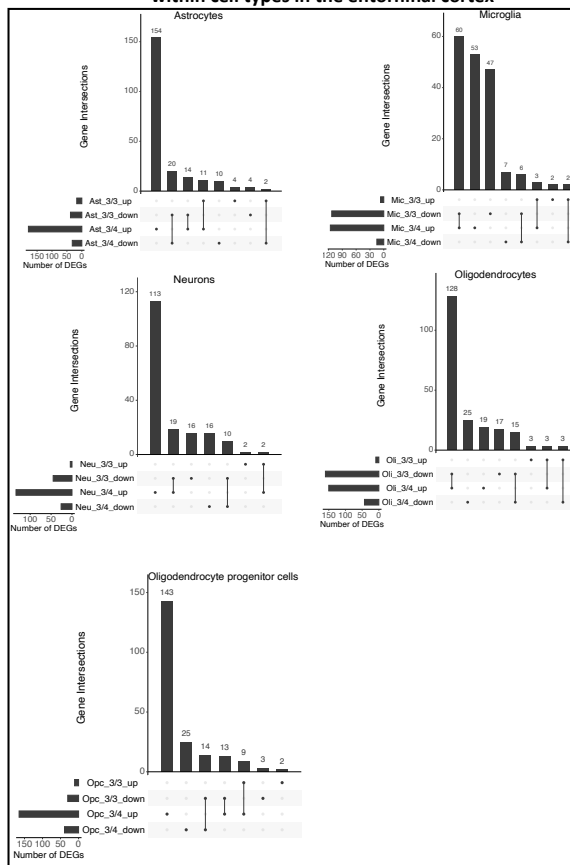
Supplementary Figure 2.2: Shared and unique disease signatures within cell types across APOE3/3 and APOE3/4 samples in the prefrontal and entorhinal cortices. Upset plots indicating intersections of AD versus non-AD differentially expressed genes (DEGs) (Benjamini Hochberg (BH) adjusted p-value < 0.05 and absolute log$_2$ fold change (FC) > 0.25) across *APOE* genotypes. Rows correspond to cell type and *APOE* genotype pairings. The bar chart shows the number of single and common sets of DEGs across *APOE* genotypes. Single filled dots represent a unique set of DEGs for the corresponding *APOE* genotype. Multiple filled black dots connected by vertical lines represent common sets of DEGs across *APOE* genotypes.
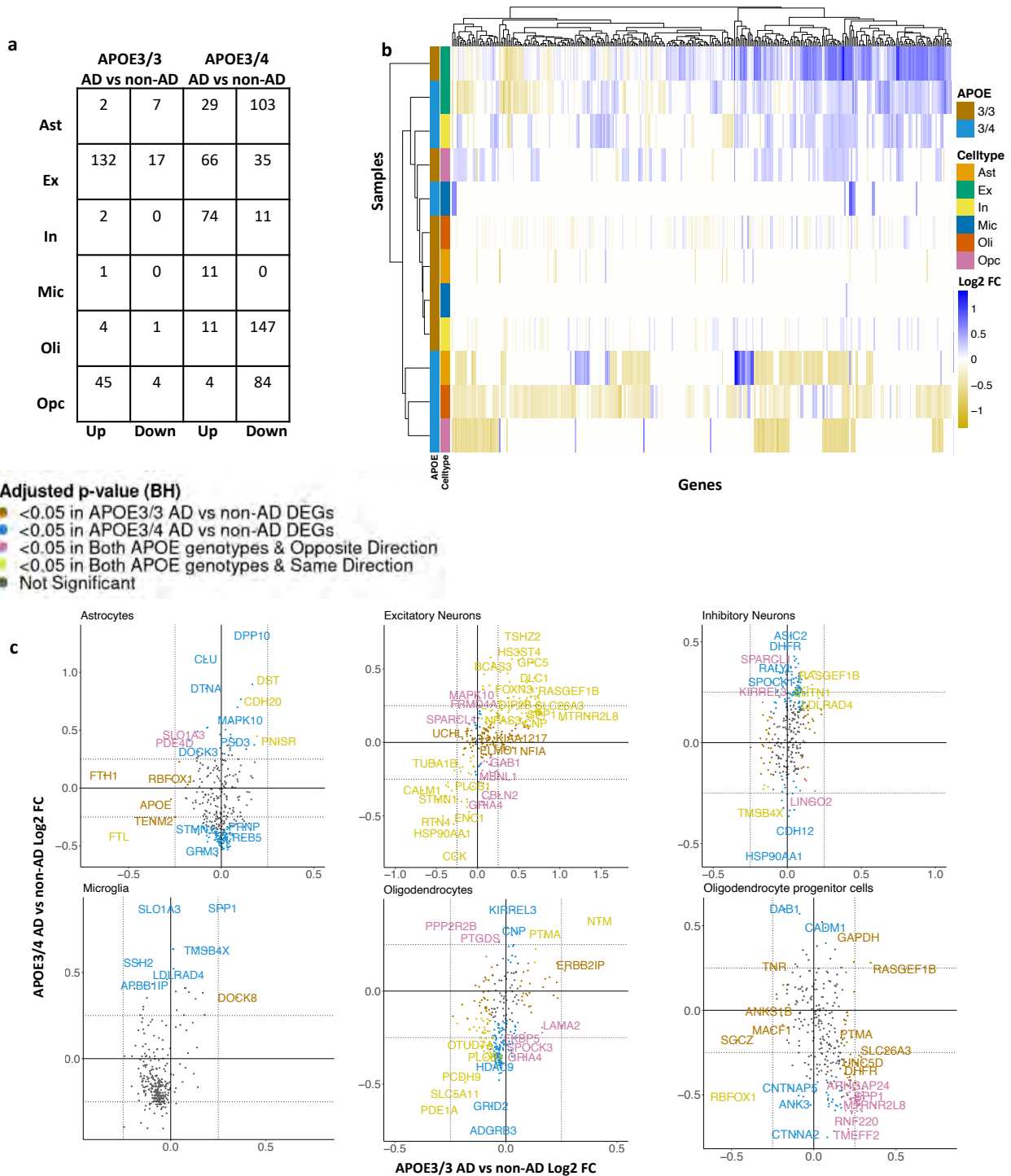
Supplementary Figure 2.3: *APOE* genotype stratified cell type specific disease signatures in male prefrontal cortex. a. AD versus non-AD differentially expressed gene (DEG) counts for astrocytes (Ast), excitatory (Ex) and inhibitory (In) neurons, microglia (Mic), oligodendrocytes (Oli), and oligodendrocyte progenitor cells (Opc) in surveyed *APOE* genotypes. DEGs were selected using a Benjamini Hochberg adjusted p-value < 0.05 and absolute log2 fold change (FC) > 0.25, b. log2 FC scores of all genes in the DE analysis clustered by cell type and *APOE* genotype, c. Pairwise DEG plots of DEGs in APOE3/3 and APOE3/4 samples using log2 FC scores. Genes shown are significant and have a log2 FC > 0.25 in at least one *APOE* genotype. Colors indicate significance level of DEGs and whether DEGs are unique or shared by *APOE* genotypes.
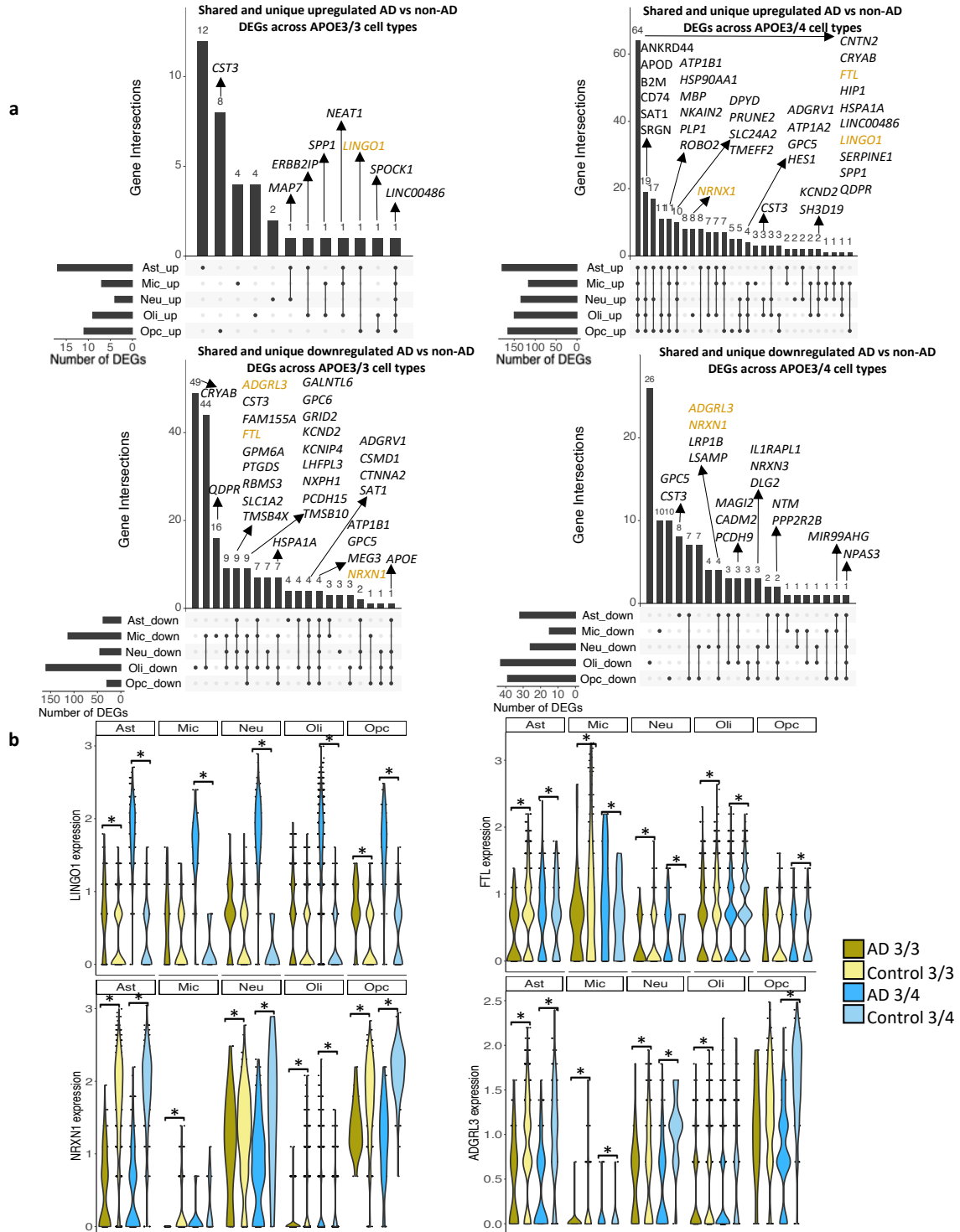
Supplementary Figure 2.4: Shared and unique disease signatures across cell types in APOE3/3 and APOE3/4 entorhinal cortex samples. a. Upset plots indicating intersections of AD versus non-AD differentially expressed genes (DEGs) (Benjamini Hochberg (BH) adjusted p-value < 0.05 and absolute log2 fold change (FC) > 0.25) across cell types. Rows correspond to cell types. The bar chart shows the number of single and common sets of DEGs across cell types. Single filled dots represent a unique set of DEGs for the corresponding cell type. Multiple filled black dots connected by vertical lines represent common sets of DEGs across cell types, b. *LINGO1*, *FTL*, *NRXN1* and *ADGRL3* expression. Asterisks represent meeting both significance (BH adjusted p-value < 0.05), and absolute log$_2$ FC (> 0.25) thresholds. Colors correspond to *APOE* genotype and AD

82
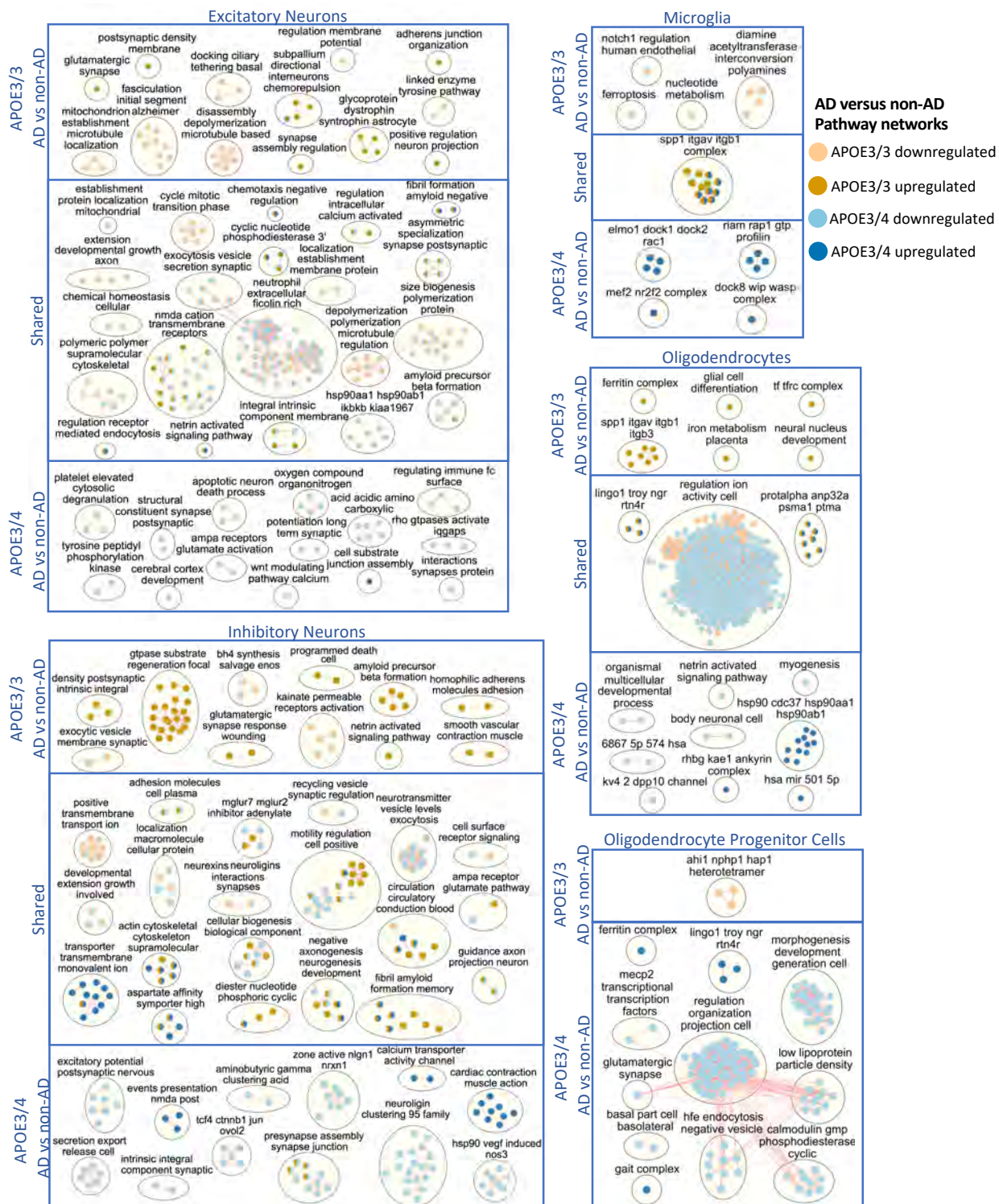
Supplementary Figure 2.5: Enriched disease pathway networks in *APOE*3/3 and *APOE*3/4 prefrontal cortex cells. AD compared to non-AD functionally enriched pathways with a (Benjamini Hochberg (BH) adjusted p-value < 0.01 clustered into biological themes for excitatory and inhibitory neurons, microglia, oligodendrocytes, and oligodendrocyte progenitor cells . Lines represent gene set overlaps with magnitude showed by thickness.
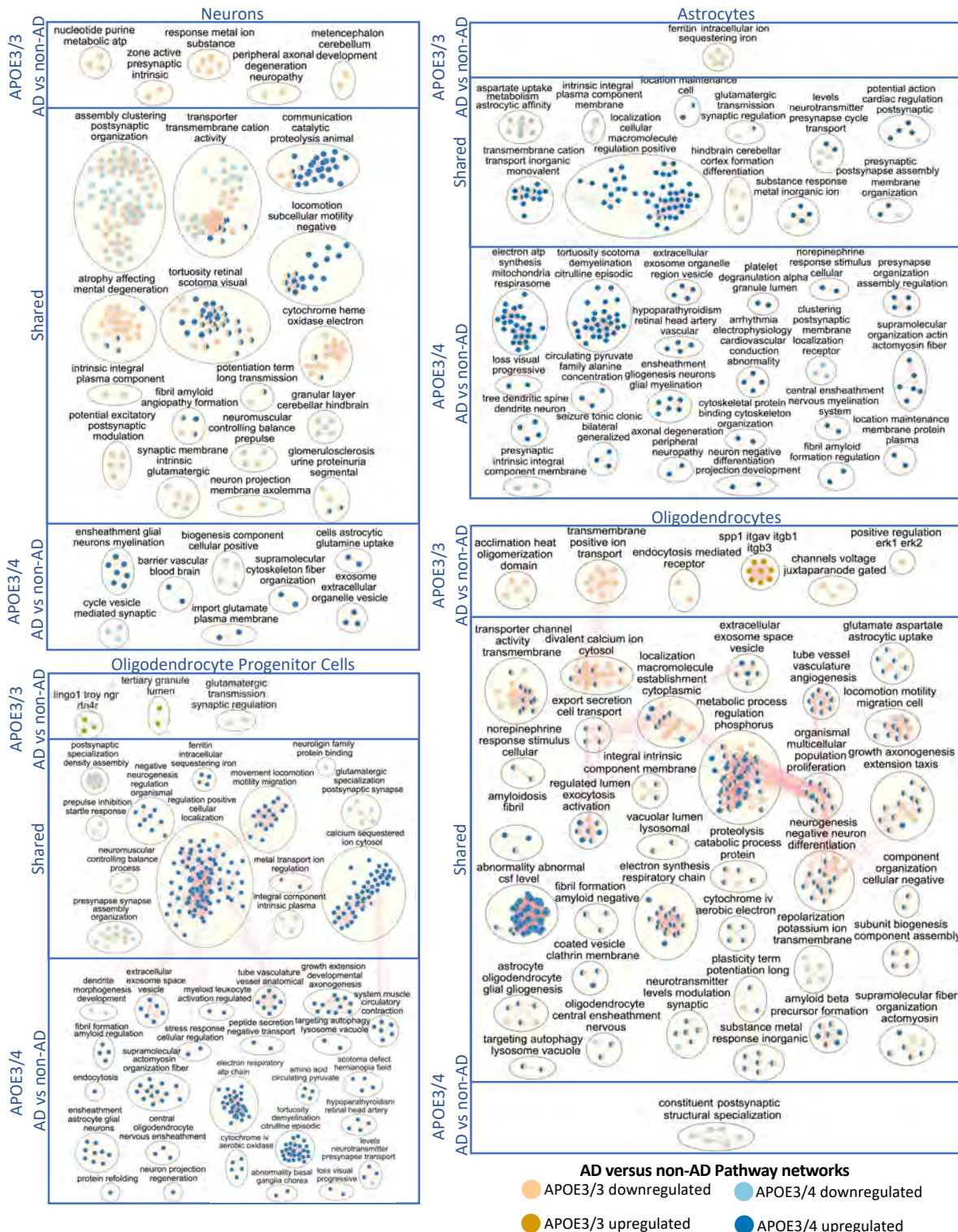
Supplementary Figure 2.6: Enriched disease pathway networks in *APOE*3/3 and *APOE*3/4 entorhinal cortex cells. AD compared to non-AD functionally enriched pathways with a (Benjamini Hochberg (BH) adjusted p-value < 0.01 clustered into biological themes for neurons, oligodendrocyte progenitor cells, astrocytes, and oligodendrocytes. Lines represent gene set overlaps with magnitude showed by thickness.

CHAPTER 3

Sex-stratified single-cell RNA-Seq analysis identifies sex-specific and cell type-specific transcriptional responses in Alzheimer's disease across two brain regions.

3.1 ABSTRACT

Alzheimer's disease (AD) is a pervasive neurodegenerative disorder that disproportionately affects women. Since neural anatomy and disease pathophysiology differ by sex, investigating sex-specific mechanisms in AD pathophysiology can inform new therapeutic approaches for both sexes. Here, we utilized nearly 74,000 cells from human prefrontal and entorhinal cortex samples from the first two publicly available single-cell RNA sequencing AD datasets to study cell type-specific sex-stratified transcriptomic perturbations in AD. Our examination at the single-cell level revealed that sex-specific gene and pathway differences in AD were most prominently observed in glial cells of the prefrontal cortex. In the entorhinal cortex, we observed the same genes and pathways to be perturbed in opposing directions between sexes in AD relative to healthy state. Our findings contribute to growing evidence of sex differences in AD-related transcriptomic changes, which can fuel the development of therapies that may prove more effective at reversing AD pathophysiology.

3.2 INTRODUCTION

Alzheimer's disease (AD) is an irreversible neurodegenerative disorder that causes progressive memory decline, cognitive deficits, and behavioral changes[1–3]. It is the most common form of dementia and is reaching epidemic proportion as a result of extended life expectancies and increased elderly populations worldwide[4,5]. It is of high priority to find disease-modifying treatments for AD, as more than five million people are diagnosed with AD currently in the United States, a number estimated to triple by 2050[6,7].

Although first described more than a century ago[8], the underlying molecular mechanisms of AD remain elusive[9]. Extensive research efforts reveal that AD is histologically characterized by pathological brain aggregates including extracellular amyloid-β (Aβ) plaques, and intracellular tau protein neurofibrillary tangles (NFTs)[10,11]. Increasing evidence suggests that neuroinflammation and brain dysfunction led by neuronal supporting cells, which include microglia, astrocytes, and oligodendrocytes, could contribute to AD pathophysiology[12,13]. These pathological features are accompanied by impaired neurotransmitter signaling, dysregulated neuronal metabolism, neuronal loss, and cerebral atrophy[14–16]. Overall, the exact pathogenesis of AD remains uncertain, which hinders the development of effective therapies.

Sex differences have been clinically documented in AD[17,18], yet the underlying cause for these differences are not well understood. Approximately two thirds of AD diagnoses are in women[19]. In addition to greater longevity in females[20], other biological differences may be responsible for the higher prevalence and accelerated cognitive decline observed in women during disease progression[18,21,22]. For instance, a longitudinal study examining a postmortem cohort of about

1,500 individuals observed that in the presence of similarly high Aβ burden, females exhibited faster cognitive decline than males[22], suggesting females might be more susceptible to Aβ toxicity. Furthermore, after adjusting for age and education, women had a higher tau tangle density[22,23]. Among genetic risk factors implicated in AD, the apolipoprotein E (*APOE*) ε4 risk allele has been observed to have a differential influence and increased risk for AD in women compared to men[24,25]. Sex hormones, especially the decline in hormone levels post-menopause, could also contribute to sex differences in AD progression. For example, after menopause, women experience an abrupt loss of progesterone[26], which was previously shown to be neuroprotective by promoting myelin repair and reducing inflammation[27,28]. In fact, compared to men, women experience more inflammation-driven symptoms and have an increased risk for autoimmune diseases[29–31]. These findings suggest that investigating sex differences in AD will not only provide insight into deciphering the fundamental biological and mechanistic causes of AD pathogenesis, but also highlight the necessity of developing personalized therapeutic strategies.

Previous studies suggest that cellular and molecular heterogeneity in AD pathogenesis[32,33] and brain immune cell dysfunction contribute to sex-specific AD pathophysiology[34]; however, sex-specific disease complexity at single-cell resolution is masked in bulk brain sequencing analysis. Recent advances in single-cell RNA sequencing technology and the increasing availability of human transcriptomic datasets present a novel opportunity to examine cell type-specific transcriptional alterations in AD brain pathology. In recent years, two single-nucleus RNA-Seq (snRNA-Seq) datasets were generated from the prefrontal[35] and entorhinal[36] cortices of age and sex-matched human AD patients and cognitively normal controls. For the prefrontal cortex

dataset, Mathys and colleagues performed differential expression analysis on single-cell transcriptomic results across 48 individuals of varying degrees of AD pathology and reported on the general sexual dimorphic transcriptional response to AD pathology; however, they did not extensively examine sex-specific DEGs in the individual brain cell types or delineate any subsequent sex-specific molecular pathway enrichments in AD. Similar to the Mathys analysis, Grubman and colleagues analyzed single-nuclei transcriptomes sequenced from the entorhinal cortex of 12 age and sex-matched human AD patients and controls. Besides investigating the likelihood of sex as a covariate factor for DEG variance observed, no sex difference analysis was performed in this study.

Understanding gene expression changes unique to each sex provides opportunities to decipher molecular underpinnings that differentially contribute to AD in males and females. In this study, we leveraged these two snRNA-Seq datasets to characterize sex-stratified cell type-specific gene expression perturbations in AD and to identify sex-specific disease-associated cellular pathways as potential precision therapeutic targets. In both brain regions, we identified sex-specific disease changes primarily in glial cells and observed samples to cluster by sex when examining gene expression changes in AD compared to controls. Our findings will be of fervent interest to the field in studying differing vulnerabilities between sexes in AD.

3.3 RESULTS

3.3.1 Sample classification and analytic workflow

Samples were categorized into cases and controls based on tau tangle and Aβ plaque burdens, using Braak clinical staging and Consortium to Establish a Registry for Alzheimer's Disease

88

(CERAD) scores[37], respectively (AD: Braak stage $\geq$ IV, CERAD score $\leq$ 2; Control: Braak stage $\leq$ III, CERAD score $\geq$ 3). This resulted in single-nucleus RNA-Seq datasets containing 17,723 genes expressed by 62,741 cells from the prefrontal cortex cohort (**Table 3.1**), and 10,846 genes expressed by 11,284 cells from the entorhinal cortex cohort (**Table 3.2**), which were acquired from different sets of individuals (**Figure 3.1**). In both brain regions, a sex-stratified differential gene expression (DGE) analysis was performed comparing AD cases to controls, with *APOE* genotype as a covariate, in astrocytes (Ast), microglia (Mic), excitatory neurons (Ex), inhibitory neurons (In), undifferentiated neurons (Neu), oligodendrocytes (Oli), and oligodendrocyte progenitor cells (OPCs) (**Supplementary Tables 3.1 and 3.2**). For the entorhinal cortex cohort, data integration was performed and *APOE* genotype was included as a sole covariate in our DGE analysis to account for batch effects and avoid collinearity in our model. Differentially expressed genes (DEGs) were determined using a Benjamini-Hochberg adjusted p-value < 0.05 and absolute log2 fold change (LFC) > 0.25 as cutoffs. DEGs were passed as inputs for pathway enrichment analysis, which provided pathways to be used as inputs for subsequent network analysis. We examined gene expression and pathway network differences in AD versus neurotypical cells to identify cell type- and brain region- specific and non-specific differences based on sex.

3.3.2 Sex-stratified DGE analysis in the prefrontal cortex reveals sex-specific disease-related changes in glial cell types

Leveraging data from Mathys et al., from our sex-stratified DGE analysis, we identified DEGs meeting significance and LFC thresholds (**Table 3.3**) in all cell types except male inhibitory neurons when comparing AD to non-AD (**Supplementary Table 3.3**). We identified 73 DEGs

across all cell types in the prefrontal cortex (**Table 3.3, Supplementary Table 3.3**). Of these

DEGs, 36 were shared in both sexes, while 8 and 29 were specific to AD compared to control

males and females, respectively. We also observed more shared DEGs in AD case versus control

female signatures versus male signatures across the cell types (**Fig. 3.2a**), which is consistent

with previous bulk tissue analysis[34]. Some of the DEGs that overlap most across cell types within

one sex or across sexes include *LINGO1*, a negative regulator of myelination[38,39], which we

found upregulated in all AD compared to control female cell types; *SLC1A3*, which encodes

excitatory amino acid transporter 1 that transports glutamate in the synaptic cleft[40] and was

perturbed in all female AD compared to control cell types except oligodendrocytes and OPCs;

and *SPP1*, a protein involved in neuroinflammation also known as Osteopontin[41] that we

observed to be upregulated in AD versus control samples of both female and male excitatory

neurons and microglia, as well as female astrocytes and inhibitory neurons. Also, clustering

samples by AD compared to control pseudo-bulk cell type gene expression (**Fig. 3.2b**) showed

samples to cluster by sex before cell type identity for all cell types except excitatory neurons.

In addition to identifying shared DEGs across cell types and sexes, we also observed a larger

range of LFC in the analysis of female AD versus control ([-0.423, 1.058], median=0.314)

compared to the analysis of male AD versus control ([-0.370, 0.620], median=0.343). Within

each cell type, we observed DEGs, a number of which are relevant to and have been studied in

AD (e.g. *NRXN1*[42], *SPP1*[41], *DHFR*[43], *SGK1*[44], *ERBB2IP*[45]), meeting significance and LFC

thresholds. These DEGs are shared by both sexes in AD versus control astrocytes, microglia, and

excitatory neurons, with consistent directionality in both sexes (**Fig. 3.2c; Fig. 3.2d, yellow**

color; Supplementary Figure 3.3). Overall, in the prefrontal cortex, we identified sex-distinct

disease-related transcriptomic changes in gene expression primarily among glial cells (**Fig. 3.2d,**
brown color for female-distinct and blue color for male-distinct).

3.3.3 Sex-stratified DGE analysis in the entorhinal cortex reveals sex-specific disease-related
changes, including opposite transcriptomic changes between sexes

Leveraging data from Grubman et al, we identified DEGs (**Table 3.4)** comparing AD to non-AD
in all cell types stratified by sex. We identified 232 DEGs across all cell types in the entorhinal
cortex (**Table 3.4, Supplementary Table 3.4).** Of these DEGs, 211 were shared in both sexes,
while 20 and 1 were specific to AD compared to control males and females, respectively. We
observed shared DEGs across cell types when comparing AD versus control samples in both
sexes (**Fig. 3.3a**). Some of the DEGs that overlap most across cell types within one sex or across
sexes include *CLU*[9,46], *HSPA1A*[47], *RBFOX1*[48], and *CST3*[49]*,* which are relevant in AD
progression. Clustering of samples by AD compared to control pseudo-bulk cell type-specific
gene expression (**Fig. 3.3b**) showed samples to cluster by sex before cell type identity for every
cell type and highlighted opposing gene expression patterns based on sex. Indeed, interestingly,
186 of the 211 DEGs shared between male and female AD were regulated in opposite directions
with respect to controls, at least in some cell types.

When comparing the magnitude of gene expression changes across sexes in AD versus control
samples, we found males to have a greater range of LFCs ([-2.174, 3.461], median=0.567)
compared to females ([-1.657, 2.649], median= -0.436). We visualized these differences in DEGs
such as *LINGO1,* which had a higher fold change difference in male astrocytes (3.415) compared
to female astrocytes (0.4); *GPM6A*, which was upregulated in male oligodendrocytes and

downregulated in female oligodendrocytes; *CST3*, which was upregulated in male neurons, male oligodendrocytes, and male and female OPCs, and downregulated in female neurons, female oligodendrocytes, and male and female astrocytes; and *LINC00486*, which was upregulated in all cell types of both sexes with an average LFC in males of 1.9 compared to 1.0 in females (**Fig. 3.3c**). Generally, directly comparing AD vs control DEGs within each cell type, we not only observe a subset of genes with directionally consistent changes among males and females (**Fig. 3.3d, yellow color; Supplementary Figure 3.3**), but we also observed numerous changes in opposing directions across sexes (**Fig. 3.3d, pink color; Supplementary Figure 3.3**), and a higher magnitude of disease-related changes in males compared to females.

3.3.4 Comparative analysis across brain regions reveals more shared transcriptomic sex differences in the entorhinal cortex

We compared DEG results from the prefrontal and entorhinal cortices to determine whether changes in each sex were consistent across brain regions. Overall, we observed more overlaps across sex DEGs to be in the entorhinal cortex (**Fig. 3.4a**). Additionally, clustering samples by AD compared to control pseudo-bulk cell type gene expression (**Fig. 3.4b**) showed some clustering by brain region and sex.

3.3.5 Pathway and network analysis reveals sex-specific transcriptomic perturbations in glial cells in the prefrontal cortex and sex-shared, but flipped AD-enriched pathways in the entorhinal cortex

Beyond identifying sex-dimorphic disease-associated genes, we performed a gene set enrichment analysis to elucidate potential biological mechanisms implicated in disease progression that are

either shared or unique to each sex and to reveal the interconnections between disease-linked pathways within AD. The pathway enrichment was performed in g:Profiler[50], a web tool that performs functional enrichment analysis from a given gene list, using separate lists of upregulated and downregulated DEGs with an adjusted p-value <0.05 and relaxed absolute LFC above 0.1 in cell types of each sex as inputs. Significantly enriched biological pathways with an adjusted p-value < 0.05 were applied to EnrichmentMap[51], a functional category grouping method from the Cytoscape software, to identify pathway network clusters annotated by associated biological processes (**Fig. 3.5, Supplementary Figures 3.3 and 3.4**).

Female and male AD compared to control excitatory neurons of the prefrontal cortex shared six common enriched clusters of pathways (**Fig. 3.5a**), which were all perturbed in the same direction for both sexes. Two of these clusters (neurotransmitter glutamate/aspartate transmembrane activity and carboxylic acid biosynthetic process) were upregulated in disease in both sexes. Of the four downregulated pathway clusters, three were related to synaptic activity (modulation of the synaptic membrane, neurotransmitter release, and synapse assembly/cell junction organization), indicating a dysregulation of synaptic plasticity in AD excitatory neurons. The other downregulated pathway cluster was plasma membrane morphogenesis, which consisted of pathways including axonogenesis, cellular projection, and plasma membrane organization. (**Supplementary Tables 3.5 and 3.6**).

In prefrontal cortex excitatory neurons, we also identified uniquely enriched disease pathway clusters for each sex (**Fig. 3.5a**). Female excitatory neurons showed upregulation of the HOXA5 factor, a DNA-binding transcription factor that regulates cell morphogenesis and tumor

suppressor that inhibits proliferation and induces apoptosis[52], and downregulation of inflammatory-mediated cell to cell interaction through adhesion and molecule binding. Interestingly, a recent epigenome-wide association study examining samples in the prefrontal cortex and superior temporal gyrus observed elevated DNA methylation of the *HOXA* gene cluster to be associated with neuropathology in AD[53]. In male excitatory neurons, we observed upregulation of axon regeneration, and downregulation of distal axonal growth cone polarization. Interestingly, we also observed downregulation of tetrahydrobiopterin (BH4) synthesis, which is important for the production of essential neurotransmitters[54], and Rho GTPase activities in male AD compared to control excitatory neurons. Overall, excitatory neurons of the prefrontal cortex shared most case vs control differentially enriched pathways between male and females, the majority of which were downregulated in AD.

Like enriched pathways in disease observed in excitatory neurons, the inhibitory neurons of the prefrontal cortex showed upregulation for glutamate/aspartate activities in both female and male AD inhibitory neurons compared to controls (**Fig. 3.5b**). Like male AD excitatory neurons, male AD inhibitory neurons also showed downregulation of axonal growth cone polarization and BH4 activities compared to controls. In addition, males specifically demonstrated upregulation in anterograde synaptic transmission and downregulation of nitric synthase, heat shock protein 90 (HSP90) complex, voltage potassium transporter, and kainite calcium-permeable receptors activities in AD. The *ITGAV-ITGB-SPP1* complex, with known function in cell adhesion[55] and without previous links to AD, was uniquely upregulated in male inhibitory neurons. Of note, the pathway cluster neuronal projection was upregulated in females and downregulated in males, consistent with the enriched upregulated pathways clusters uniquely observed in females, which

were modulation of spine morphogenesis and synaptic membranes. Lastly, the transcription factors, nuclear receptor TLX (essential for the regulation of self-renewal, neurogenesis and maintenance in neuron stem cell)[56] and nuclear protein HOXB2 (involved in cellular development)[57], were upregulated only in AD female inhibitory neurons.

Unlike in neurons in the prefrontal cortex, we identified a variety of commonly enriched disease pathway networks in entorhinal cortex neurons that were regulated in opposite directions for the sexes (**Fig. 3.5c**). For instance, amyloid-beta binding/fibril formation, mitochondrial abnormality, coupled electron ATP metabolic process, demyelination/remyelination, cellular metabolism, extracellular organelle exosome vesicle and cation transmembrane transport were among the clusters downregulated in females and upregulated in males. We did not observe any pathway networks unique to female neurons; however, for the AD male neurons in the entorhinal cortex, we identified pathways in maintaining cellular metabolism and homeostasis, through the upregulation of genes involved in axon myelination, regulation of the metabolic process, cell component locomotion, cytoskeleton organization, and intracellular ferritin complex (iron storage). In male neurons, we also observed synaptic activity deficiency, indicated by the downregulation of pathways in synaptic vesicle transport, presynaptic assembly at cell junction, synaptic membrane clustering, postsynaptic membrane morphogenesis, chemical regulation at the synapse, neuroligin family protein binding, and ionotropic receptor signaling. Additionally, male AD neurons compared to controls also showed downregulation in plasma membrane regulation, cell projection, and developmental process in differentiation. While sex differences are minimal in the neurons of the prefrontal cortex, we observed overwhelmingly shared but inversely regulated enrichment pathways in the neurons of the entorhinal cortex.

Microglia, the resident immune cells of the brain, have gained growing recognition as being critically involved in AD pathogenesis due to their key role contributing to neuroinflammation, a prominent feature of AD[58]. Only a few significantly enriched disease pathways were observed in microglial cells of the prefrontal cortex and none were shared across sexes (**Fig. 3.5d**). We observed upregulation of axon sprouting in response to injury in males, as well as an enriched upregulated pathway in axonogenesis regulation in females (**Supplementary Table 3.6**). Interestingly, a cluster of the *PDE4B-DISC1*-complex, with important functions in cAMP-regulated signal transduction and synaptic plasticity,[59] was downregulated in females. The phosphodiesterase 4B (PDE4B) enzyme was previously shown to be pro-inflammatory in microglia and is currently under study as a therapeutic target for neuroinflammation and cognitive function impairment[59].

Microglia in the entorhinal cortex had mostly downregulated pathway clusters in females and upregulated pathway clusters in males (**Fig. 3.5e**). Amyloid fibril formation, chaperone-mediated autophagy, protein folding, protein stability regulation, cell junction synapse, neurogenesis structure development, and cell body assembly were among the clusters shared by both sexes but downregulated in females and upregulated in males. Protein homeostasis was altered in disease for females, as shown by downregulation of tau protein kinase activity, tau protein binding, protein folding chaperone, and histone deacetylase binding. Protein degradation and secretion were also downregulated in females with AD compared to controls, as indicated through downregulation of lytic vacuole lysosome and secretory granule vesicle exocytosis respectively. Interestingly, nitric oxide synthase 3 (NOS3), which is involved in a complex cascade of events in oxidative stress that may induce cellular injury and accelerate neurodegenerative changes[60],

and its chaperone, HSP90[61], were downregulated in AD females compared to controls. In males, myelination in axon ensheathment, synaptic signaling transmission, and energy coupled proton transport were upregulated. We also identified downregulation of two microRNA clusters, hsa-miR-190a and hsa-miR-3605, in AD males compared to healthy controls. These are potentially important findings because epigenetic modulation by microRNAs has the capacity to modify microglial behavior in physiological conditions, and dysregulation of microRNAs could mediate microglial hyper-activation and persistent neuroinflammation in neurological diseases[62]. Overall, we observed extensive sex-specific pathway enrichments in microglial populations of AD compared to controls for both brain regions, but especially pronounced in entorhinal cortex.

Furthermore, astrocytes, oligodendrocytes, and OPCs also demonstrated sex-specific pathway perturbations in both prefrontal and entorhinal cortices (**Supplementary Fig. 3.3 and 3.4**). In astrocytes, which normally function to maintain overall brain homeostasis, we observed downregulated plasma and presynaptic membrane components and upregulated postsynaptic asymmetric synapse density in the prefrontal cortex of AD compared to controls in both sexes. In female AD astrocytes, we observed downregulation in pathways related to amino acid transport and vascular transport across the blood brain barrier. Although the downregulation of these pathways was not observed in males, a related pathway cluster, presynaptic filopodia activities, was downregulated. These observed pathway networks suggest that the same biological process, regulation of synaptic activities, was disrupted in both sexes but via different mechanisms.

In oligodendrocytes, which provide support and insulation to axons in the brain, we observed downregulation in pathways related to regulation of synaptic activity in both female and male

AD compared to controls, indicated by the downregulated clusters of cleft regulation, presynaptic assembly, and transmembrane transport channel in females, and neurotransmitter secretion, transmembrane ion transporter, and postsynaptic membrane potential regulation in males. Interestingly, pathways related to cell morphological changes and energy production were upregulated in males and downregulated in females, such as pathway clusters of neuron projection organization, cell migration/locomotion, cellular component organization, ATP coupled electron transport, mitochondrial NADH dehydrogenase, suggesting oligodendrocyte responses were sex-specific when challenged by disease.

Lastly, we observed upregulation of membrane morphogenesis in female OPCs in the prefrontal cortex, as well as related pathway cluster, *TROY-NGR-LINGO1-NGFR* complex, which plays essential roles in the inhibition of axonal regeneration[63]. In the entorhinal cortex, a few pathways were downregulated in female and male OPCs, including cell junction synapse assembly, glutamatergic synapse, and plasma membrane intrinsic component. The male OPCs of the entorhinal cortex were overwhelmingly enriched with upregulation in neuronal development, axon ensheathment, neuron myelination, metabolic protein regulation, as well as ion and vesicle transport, with the exception that synaptic membrane adhesion molecules were downregulated. Although inconclusive due to the unbalanced numbers of significantly enriched pathways obtained in OPCs from both sexes, our observations suggest that AD female OPCs in the prefrontal cortex diverge more from controls compared to male OPCs, whereas in the entorhinal cortex, AD male OPCs were more perturbed by disease status compared to females.

3.4 DISCUSSION

Men and women show differing vulnerabilities to AD, with increased longevity and prevalence in women, and decreased tau and possibly cognitive decline in men[17–19,21,22]. To understand how AD presents in each sex on a cell type-specific level, we performed a sex-stratified differential gene expression (DGE) and pathway network analysis on the five main brain cell types using the first two publicly available human single nucleus RNA-Seq datasets. The two datasets target two separate brain regions, the entorhinal and prefrontal cortices, and we analyzed each in a sex-stratified manner, then compared findings across sexes and brain regions to highlight both general and cell type-, region-, and sex- specific transcriptional phenotypes of AD (**Fig. 3.1**).

Our gene level analysis in the prefrontal cortex demonstrated sex-distinct disease-related transcriptomic changes when comparing their respective cases to controls (**Fig. 3.2**). We observed more DEGs shared among cell types in females versus males (e.g. *LINGO1*[38,39], *SLC1A3*[40], *SPP1*[41]), and a larger range of fold change in our female DGE analysis. Additionally, through clustering prefrontal cortex samples based on AD compared to control pseudo-bulk gene expression, we observed samples to cluster first by sex in all cell types except excitatory neurons. In the entorhinal cortex, compared to the prefrontal cortex, we observed more DEGs and many global changes across cell types of both sexes (**Fig. 3.3a**). Through clustering entorhinal cortex samples by AD compared to control pseudo-bulk gene expression, we observed samples to cluster by sex for all cell types and observed opposing expression patterns across sex (**Fig. 3.3b-d**), implying sex-distinct mechanisms of neurodegeneration in the entorhinal cortex. Moreover, our comparative analysis across brain regions showed more DEG overlaps across sex

in the entorhinal cortex, and disease-related changes of gene expression to be influenced by brain region and sex (**Fig. 3.3, Fig. 3.4**).

From the gene-set enrichment and pathway clustering network analysis, we identified sex-specific pathway network changes, which are potentially involved in AD pathogenesis through mechanisms unique to each sex (**Fig. 3.5, Supplementary Figure 3.3, and Supplementary Figure 3.4**). Our results demonstrated that diseased neurons in the prefrontal cortex shared more enriched pathways compared to glial cells in both sexes, indicated by the proportion and directionality of the shared pathways. This may suggest that neuronal pathophysiology is similar in female versus male, and glial pathophysiological changes are more distinctive in contributing to sex-specific disease progression in AD. Despite neurons being more similar than glial cells, interesting sex-specific biological perturbations were revealed in neurons of females and males separately. Diseased female neurons showed increased activation in cell membrane morphogenesis but reduction in the production of tight junction complexes. A few transcriptional factors were uniquely upregulated in females, such as HOXA5, HOXB2, and TLX. Future studies investigating the role of overactivation of these genes in AD, especially in females, could lead to better mechanistic understanding of AD pathogenesis and potential therapies targeting these transcriptional factors in females. In diseased male neurons, nitric oxide synthase (NOS) activity was downregulated, as well as its regulating factors, the HSP90 complex and co-factor BH4. BH4 has been extensively studied in its role of regulating nitric oxide production from nitric oxide synthases and superoxide anion radical (O_2^{*-}) release in the endothelium[64]. Our pathway enrichment analysis suggests that perhaps excessive O_2^{*-} in diseased male neurons due to dysregulated NOS activities and BH4 levels could lead to neuronal stress and death.

Therefore, resolving the chronic BH4 deficiency and change in redox state of neurons pharmacologically could be a beneficial therapy for AD male patients.

The glial cells in the prefrontal cortex shared just a few enriched pathways, out of hundreds detected collectively, between AD males and females: nine in astrocytes (**Supplementary Figure 3.3a**), two in oligodendrocytes (**Supplementary Figure 3.4a**), and none in microglia and OPCs (**Fig. 3.5d and Supplementary Figure 3.4c**) (these numbers do not include shared pathways regulated in opposite directions in female vs male). Besides downregulation of membrane morphogenesis, both female and male diseased astrocytes demonstrated decreased synaptic regulation, but different pathways for different components were involved. In females, we observed a decrease in glutamate transmembrane transport, vascular transport, and organic acid symporter activities. In males, we observed a decrease in presynaptic intrinsic component filopodia activities. These pathways are interconnected, indicating that they belong to related biological processes, which suggests that similar resulting synaptic deficiencies were observed in both sexes but resulted from different pathway mechanisms. These present compelling evidence for focusing on glial cell pathophysiological changes in studying sex differences in AD pathogenesis.

In the entorhinal cortex, while like in the prefrontal cortex, we identified sex-specific perturbed pathway networks in all cell types, where the pathways shared across sexes were overwhelmingly of opposite direction, with most pathways downregulated in female and upregulated in males (**Fig. 3.5, Supplementary Figure 3.3, and Supplementary Figure 3.4**). Out of the five cell types investigated, two were dominated by enriched pathways detected in

males (neurons and OPCs), one was dominated by enriched pathways detected in females (oligodendrocytes), and two were more evenly distributed (microglia and astrocytes). The diseased female microglia demonstrated deficiency in tau protein processing uniquely, by downregulation of tau kinase activity and tau protein binding. Additionally, disruption of cellular protein homeostasis was also observed in female microglia, indicated by downregulation of protein folding chaperone, histone deacetylase binding, lysosomal activity, and exocytosis vesicle secretion. The female microglia were perceived as deficient in dealing with the degradation of the debris and cellular waste that they phagocytosed while the male microglia were active at combating the disease environment by upregulating axonal myelination, synaptic transmission signaling, cellular component assembly and energy production through energy coupled proton transport. As immune cells are critical for repair after injury, this may indicate that female AD risk relates to decreased ability to properly recover after deleterious events over time.

While we observed evidence of sex-dimorphic disease changes in glial cells in AD, it is important to note some limitations in the study. First, the data sets were limited in sample size. The entorhinal cohort consisted of six cases (two female, four male) and five controls (two female, three male) (**Fig. 3.1, Table 3.2),** while the prefrontal cohort consisted of 20 cases (10 female, 10 male) and 22 controls (10 female, 12 male) (**Fig. 3.1, Table 3.1**). Second, there were batch effects in the entorhinal cortex data introduced by the study design. This was overcome by performing data integration and including *APOE* genotype as a covariate in our DGE analysis to account for batch and avoid collinearity in our model. Next, although both datasets were age-matched, they were not *APOE* genotype matched. APOE4 is the largest risk factor in AD, and as

a result, we would expect some transcriptional differences based on the *APOE* genotype of a sample[66]. In the prefrontal cortex cohort, female samples had cases but not controls with the ε4 allele of *APOE*, and male samples had cases and only one control sample with ε4 allele of *APOE* (**Table 3.1**)*.* In the entorhinal cortex cohort, female samples included one of two cases and no controls with an ε4 allele of *APOE*, and all male cases had at least one ε4 allele of *APOE*, and one of three control samples had an ε4 allele of *APOE* (**Table 3.2**). While we accounted for *APOE* genotype as a covariate in the DGE analysis, the interactions of sex and *APOE* genotype may still explain trends that we observe.

Additionally, interpretation at the DEG level (**Fig. 3.2a, Fig. 3.3a, and Fig. 3.4a**) was limiting without disease-relevant mechanistic insights. To summarize novel and previously studied DEGs in AD, we extended our analysis to include pathway and network enrichment. When comparing our results across brain regions, noting the limitations of each dataset, we were cautious to not further explore unique molecular profiles in each region, which could give insight into the spread of AD pathology as it relates to sex. Moreover, literal biological sex could be a misleading classifier for trans* individuals. A properly powered study of differences between male versus female versus recipients of testosterone- versus estrogen-focused hormone replacement therapy might help narrow down a genetic versus hormonal basis of DEGs deemed sexually dimorphic. Overall, we hope that larger and different types of omics datasets from more brain regions of individuals with diverse age groups, racial and ethnic backgrounds, and *APOE* genotypes become available to allow for future explorations of sex-specific multiomic changes in AD that will address these points.

In general, our findings suggest that AD signatures in neurons in the prefrontal cortex were more similar in females and males compared to glial cells, as indicated by the proportions of sex-shared genes and pathways with directionally similar regulation in each cell type (**Fig. 3.5, Supplementary Figure 3.3, and Supplementary Figure 3.4**). In the entorhinal cortex, while we identified sex-specific perturbed pathways in each cell type, the sex-shared pathways were overwhelmingly opposite in the direction of regulation, with most pathways downregulated in female and upregulated in males or conversely regulated for a few other pathways, suggesting differential mechanisms of neurodegeneration between sexes. Sex-stratified findings in the entorhinal cortex could relate to recent observations that women show more tau deposition early on in the AD trajectory, specifically in this area[67]. Perhaps future studies could also explore the specific association between the gene changes in the entorhinal region with tau burden. Collectively, these observed sex-specific transcriptomic changes provide a valuable resource to study sex-specific cell type-specific pathophysiology of AD. Although expression differences in all cell types may be relevant to disease mechanisms in AD, we focused on discussing the cell types with the most compelling findings in our study: neurons, astrocytes, and microglia. We hope this work serves as a resource for follow-up studies that will examine more deeply all the cell types and their specific roles leading to sex-specific AD pathophysiology.

3.5 METHODS

3.5.1 Materials Availability

This study did not generate new unique reagents.

3.5.2 Data and Code Availability

We accessed single nuclei RNA-Seq counts data from the prefrontal cortex via the Accelerating

Medicines Partnership Alzheimer's Disease Project (AMP-AD) Knowledge Portal under the

Religious Orders Study and Memory and Aging Project (ROSMAP)

(https://www.synapse.org/#!Synapse:syn18485175;

https://www.synapse.org/#!Synapse:syn3157322), and from the entorhinal cortex via a data

repository provided by Grubman et al. (http://adsn.ddnetbio.com/). The entorhinal cortex dataset

and supporting materials may also be accessed via the Gene Expression Omnibus (GEO) under

the accession number GSE138852. Access to the prefrontal cortex dataset requires a formal

request to ROSMAP. All code necessary for recreating the reported analyses and figures within

R are available at: https://github.com/stebel5/AD_SexDiff_snRNAseq.

3.5.3 Identification of Study Cohorts

The prefrontal cortex cohort comprised age and sex matched samples from 24 males and 24

females with varying degrees of AD pathology. We reclassified samples based on tau and

amyloid β (Aβ) plaque burden, using Braak clinical staging and Consortium to Establish a

Registry for Alzheimer's Disease (CERAD) scores, respectively. We defined cases as individuals

with severe tau deposition (Braak $\geq$ IV), and high Aβ load (CERAD $\leq$ 2), and non-AD controls

as individuals with low tau (Braak $\leq$ III) and low Aβ load (CERAD $\geq$ 3). For our sex-stratified

analysis, we focused on 20 cases (10 female, 10 male) and 22 controls (10 female, 12 male) (**Fig.**

3.1, Table 3.1).

The entorhinal cortex cohort consisted of age matched 6 (2 female, 4 male) AD patients and 6 (2 female, 4 male) control subjects, as indicated by Grubman et al. All cases had a history of AD, while controls had no history of AD or cognitive impairment, as reported by treating general practitioners. For pathological scores used in categorizing samples, Braak staging scores were provided only for cases, and amyloid pathology information was provided for all samples using the categories: "Numerous diffuse and neuritic Aβ plaque," "Occasional diffuse plaque in cortex," and "None." To use the same scoring system for identifying cases and controls in both datasets, we used criteria from the Rush Alzheimer's Disease Center clinical codebook provided with the prefrontal cortex dataset to convert these measures of neuritic plaques into CERAD scores of 1 (Definite), 3 (Possible), and 4 (No AD), respectively. We excluded one control male sample with the APOE2/4 genotype. For our sex-stratified analysis, we focused on 6 cases (2 female, 4 male) and 5 controls (2 female, 3 male) (**Fig. 3.1, Table 3.2**).

3.5.4 Single Cell Data Processing, Cell Type Identification and Batch Correction

Data processing and analysis was performed separately for each dataset with R[68] version 4.0.0 (2020-04-24) via RStudio[69], using Seurat[65] (v3.1.5). Visualizations were created with BioRender (https://biorender.com/) (**Fig. 3.1**), dittoSeq (v1.0.2) (https://github.com/dtm2451/dittoSeq/), a package for analysis and visualization of bulk and single-cell transcriptomic data in a color blind friendly manner, ggplot2[70], and UpsetR[71]**.**

Prefrontal Cortex
Seurat's Read10X function was used to generate a count data matrix using the filtered count matrix of 17,296 genes and 70,634 cells, gene names, and barcodes files provided by 10X. A Seurat object was created with the count data matrix and metadata and filtered to keep genes

present in at least 3 cells, and cells meeting cohort selection criteria of at least 200 genes. Log normalization was performed using Seurat's NormalizeData function with a scale factor of 10,000, and highly variable features were identified using Seurat's FindVariableFeatures, returning 3,188 features, as specified in the original paper. The data matrix was then scaled using Seurat's ScaleData function with *nCount_RNA* regressed out, and dimensionality reduction through Uniform Manifold Approximation and Projection (UMAP) was performed with the appropriate dimensions selected based on the corresponding principal component analysis (PCA) elbow plot. UMAP plots confirmed that there were no confounding variables (**Supplementary Figure 3.1**).

To identify cell types, following similar steps as Grubman and colleagues[36], we applied Seurat's *AddModuleScore* function to lists of 200 brain cell type markers from the BRETIGEA[72] package to identify each cell type. Cell types assessed included astrocytes, neurons, microglia, oligodendrocytes, oligodendrocyte progenitor cells, pericytes, and endothelial cells. Cells with the highest score across brain cell type markers were labeled the corresponding cell type, and if the highest and second highest score were within 20%, cells were deemed hybrids and excluded from further analysis. We further confirmed successful cell type identification by assessing homogeneity and separation of clusters in UMAP plots, and by examining expression of top marker genes across cell types. While cell type identification with BRETIGEA package's cell type markers was comparable to the original paper's identification, we found the original paper's cell types more comprehensive as it distinguished excitatory from inhibitory neurons. Thus, we used the original paper's cell type labels for the further analysis (**Supplementary Table 3.1**).

Due to low cell counts, we did not analyze pericytes and endothelial cells. The final Seurat object contained 17,723 genes and 62,741 cells.

Entorhinal Cortex

We acquired a filtered raw expression matrix of 10,850 genes and 13,214 cells, which was originally composed of 33,694 genes and 14,876 cells and filtered as described by Grubman and colleagues. A Seurat object was created and consisted of genes in at least 3 cells, and cells with at least 200 genes. Normalization was performed using Seurat's SCTransform[73] method, and Seurat's integration workflow was performed to correct the confounded batches introduced by the original study's experimental design.

Dimensionality reduction was performed using values from the integrated assay to assess successful batch correction (**Supplementary Figure 3.1**). Using the method for cell type identification described for the former cohort, we identified astrocytes, endothelial cells, neurons, microglia, oligodendrocytes, and oligodendrocyte progenitor cells. We further confirmed successful cell type identification by assessing homogeneity and separation of clusters in UMAP plots. Due to limitations in the number of cells, we excluded endothelial cells from further analyses. The final Seurat object contained 10,846 genes and 11,284 cells (**Supplementary Table 3.2**).

3.5.5 Cell Type-Specific Sex-stratified Differential Expression Analysis

To generate molecular signatures relative to sex in each cell type, we used the Limma[74,75] package's Voom[76] pipeline for RNA-seq. For the prefrontal and entorhinal cortices, we

performed a sex-stratified analysis including *APOE* genotype as a covariate. For the entorhinal cortex cohort, while we integrated batches in our pre-processing, we were not able to include batch as a covariate, as its collinearity did not allow for an appropriate model fit.

After the design formulas were established, the DGEList object was created from a matrix of counts extracted from the corresponding Seurat objects. To improve the accuracy of mean-variance trend modeling and lower the severity of multiple testing correction, lowly expressed genes were filtered out using edgeR's FilterByExpr function with default parameters. Normalization was performed with Trimmed Mean of M-values with singleton pairing (TMMwsp), followed by voom, model fitting with a contrast matrix of each defined case-control comparison, and Empirical Bayes fitting of standard errors. We determined differentially expressed genes (DEGs) as those with a Benjamini-Hochberg corrected p-value less than 0.05, and an absolute LFC greater than 0.25. We then examined AD compared to control gene expression changes in all cell types of each sex using pairwise gene expression plots, violin plots of gene expression, hierarchical clustering of samples using AD compared to control pseudobulk cell type gene expression, and Upset plots, which prioritized labeling DEGs with more overlaps across the groups compared.

3.5.6 Pathway Analysis

We performed an overrepresentation analysis of DEGs from the cell type-specific sex-stratified analysis of cells from the prefrontal and entorhinal cortex using g:Profiler[50], a web tool that performs functional enrichment analysis from a given gene list. We queried DEGs split by upregulated and down-regulated expression and selected enriched pathways with a Benjamini-

Hochberg adjusted p-value cutoff of 0.05. In addition to Gene Ontology cellular components, biological processes, and molecular functions, our enrichment analysis also provided pathways from the Human Protein Atlas, Human Phenotype Ontology, KEGG, Reactome, and Wiki pathways.

3.5.7 Network Visualization of Enrichment Results

We followed a previously established protocol[51] for network enrichment analysis on pathway results derived from our cell type-specific DEGs. Briefly, pathway results were imported into the Cytoscape visualization application, EnrichmentMap. Then, redundant, and related pathways were collapsed into single biological themes using the AutoAnnotate Cytoscape application.

3.6 REFERENCES

1.   Masters, C. L. *et al.* Alzheimer's disease. *Nature Reviews Disease Primers* **1**, 1–18 (2015).

2.   Murray, M. E. *et al.* Neuropathologically defined subtypes of Alzheimer's disease with distinct clinical characteristics: a retrospective study. *Lancet Neurol* **10**, 785–796 (2011).

3.   Scheltens, P. *et al.* Alzheimer's disease. *Lancet* **388**, 505–517 (2016).

4.   Prince, M. J., Comas-Herrera, A., Knapp, M., Guerchet, M. M. & Karagiannidou, M. World Alzheimer Report 2016 - Improving healthcare for people living with dementia: Coverage, quality and costs now and in the future. (2016).

5.   Bureau, U. C. An Aging World: 2015. *The United States Census Bureau* https://www.census.gov/library/publications/2016/demo/P95-16-1.html.

6.   Plassman, B. L. *et al.* Prevalence of dementia in the United States: the aging, demographics, and memory study. *Neuroepidemiology* **29**, 125–132 (2007).

7.   Hebert, L. E., Weuve, J., Scherr, P. A. & Evans, D. A. Alzheimer disease in the United States (2010-2050) estimated using the 2010 census. *Neurology* **80**, 1778–1783 (2013).

8.   Goedert, M. & Spillantini, M. G. A century of Alzheimer's disease. *Science* **314**, 777–781 (2006).

9.   Karch, C. M. & Goate, A. M. Alzheimer's disease risk genes and mechanisms of disease pathogenesis. *Biological Psychiatry* **77**, 43–51 (2015).

10.   Bloom, G. S. Amyloid-β and tau: the trigger and bullet in Alzheimer disease pathogenesis. *JAMA Neurol* **71**, 505–508 (2014).

11.   Busche, M. A. & Hyman, B. T. Synergy between amyloid-β and tau in Alzheimer's disease. *Nat Neurosci* **23**, 1183–1193 (2020).

12. Heneka, M. T. *et al.* Neuroinflammation in Alzheimer's disease. *Lancet Neurol* **14**, 388–405 (2015).

13. Hebert, L. E. *et al.* Change in risk of Alzheimer disease over time. *Neurology* **75**, 786–791 (2010).

14. Xu, Y. *et al.* Neurotransmitter receptors and cognitive dysfunction in Alzheimer's disease and Parkinson's disease. *Prog Neurobiol* **97**, 1–13 (2012).

15. de la Monte, S. M. & Tong, M. Brain metabolic dysfunction at the core of Alzheimer's disease. *Biochem Pharmacol* **88**, 548–559 (2014).

16. Donev, R., Kolev, M., Millet, B. & Thome, J. Neuronal death in Alzheimer's disease and therapeutic opportunities. *J Cell Mol Med* **13**, 4329–4348 (2009).

17. Dubal, D. B. Sex difference in Alzheimer's disease: An updated, balanced and emerging perspective on differing vulnerabilities. *Handb Clin Neurol* **175**, 261–273 (2020).

18. Mielke, M. M., Vemuri, P. & Rocca, W. A. Clinical epidemiology of Alzheimer's disease: assessing sex and gender differences. *Clin Epidemiol* **6**, 37–48 (2014).

19. Medeiros, A. de M. & Silva, R. H. Sex Differences in Alzheimer's Disease: Where Do We Stand? *J Alzheimers Dis* **67**, 35–60 (2019).

20. Davis, E. J. *et al.* A second X chromosome contributes to resilience in a mouse model of Alzheimer's disease. *Sci Transl Med* **12**, (2020).

21. Barnes, L. L. *et al.* Sex Differences in the Clinical Manifestations of Alzheimer Disease Pathology. *Archives of General Psychiatry* **62**, 685–691 (2005).

22. Oveisgharan, S. *et al.* Sex differences in Alzheimer's disease and common neuropathologies of aging. *Acta Neuropathol* **136**, 887–900 (2018).

23.   Hohman, T. J. *et al.* Sex-Specific Association of Apolipoprotein E With Cerebrospinal Fluid Levels of Tau. *JAMA Neurol* **75**, 989–998 (2018).

24.   Guerreiro, R. J., Gustafson, D. R. & Hardy, J. The genetic architecture of Alzheimer's disease: beyond APP, PSENs and APOE. *Neurobiol Aging* **33**, 437–456 (2012).

25.   Altmann, A., Tian, L., Henderson, V. W., Greicius, M. D. & Alzheimer's Disease Neuroimaging Initiative Investigators. Sex modifies the APOE-related risk of developing Alzheimer disease. *Ann Neurol* **75**, 563–573 (2014).

26.   Barron, A. M. & Pike, C. J. Sex hormones, aging, and Alzheimer's disease. *Front Biosci (Elite Ed)* **4**, 976–997 (2012).

27.   Ibanez, C. *et al.* Systemic progesterone administration results in a partial reversal of the age-associated decline in CNS remyelination following toxin-induced demyelination in male rats. *Neuropathol Appl Neurobiol* **30**, 80–89 (2004).

28.   Ghoumari, A. M., Baulieu, E. E. & Schumacher, M. Progesterone increases oligodendroglial cell proliferation in rat cerebellar slice cultures. *Neuroscience* **135**, 47–58 (2005).

29.   Lasselin, J., Lekander, M., Axelsson, J. & Karshikoff, B. Sex differences in how inflammation affects behavior: What we can learn from experimental inflammatory models in humans. *Frontiers in Neuroendocrinology* **50**, 91–106 (2018).

30.   Quintero, O. L., Amador-Patarroyo, M. J., Montoya-Ortiz, G., Rojas-Villarraga, A. & Anaya, J.-M. Autoimmune disease and gender: Plausible mechanisms for the female predominance of autoimmunity. *Journal of Autoimmunity* **38**, J109–J119 (2012).

31.   Di Florio, D. N., Sin, J., Coronado, M. J., Atwal, P. S. & Fairweather, D. Sex differences in inflammation, redox biology, mitochondria and autoimmunity. *Redox Biol* **31**, 101482 (2020).

32.   Nirzhor, S. S. R., Khan, R. I. & Neelotpol, S. The Biology of Glial Cells and Their

Complex Roles in Alzheimer's Disease: New Opportunities in Therapy. *Biomolecules* **8**, (2018).

33.   Luchena, C., Zuazo-Ibarra, J., Alberdi, E., Matute, C. & Capetillo-Zarate, E. Contribution

of Neurons and Glial Cells to Complement-Mediated Synapse Removal during Development,

Aging and in Alzheimer's Disease. *Mediators of Inflammation* **2018**, e2530414 (2018).

34.   Paranjpe, M. D. *et al.* Sex-Specific Cross Tissue Meta-Analysis Identifies Immune

Dysregulation in Women with Alzheimer's Disease. *bioRxiv* 2020.04.24.060558 (2020)

doi:10.1101/2020.04.24.060558.

35.   Mathys, H. *et al.* Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**,

332–337 (2019).

36.   Grubman, A. *et al.* A single-cell atlas of entorhinal cortex from individuals with

Alzheimer's disease reveals cell-type-specific gene expression regulation. *Nat Neurosci* **22**,

2087–2097 (2019).

37.   Mirra, S. S. *et al.* The Consortium to Establish a Registry for Alzheimer's Disease

(CERAD). Part II. Standardization of the neuropathologic assessment of Alzheimer's disease.

Neurology **41**, 479–486 (1991).

38.   Fernandez-Enright, F. & Andrews, J. L. Lingo-1: a novel target in therapy for

Alzheimer's disease? *Neural Regen Res* **11**, 88–89 (2016).

39.   Mi, S. *et al.* LINGO-1 negatively regulates myelination by oligodendrocytes. *Nature

Neuroscience* **8**, 745–751 (2005).

40.   Iwama, K. *et al.* A novel mutation in SLC1A3 causes episodic ataxia. *J Hum Genet* **63**,

207–211 (2018).

41.   Mahmud, F. J. *et al.* Osteopontin/secreted phosphoprotein-1 behaves as a molecular brake regulating the neuroinflammatory response to chronic viral infection. *J Neuroinflammation* **17**, 273 (2020).

42.   Mozhui, K. *et al.* Genetic regulation of Nrnx1 expression: an integrative cross-species analysis of schizophrenia candidate genes. *Transl Psychiatry* **1**, e25 (2011).

43.   Cario, H. *et al.* Dihydrofolate Reductase Deficiency Due to a Homozygous DHFR Mutation Causes Megaloblastic Anemia and Cerebral Folate Deficiency Leading to Severe Neurologic Disease. *The American Journal of Human Genetics* **88**, 226–231 (2011).

44.   Sahin, P., McCaig, C., Jeevahan, J., Murray, J. T. & Hainsworth, A. H. The cell survival kinase SGK1 and its targets FOXO3a and NDRG1 in aged human brain. *Neuropathology and Applied Neurobiology* **39**, 623–633 (2013).

45.   Wang, B.-J. *et al.* ErbB2 regulates autophagic flux to modulate the proteostasis of APP-CTFs in Alzheimer's disease. *PNAS* **114**, E3129–E3138 (2017).

46.   Kok, E. H. *et al.* CLU, CR1 and PICALM genes associate with Alzheimer's-related senile plaques. *Alzheimer's Research & Therapy* **3**, 12 (2011).

47.   Muraoka, S. *et al.* Proteomic Profiling of Extracellular Vesicles Derived from Cerebrospinal Fluid of Alzheimer's Disease Patients: A Pilot Study. *Cells* **9**, (2020).

48.   Raghavan, N. S. *et al.* Association Between Common Variants in *RBFOX1* , an RNA-Binding Protein, and Brain Amyloidosis in Early and Preclinical Alzheimer Disease. *JAMA Neurol* **77**, 1288 (2020).

49.   Hua, Y., Zhao, H., Lu, X., Kong, Y. & Jin, H. Meta-Analysis of the Cystatin C(CST3) Gene G73A Polymorphism and Susceptibility to Alzheimer's Disease. *International Journal of Neuroscience* **122**, 431–438 (2012).

50.   Raudvere, U. *et al.* g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res* **47**, W191–W198 (2019).

51.   Reimand, J. *et al.* Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. *Nature Protocols* **14**, 482–517 (2019).

52.   Chen, H., Chung, S. & Sukumar, S. HOXA5-Induced Apoptosis in Breast Cancer Cells Is Mediated by Caspases 2 and 8. *Mol. Cell. Biol.* **24**, 924 (2004).

53.   Smith, R. G. *et al.* Elevated DNA methylation across a 48-kb region spanning the HOXA gene cluster is associated with Alzheimer's disease neuropathology. *Alzheimer's & Dementia* **14**, 1580–1588 (2018).

54.   Kapatos, G. The neurobiology of tetrahydrobiopterin biosynthesis: A model for regulation of GTP cyclohydrolase I gene transcription within nigrostriatal dopamine neurons. *IUBMB Life* **65**, 323–333 (2013).

55.   Frank, J. W., Seo, H., Burghardt, R. C., Bayless, K. J. & Johnson, G. A. ITGAV (alpha v integrins) bind SPP1 (osteopontin) to support trophoblast cell adhesion. *Reproduction* **153**, 695–706 (2017).

56.   Sun, G., Cui, Q. & Shi, Y. Chapter Nine - Nuclear Receptor TLX in Development and Diseases. in *Current Topics in Developmental Biology* (eds. Forrest, D. & Tsai, S.) vol. 125 257–273 (Academic Press, 2017).

57.   Davenne, M. *et al.* Hoxa2 and Hoxb2 control dorsoventral patterns of neuronal development in the rostral hindbrain. *Neuron* **22**, 677–691 (1999).

58.   Hansen, D. V., Hanson, J. E. & Sheng, M. Microglia in Alzheimer's disease. *J Cell Biol* **217**, 459–472 (2018).

59.   Tibbo, A. J. & Baillie, G. S. Phosphodiesterase 4B: Master Regulator of Brain Signaling. *Cells* **9**, 1254 (2020).

60.   Yun, H.-Y., Dawson, V. & Dawson, T. Nitric oxide in health and disease of the nervous system. *Nitric oxide in the nervous system* 12.

61.   Miao Robert Q. *et al.* Dominant-Negative Hsp90 Reduces VEGF-Stimulated Nitric Oxide Release and Migration in Endothelial Cells. *Arteriosclerosis, Thrombosis, and Vascular Biology* **28**, 105–111 (2008).

62.   Guo, Y. *et al.* MicroRNAs in Microglia: How do MicroRNAs Affect Activation, Inflammation, Polarization of Microglia and Mediate the Interaction Between Microglia and Glioma? *Front. Mol. Neurosci.* **12**, (2019).

63.   Theotokis, P. *et al.* Nogo receptor complex expression dynamics in the inflammatory foci of central nervous system experimental autoimmune demyelination. *Journal of Neuroinflammation* **13**, 265 (2016).

64.   Whitsett, J. *et al.* Endothelial cell superoxide anion radical generation is not dependent on endothelial nitric oxide synthase-serine 1179 phosphorylation and endothelial nitric oxide synthase dimer/monomer distribution. *Free Radic Biol Med* **40**, 2056–2068 (2006).

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

66.   Belonwu, S. *et al. Single-cell transcriptomic analysis elucidates APOE genotype specific changes across cell types in two brain regions in Alzheimer's disease.* https://www.researchsquare.com/article/rs-291648/v1 (2021) doi:10.21203/rs.3.rs-291648/v1.

67.   Buckley, R. F. *et al.* Sex Differences in the Association of Global Amyloid and Regional

Tau Deposition Measured by Positron Emission Tomography in Clinically Normal Older Adults.

JAMA Neurol **76**, 542–551 (2019).

68.   R Core Team. R: A Language and Environment for Statistical Computing. https://www.r-

project.org/ (2020).

69.   R Studio Team. RStudio: Integrated Development Environment for R.

https://rstudio.com/ (2020).

70.   Ginestet, C. ggplot2: Elegant Graphics for Data Analysis. *Journal of the Royal Statistical

Society: Series A (Statistics in Society)* (2011) doi:10.1111/j.1467-985x.2010.00676_9.x.

71.   Conway, J. R., Lex, A. & Gehlenborg, N. UpSetR: an R package for the visualization of

intersecting sets and their properties. *Bioinformatics* **33**, 2938–2940 (2017).

72.   McKenzie, A. T. *et al.* Brain Cell Type Specific Gene Expression and Co-expression

Network Architectures. *Scientific Reports* **8**, 1–19 (2018).

73.   Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-

seq data using regularized negative binomial regression. *Genome Biology* **20**, 296 (2019).

74.   Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing

and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).

75.   Phipson, B., Lee, S., Majewski, I. J., Alexander, W. S. & Smyth, G. K. Robust

hyperparameter estimation protects against hypervariable genes and improves power to detect

differential expression. *Ann Appl Stat* **10**, 946–963 (2016).

76.   Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: precision weights unlock linear

model analysis tools for RNA-seq read counts. *Genome Biol* **15**, R29 (2014).

3.7 TABLES

Table 3.1: Prefrontal cortex cohort

ID	Sex	APOE	Age	Diagnosis	Batch
ADF1	Female	2/3	90	AD	7
ADF2	Female	3/3	90	AD	3
ADF3	Female	3/4	90	AD	11
ADF4	Female	3/3	90	AD	5
ADF5	Female	3/3	90	AD	8
ADF6	Female	3/3	87	AD	10
ADF7	Female	3/4	76	AD	3
ADF8	Female	3/4	74	AD	2
ADF9	Female	4/4	82	AD	6
ADF10	Female	4/4	83	AD	7
ADM1	Male	2/3	86	AD	12
ADM2	Male	3/3	89	AD	5
ADM3	Male	3/4	89	AD	1
ADM4	Male	3/3	87	AD	10
ADM5	Male	3/4	85	AD	4
ADM6	Male	3/4	86	AD	9
ADM7	Male	3/3	83	AD	4
ADM8	Male	3/3	86	AD	2
ADM9	Male	3/3	80	AD	8
ADM10	Male	4/4	83	AD	11
CF1	Female	2/3	90	Control	2
CF2	Female	2/3	90	Control	3
CF3	Female	2/3	90	Control	5
CF4	Female	2/3	90	Control	7
CF5	Female	3/3	90	Control	11
CF6	Female	3/3	87	Control	9
CF7	Female	3/3	87	Control	10
CF8	Female	3/3	83	Control	6
CF9	Female	3/3	81	Control	7
CF10	Female	3/3	79	Control	3
CM1	Male	2/3	86	Control	10
CM2	Male	2/3	83	Control	11
CM3	Male	2/3	86	Control	12
CM4	Male	3/3	90	Control	5
CM5	Male	3/4	88	Control	1
CM6	Male	3/3	87	Control	12
CM7	Male	3/3	84	Control	4
CM8	Male	3/3	80	Control	8
CM9	Male	3/3	80	Control	1
CM10	Male	3/3	80	Control	9
CM11	Male	3/3	79	Control	2
CM12	Male	3/3	76	Control	4

Table 3.2: Entorhinal cortex cohort

ID	Sex	APOE	Age	Diagnosis	Batch
ADF1	Female	4/4	67.8	AD	AD3_AD4
ADF2	Female	3/3	83.0	AD	AD3_AD4
ADM1	Male	3/4	91	AD	AD1_AD2
ADM2	Male	3/4	83.8	AD	AD1_AD2
ADM3	Male	4/4	73	AD	AD5_AD6
ADM4	Male	3/4	74.6	AD	AD5_AD6
CF1	Female	3/3	67.3	Control	Ct1_Ct2
CF2	Female	3/3	82.7	Control	Ct1_Ct2
CM1	Male	3/3	72.6	Control	Ct3_Ct4
CM2	Male	3/4	75.6	Control	Ct3_Ct4
CM3	Male	3/3	77.5	Control	Ct5_Ct6

Table 3.3: Number of differentially expressed genes in both sexes per cell type in the prefrontal cortex

	Male		Female	
	Up	Down	Up	Down
Ast	5	2	23	7
Ex	33	3	30	5
In	0	0	32	1
Mic	1	0	6	2
Oli	1	2	3	2
Opc	1	3	7	0

Table 3.4: Number of differentially expressed genes in both sexes per cell type in the entorhinal cortex

	Male		Female	
	Up	Down	Up	Down
Ast	175	36	19	109
Mic	101	23	9	15
Neu	155	30	13	46
Oli	150	40	22	163
Opc	141	31	13	12
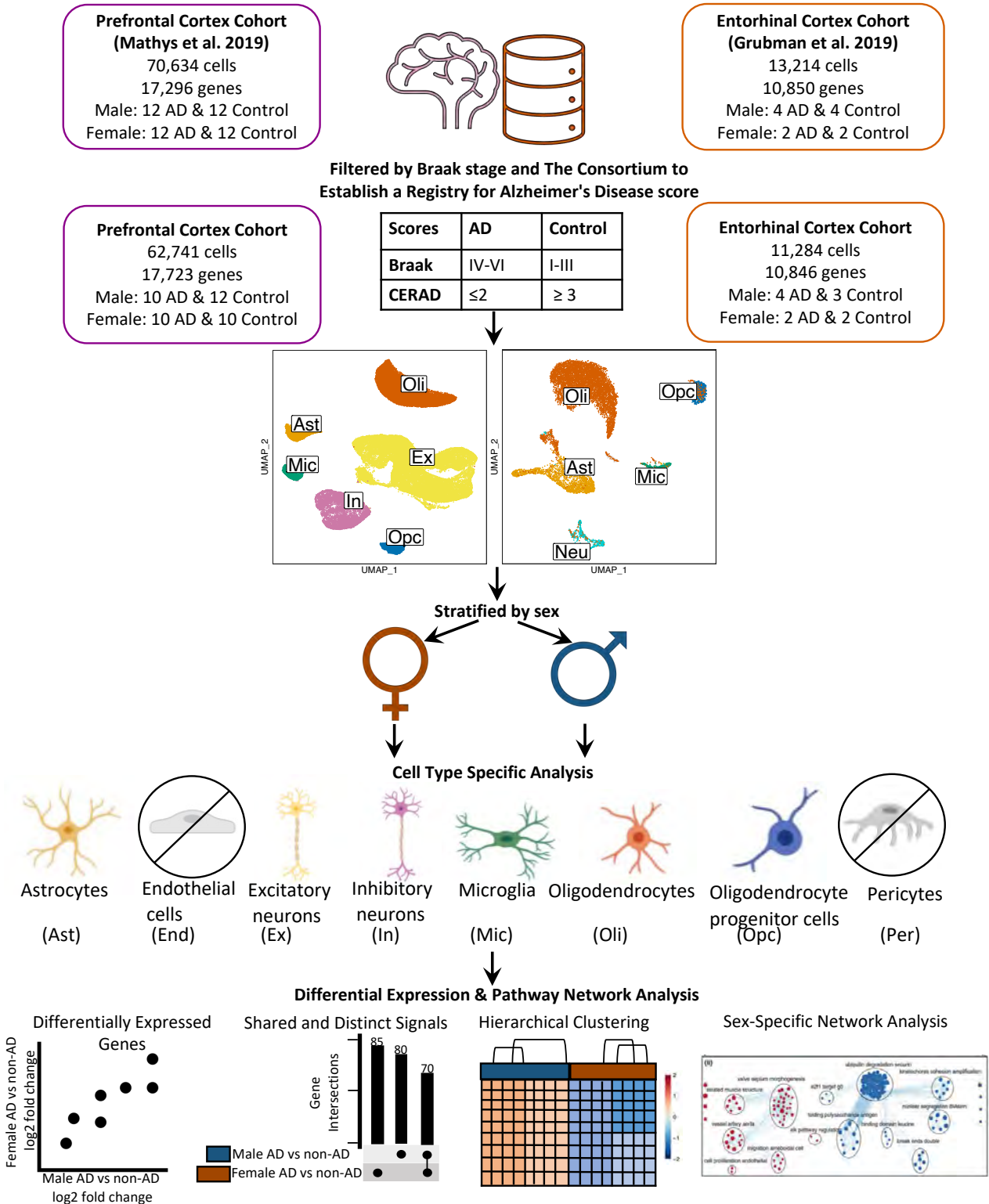
Figure 3.1: Overview of cohort sample definition and workflow for sex-stratified cell type-specific differential gene expression and functional enrichment. AD and non-AD cells were determined based on tau (Braak) and amyloid β plaque (CERAD) burdens. Cell types were identified, and AD versus non-AD differential expression and pathway network enrichment analyses were performed separately for each sex in each cell type.
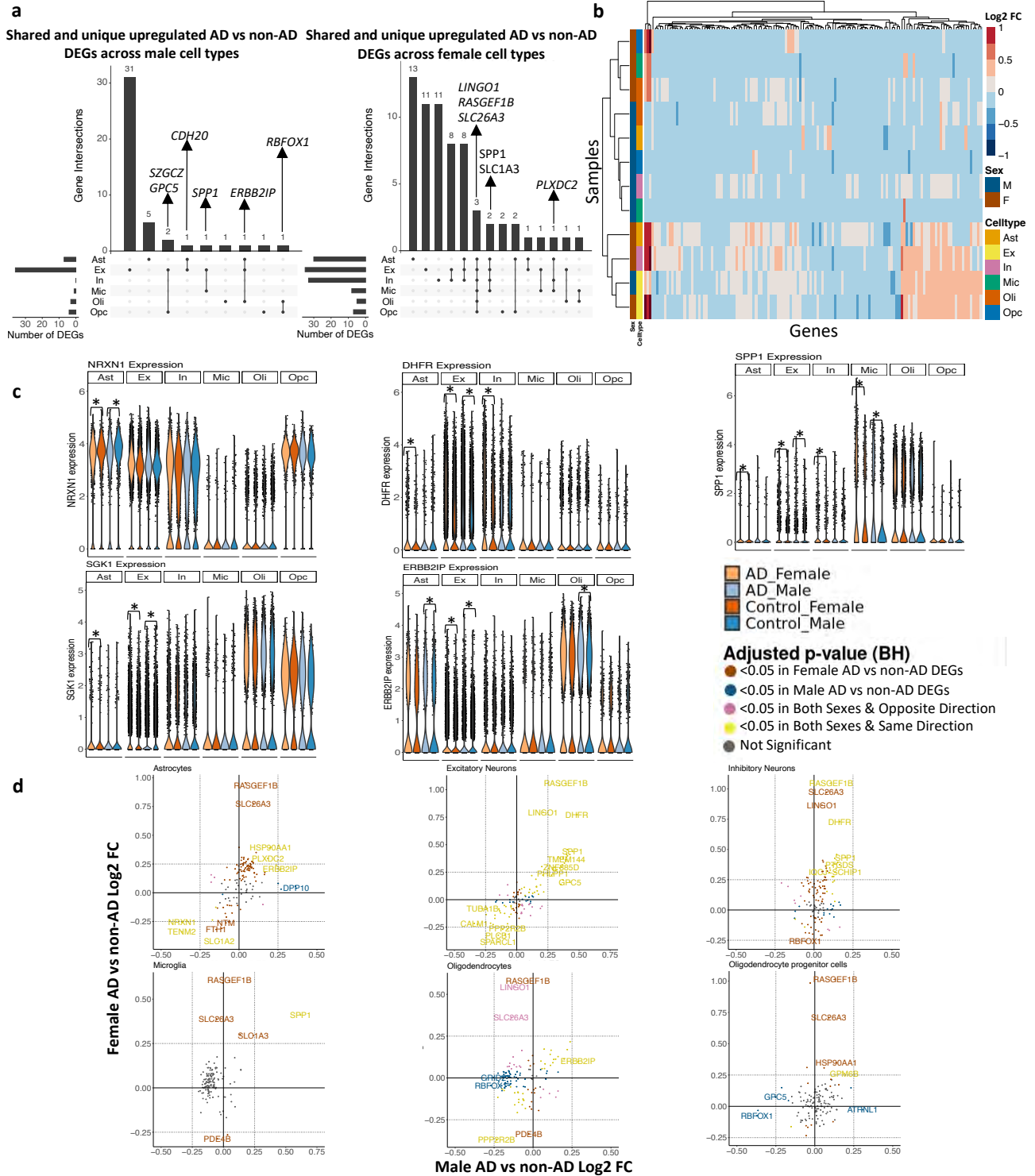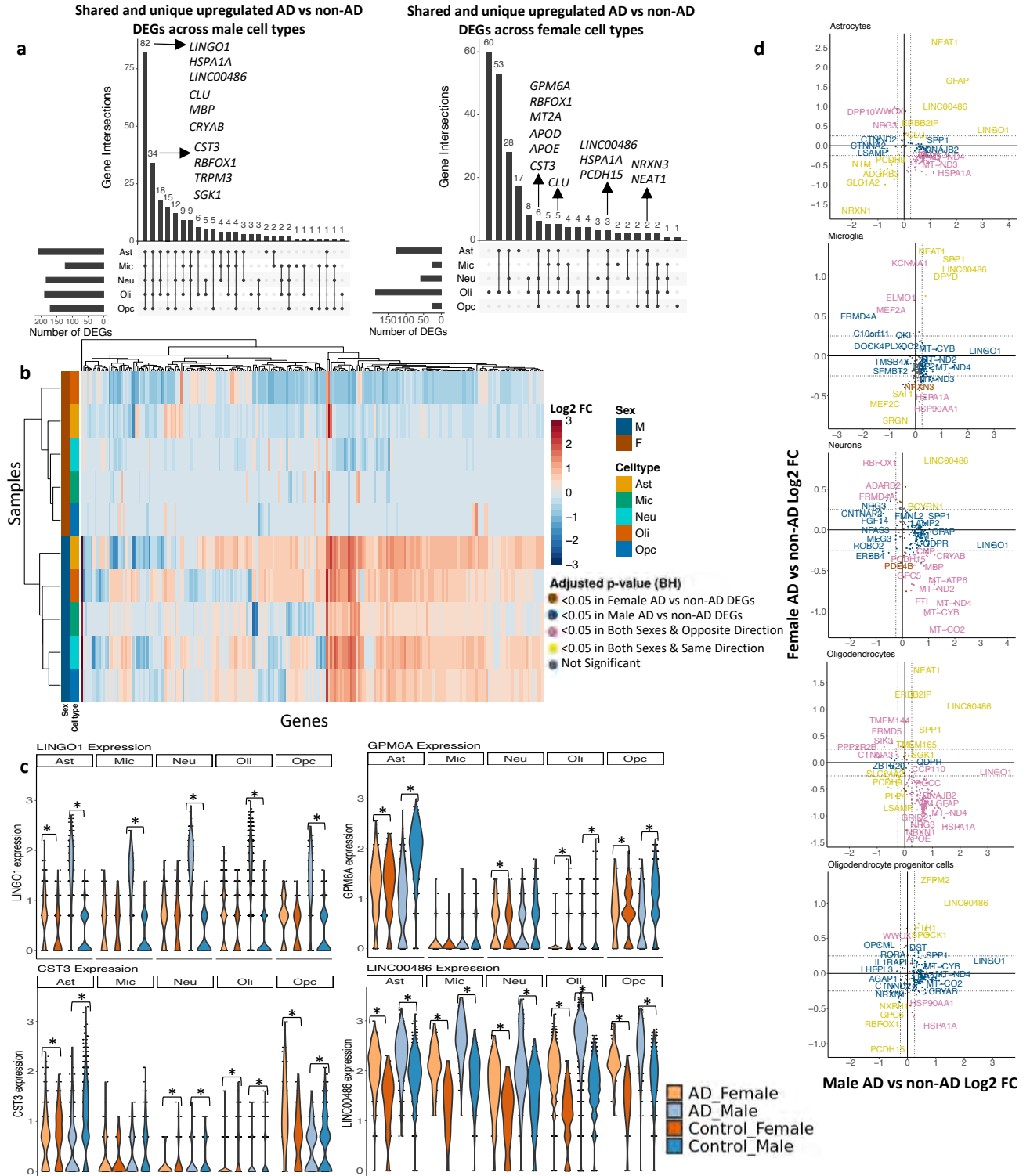
Figure 3.2: Sex-stratified cell type-specific differential gene expression signatures in the prefrontal cortex. a. Upset plots indicating intersections of AD versus non-AD differentially expressed genes (DEGs) (Benjamini-Hochberg (BH) adjusted p-value < 0.05 and absolute log2 fold change (FC) > 0.25) across cell types. Rows correspond to cell types. The bar chart shows the number of single and common sets of DEGs across cell types. Single filled dots represent a unique set of DEGs for the corresponding cell type. Multiple filled black dots connected by vertical lines represent common sets of DEGs across cell types, b. log2 FC scores of all genes in the DE analysis clustered by cell type and sex, c. *LINGO1, PLXDC2, SPP1, RBFOX1, ERBB21P* expression. Asterisks represent meeting both significance and absolute log2 FC thresholds. Colors correspond to sex and AD status, d. Pairwise DEG plots of DEGs in male and female samples using log2 FC scores. Genes shown are significant and have a log2 FC > 0.25 in at least one sex. Colors indicate significance level of DEGs and whether DEGs are unique or shared by both sexes.
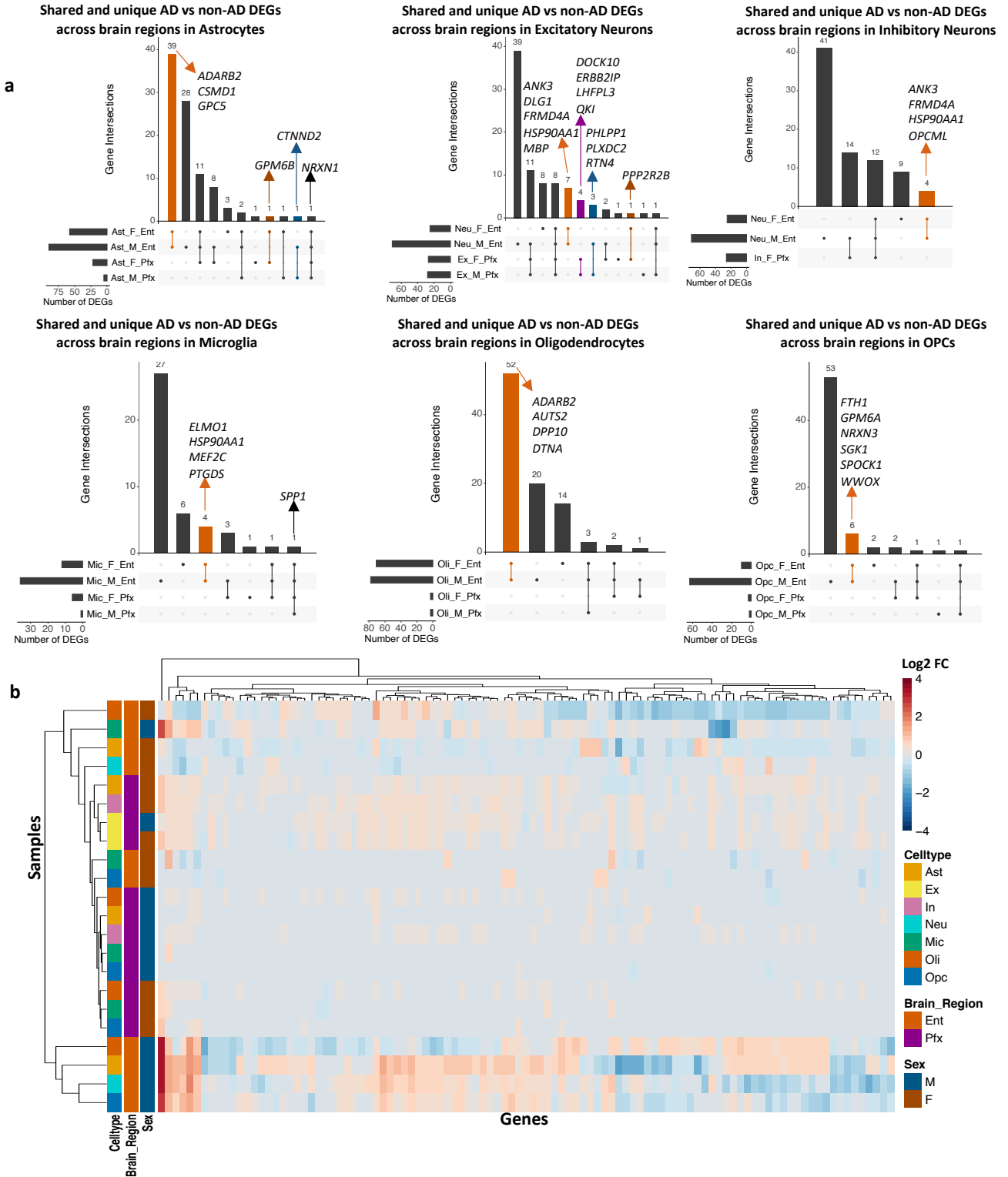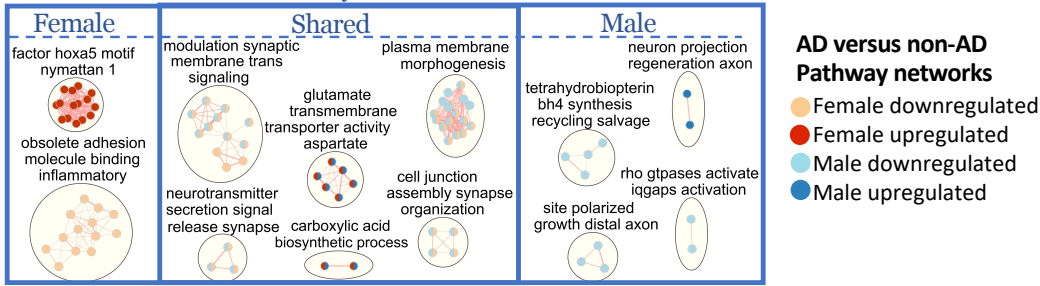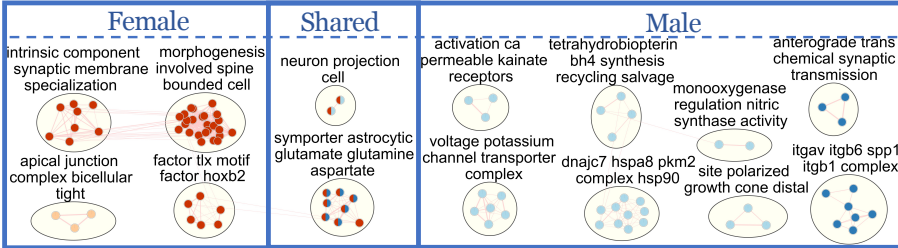
122

Figure 3.3: Sex-stratified cell type-specific differential gene expression signatures in the entorhinal cortex. a. Upset plots indicating intersections of AD versus non-AD differentially expressed genes (DEGs) (Benjamini-Hochberg (BH) adjusted p-value < 0.05 and absolute log2 fold change (FC) > 0.25) across cell types. Rows correspond to cell types. The bar chart shows the number of single and common sets of DEGs across cell types. Single filled dots represent a unique set of DEGs for the corresponding cell type. Multiple filled black dots connected by vertical lines represent common sets of DEGs across cell types, b. log2 FC scores of all genes in the DE analysis clustered by cell type and sex, c. *LINGO1*, *GPM6A, CST3, LINC00486* expression. Asterisks represent meeting both significance and absolute log2 FC thresholds. Colors correspond to sex and AD status, d. Pairwise DEG plots of DEGs in male and female samples using log2 FC scores. Genes shown are significant and have a log2 FC > 0.25 in at least one sex. Colors indicate significance level of DEGs and whether DEGs are unique or shared by both sexes.

Figure 3.4: Sex-stratified cell type-specific disease signatures across brain regions. a. Upset plots indicating intersections of AD versus non-AD differentially expressed genes (DEGs) (Benjamini-Hochberg (BH) adjusted p-value < 0.05 and absolute log2 fold change (FC) > 0.25) within cell types across brain region and sex. Rows correspond to brain region and sex pairings. The bar chart shows the number of single and common sets of DEGs across brain regions and sex. Single filled dots represent a unique set of DEGs for the corresponding brain region and sex. Multiple filled black dots connected by vertical lines represent common sets of DEGs across brain region and sex. Bar chart colors correspond to whether DEGs are shared by brain regions or sex using the bottom right key, b. log2 FC scores of all genes in the DE analysis of both brain regions clustered by cell type, brain region, and sex.
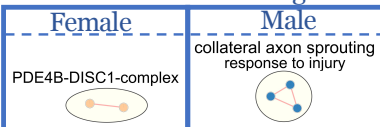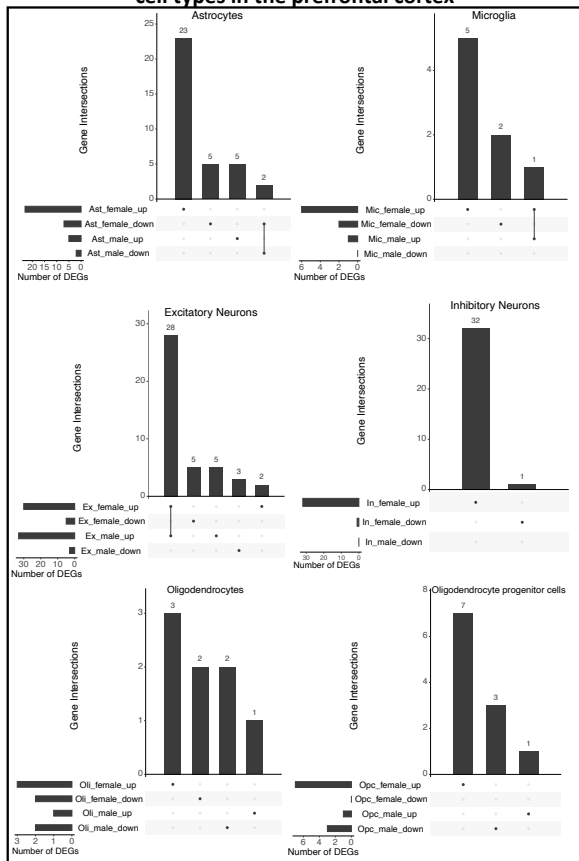
Figure 3.5: Enriched disease pathway networks in female and male neurons and microglia. AD compared to non-AD functionally enriched pathways with a Benjamini-Hochberg (BH) adjusted p-value < 0.05 clustered into biological themes for a. excitatory and b. inhibitory neurons from the prefrontal cortex, c. neurons from the entorhinal cortex, and microglia from the d. prefrontal, and e. entorhinal cortices. Lines represent gene set overlaps with magnitude showed by thickness.
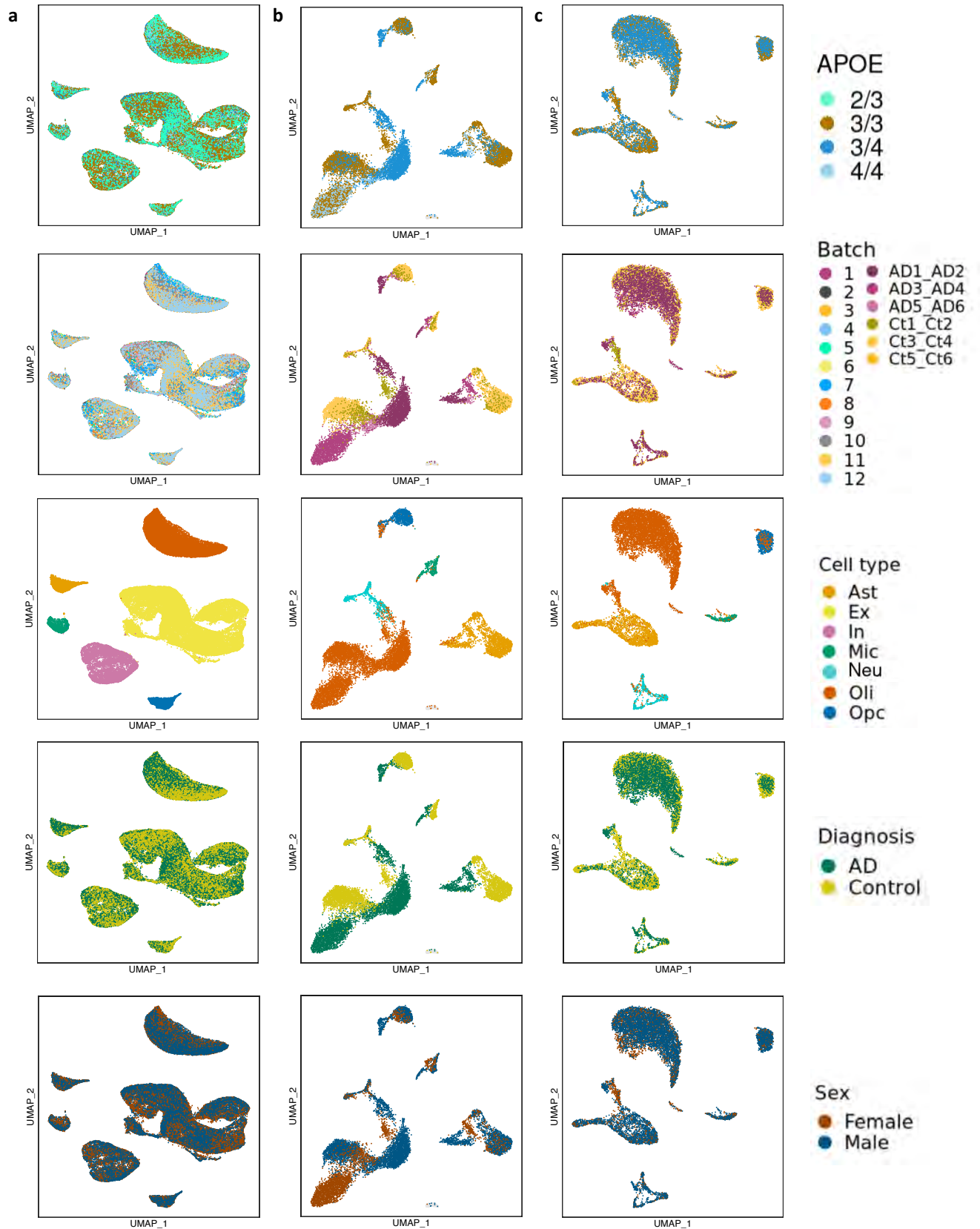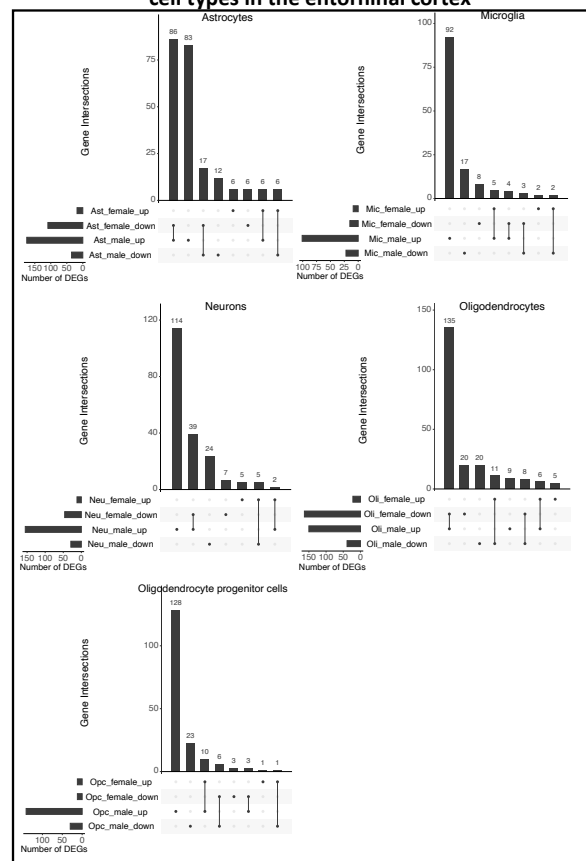
Supplementary Figure 3.1: Dimensionality reduction of prefrontal and entorhinal cortices cohort cells by *APOE* genotype, batch, cell type, diagnosis, and sex. a. prefrontal cortex UMAP, b. entorhinal cortex UMAP before batch correction, c. entorhinal cortex UMAP after batch correction.
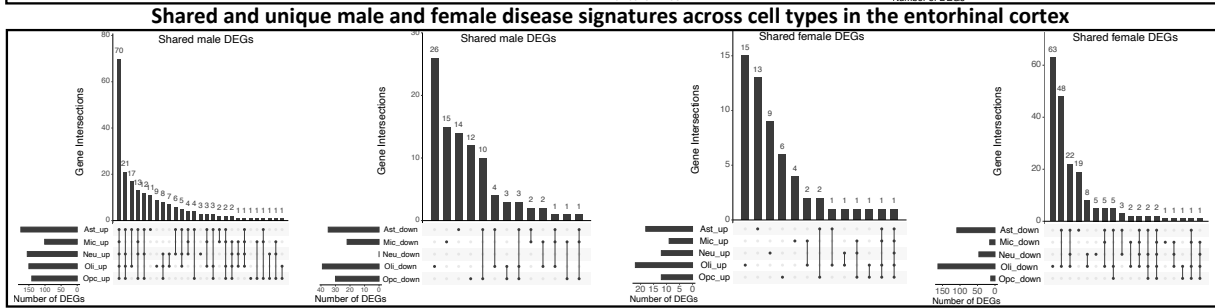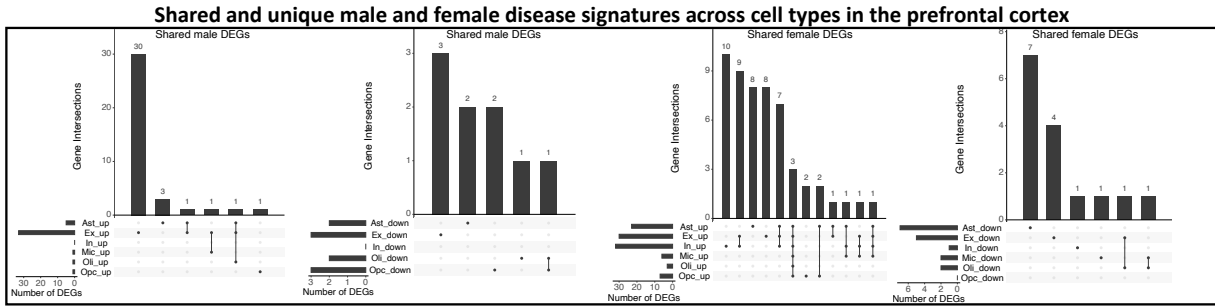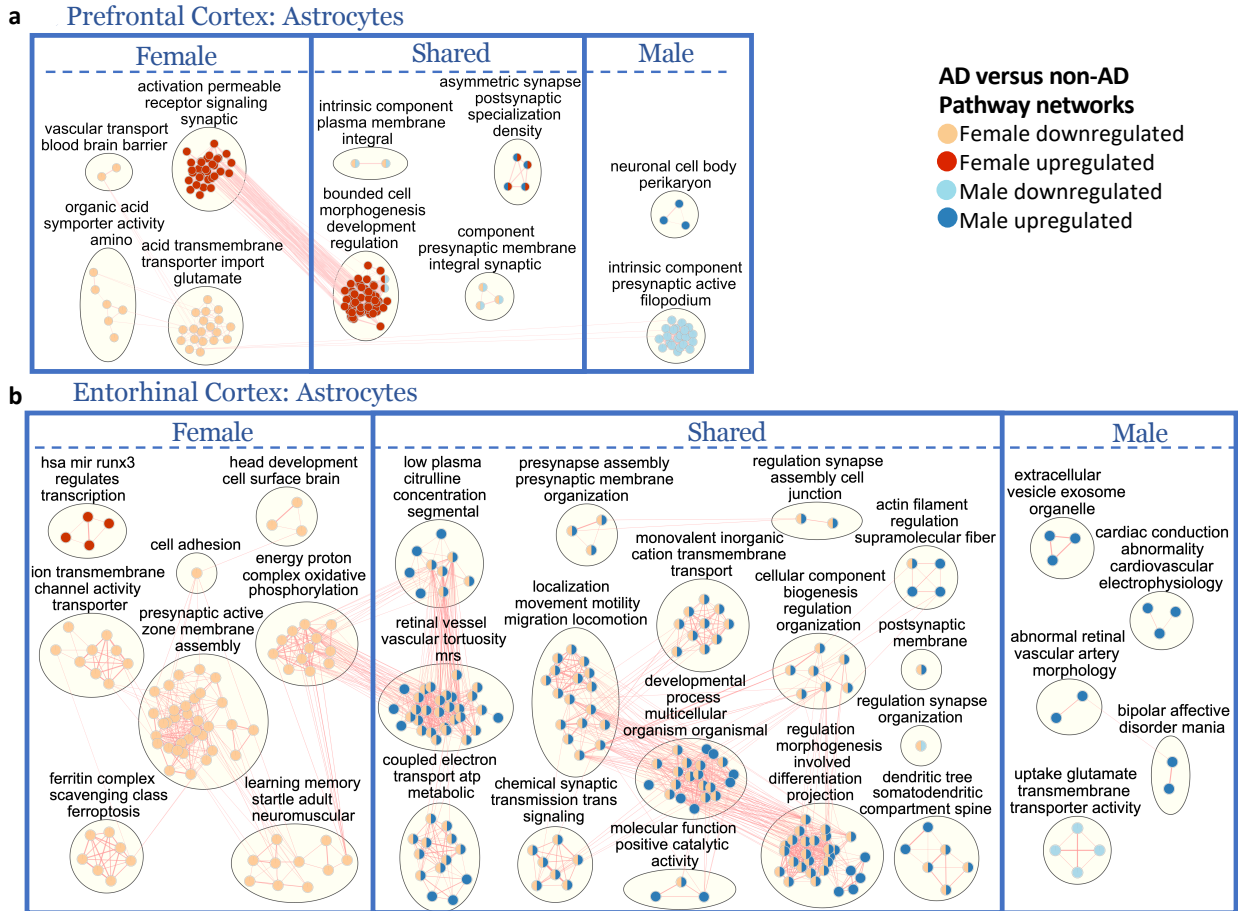
Supplementary Figure 3.2: Shared and unique male and female disease signatures across and within cell types in the prefrontal and entorhinal cortices. Upset plots indicating intersections of AD versus non-AD differentially expressed genes (DEGs) (Benjamini-Hochberg (BH) adjusted p-value < 0.05 and absolute log2 fold change (FC) > 0.25). Rows correspond to cell type, direction of gene expression change, and sex in respective plots. The bar chart shows the number of single and common DEGs. Single filled dots represent a unique set of DEGs, and multiple filled black dots connected by vertical lines represent common DEGs.

a Prefrontal Cortex: Astrocytes

Female | Shared | Male

AD versus non-AD
Pathway networks
- Female downregulated
- Female upregulated
- Male downregulated
- Male upregulated

b Entorhinal Cortex: Astrocytes
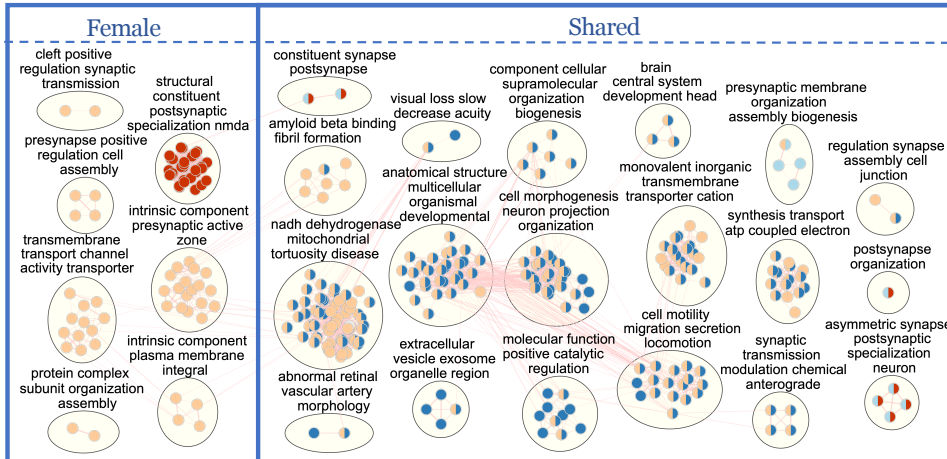
Female | Shared | Male

Supplementary Figure 3.3: Enriched disease pathway networks in female and male astrocytes. AD compared to non-AD functionally enriched pathways with a Benjamini-Hochberg (BH) adjusted p-value < 0.05 clustered into biological themes for astrocytes in a. prefrontal, and b. entorhinal cortices. Lines represent gene set overlaps with magnitude showed by thickness.
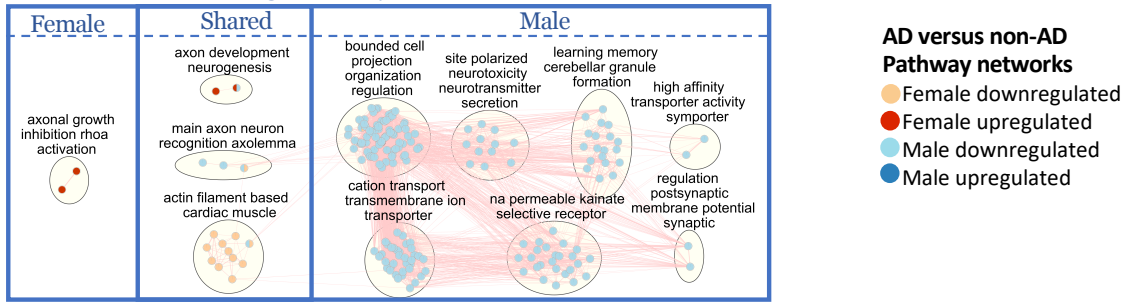
Supplementary Figure 3.4: Enriched disease pathway networks in female and male oligodendrocytes and OPCs. AD compared to non-AD functionally enriched pathways with a Benjamini-Hochberg (BH) adjusted p-value < 0.05 clustered into biological themes for a, b. oligodendrocytes and c, d. oligodendrocyte progenitor cells in prefrontal and entorhinal cortices. Lines represent gene set overlaps with magnitude showed by thickness.

CHAPTER 4

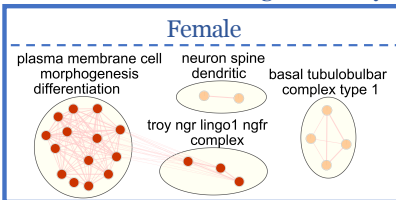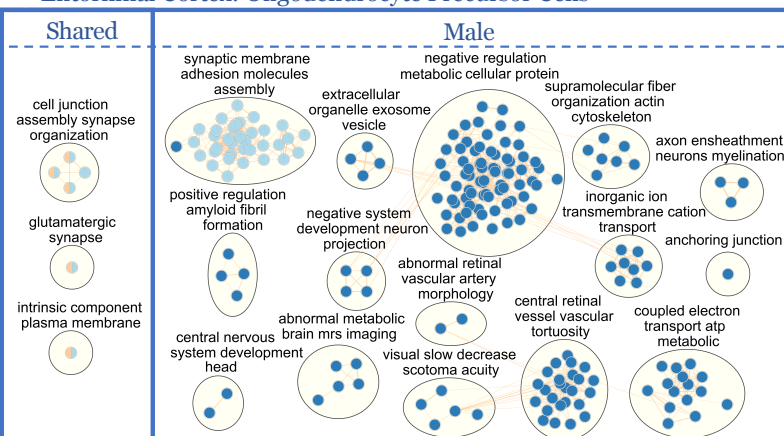Conclusions and Perspectives

A primary challenge in developing effective treatments for Alzheimer's disease (AD) is its heterogeneity in clinical phenotypes and genetic risk factors. Despite evidence of APOE4 being identified as the largest genetic risk factor, and sex differences being documented as a major source of heterogeneity in AD, molecular signatures underlying these findings remain elusive. This dissertation presents great examples of approaches to uncover the molecular basis of Alzheimer's Disease (AD) as it relates to *APOE* genotype and sex. Here, we leverage publicly available transcriptomic datasets to explore *APOE* genotype-specific disease-related changes on a bulk and single-cell level, and sex-specific disease-related changes on a single-cell level.

In chapter 1, we performed a case versus control APOE4-stratified analysis in separate bulk RNA-sequencing (RNA-Seq) datasets spanning 7 brain regions (temporal cortex, cerebellum, dorsolateral prefrontal cortex, anterior prefrontal cortex, posterior superior temporal gyrus or Wernicke's area, perirhinal cortex, and inferior frontal gyrus or Broca's area) containing 494 AD and 262 non-demented controls. For this stratified analysis, we examine disease-related changes in APOE4-negative ("E4NEG": APOE3/3 (homozygous for allele ε3)) and APOE4-positive ("E4POS": APOE3/4 (heterozygous ε3/ε4) and APOE4/4 (homozygous for allele ε4)) samples by comparing AD to control samples within each subgroup. In our observations, we identified transcriptomic changes based on the presence of APOE4, some of which are shared across brain regions. We observed neuroinflammatory pathways to be present in all samples. We observed alterations in stress response, hormone and receptor signaling, and epigenetic regulation in

E4NEG samples, and alterations in metabolic changes, lipid metabolism, clearance and recovery from deleterious events, and ion, iron, and vitamin homeostasis in E4POS samples.

In chapter 2, we sought to uncover AD-related changes based on *APOE* genotype on a single cell level, which can unmask the more complex interactions and contributions of different brain cell types in AD. In the past years, two papers (Mathys et al. Nature, 2019 & Grubman et al., Nature Neuroscience, 2019) published the first single-cell transcriptomics human AD datasets. These papers provided necessary insights into AD molecular signatures in astrocytes, microglia, neurons, oligodendrocytes, and oligodendrocyte progenitor cells; however, neither of them examined expression differences considering different configurations of *APOE* variants. In this chapter, we leverage these single-cell AD datasets from human brain samples, including nearly 55,000 cells from the prefrontal and entorhinal cortices, and take a stratified analysis approach to elucidate *APOE* genotype-specific transcriptomic signals in AD. We test for AD relative to neurotypical transcriptomic differences in APOE3/3 and APOE3/4 samples. We then examined differences in the perturbed genes and pathway networks, whether these changes were shared by more than one cell type, whether these changes were specific to one of the stratified *APOE* genotypes, and whether these patterns exist in more than one brain region. Through exploring *APOE* in this manner and comparatively observing patterns in more than one brain region, we observed changes in biological processes pertaining neuroinflammatory, metabolic, synaptic regulation, myelination, and stress response activities. We observed consistent transcriptomic differences relative to the *APOE* variants studied to take place primarily in non-neuronal cell types, and we observed more gene signatures to be shared by more than one cell type in APOE3/4 samples.

In chapter 3, we leveraged the same two single-cell from chapter 2 to uncover sex-specific precise molecular targets in each functionally distinct brain cell type across different brain regions. Previous bulk human brain sequencing studies have revealed transcriptomic differences in dysregulated molecular pathways related to energy production, neuronal function, and immune response between the two sexes (**Paranjape et al, 2020 (preprint); Appendix A**); however, the sex differences in disease mechanisms are yet to be examined comprehensively on a single-cell level. Here, we utilized nearly 74,000 cells from human prefrontal and entorhinal cortex samples and performed a case vs control sex-stratified analysis in the same cell-types as in chapter 2. We identified gene expression changes and subsequent perturbed functional pathways in AD relative to neurotypical controls in each cell type separately for females and males while accounting for the *APOE* genotype as a covariate. We reported disease transcriptomic changes and enriched pathway networks that were either shared or uniquely observed in males and females and compared sex differences between both brain regions. We observed prominent sex-dependent transcriptomic perturbations in AD relative to healthy state glial cells of the prefrontal cortex and relatively similar perturbed states between sexes in AD relative to healthy state neurons, indicating that glial cell biology in disease progression might account for most of the sex differences in clinical phenotypes. In the entorhinal cortex, we observed shared DEGs by both sexes but in opposite direction of regulation involving biological processes that pertain synaptic regulation, cellular morphogenesis, myelination/remyelination, and metabolic energy production, indicating different mechanisms of neurodegeneration based on sex.

While we were able to examine *APOE* genotype- and sex- specific changes across multiple brain regions, some limitations still exist. In chapter 1, we encountered batch effects in five out of our

seven datasets. To mitigate this limitation, we performed batch correction, confirmed batch correction using clustering and dimensionality reduction, and included batch as the first covariate in all differential gene expression analyses. We also observed a limited distribution of samples across *APOE* genotypes and diagnosis, where we had more E4NEG samples, and much fewer controls compared to cases especially in the E4POS group. Finally, with our data type, we were could not observe more complex disease-related transcriptomic changes in AD.

In chapter 2, we also observed limited *APOE* genotypes, which restricted our analysis to APOE3/3 and APOE3/4 samples. Each dataset contained only one APOE3/4 control, which was a male sample in both cases. So, we performed a sensitivity analysis in males of the prefrontal cortex cohort and were able to confirm similar perturbed gene profiles and stronger clustering by *APOE* genotype than cell type identity. We also encountered constraints in the entorhinal cortex dataset: small sample size, imbalanced *APOE* genotype and sex covariates across diagnoses, and a batch effect due to the design of the Grubman et al. study. To mitigate these limitations, we followed an integration workflow, confirmed appropriate batch correction by dimensionality reduction, and included sex as a covariate in our model for differential expression to account for batch while avoiding the collinearity observed with including batch.

In chapter 3, we also faced limitations like chapter 2, which utilized the same datasets. These involved small sample sizes and batch effects in the entorhinal cortex data introduced by the study design, which we overcame by performing data integration and including *APOE* genotype as a covariate in our DGE analysis to account for batch and avoid collinearity in our model. Additionally, the datasets were not *APOE* genotype matched, and although we accounted for

APOE genotype as a covariate in the DGE analysis, the interactions of sex and *APOE* genotype may still explain trends that we observe. Also, literal biological sex could be a misleading classifier for trans* individuals. Thus, in future studies, a properly powered study of differences between male versus female versus recipients of testosterone- versus estrogen-focused hormone replacement therapy might help narrow down a genetic versus hormonal basis of DEGs deemed sexually dimorphic. In both chapter 2 and 3, due to constraints in the datasets, we did not explore further molecular profiles unique to each brain region and their implications for the spread of AD pathology.

Moreover, all datasets we surveyed contained samples primarily from individuals of Caucasian descent, which limits translating findings to other populations. Also, with using publicly available datasets, samples were processed and annotated differently, so we tried to combat this by identifying cases and controls based on similar criteria. We also did not stratify our analysis by age or disease severity, so we could not infer disease related transcriptomic changes based on them. Overall, the nature of our analyses only allowed for association of transcriptomic changes with *APOE* genotype, so links to causality might be hypothesized, but additional follow-up are needed to prove any such potential links.

Collectively, we evaluated cell-type-specific molecular heterogeneity in AD across *APOE* variants and sex to better understand the complex AD molecular profiles. We hope that more types of transcriptomic datasets become available from diverse populations from more brain regions of diverse sets of individuals, across different ages, racial and ethnic backgrounds, with a greater diversity of *APOE* genotypes and disease severity. With more robust datasets, future

investigations may: 1) integrate datasets from multiple sources and brain regions to examine AD-relevant changes based on covariates like *APOE* genotype, age, sex, and severity of pathology across brain regions. 2) identify molecular profiles associated with the spread of AD-pathology, and 3) validate findings with relevant multiomic studies through *in silico*, and subsequently *in vitro* and *in vivo* methods.

Furthermore, advances in sequencing technology and the influx of biomedical data have increased our potential to strengthen clinical medicine and drug development. With approaches such as integrating different omics datasets, exploring electronic health medical records, and predictive modeling using artificial intelligence, we can facilitate improvements in diagnostics and treatment strategies for AD. Despite the limitations highlighted, this dissertation identifies disease-relevant transcriptomic perturbations on a bulk and single-cell level that suggest possible mechanisms and vulnerable cell subpopulations pertaining *APOE* genotype and sex in AD pathogenesis, which have implications for diagnosing and treating AD. Plainly, with incorporating these risk factors into future investigations, we can identify novel biomarkers related to the differences in neurodegeneration that arise based on *APOE* genotype and sex. Additionally, by including risk factors like *APOE* genotype and sex, and exploring tailored therapies in drug discovery efforts, we can prioritize worthwhile targets, reduce the length of time, failures and costs associated with traditional drug development, and overtime deliver more efficacious therapies. Overall, the findings here highlight the importance of incorporating *APOE* genotype and sex in future multiomic exploration of AD pathogenesis and progression. We hope this work spurs further investigations as more robust datasets become available.

APPENDICES

Here are some of the others works I contributed to during my time at UCSF:

APPENDIX A: RESEARCH COLLABORATIONS

1. Paranjpe MD, **Belonwu S**, Wang JK, Oskotsky T, Gupta A, Taubes A, Zalokusky K, Paranjpe I, Glicksberg BS, Huang Y, Sirota M. Sex-Specific Cross Tissue Meta-Analysis Identifies Immune Dysregulation in Women with Alzheimer's Disease. (Under Review) https://www.biorxiv.org/content/10.1101/2020.04.24.060558v1

   Meta-analysis of RNA-sequencing data from blood and brain samples to study molecular signals in Alzheimer's disease based on sex.

APPENDIX B: ARTICLES
1. Belonwu, Stella. "How Artificial Intelligence is Bolstering Healthcare Advances in Therapeutics Development." BCBA Viewpoints. June 14, 2019. https://biotechconnectionbay.org/viewpoint/how-artificial-intelligence-is-bolstering-healthcare-advances-in-therapeutics-development/

   Wrote an article on the intersection of Artificial Intelligence (AI) and healthcare advancement based on interviews with four professionals in the AI and drug discovery space.

APPENDIX C: PODCAST EPISODES
1. "CRISPR: The Unauthorized Biography" by Liron Noiman, **Stella Belonwu**, and Ben Mansky. June 3, 2019. https://carrytheoneradio.com/episodes/crispr

   Interviewed Dr. Martin Kampmann and Dr. Joseph Bondy-Denomy about the gene-editing tool CRISPR.

2. "Whatever Happened with Zika?" by Katie Cabral and **Stella Belonwu**. September 3, 2019. https://carrytheoneradio.com/episodes/zika

   Episode discussing the history of Zika covering its origins, to the widespread epidemic of 2015-2016, to the personal impact it had on those who were affected, to its disappearance, and to the possibility of preventing a future Zika epidemic through a vaccine.

3. "The Mosquito Menace" by Katie Cabral and **Stella Belonwu**. September 19, 2019. https://carrytheoneradio.com/episodes/mosquitoes

Episode about mosquito-borne illnesses, where we spoke with a variety of experts that are working on mosquito vector control in the Caribbean and Latin America.

4. "Shedding Light on Dark Matter" by **Stella Belonwu**, Anna Lipkin, Cindy Liu and Liron Noiman. November 12, 2019. https://carrytheoneradio.com/episodes/dark-matter

   Talked to Dr. Neta Bahcall, the Eugene Higgins Professor of Astrophysics at Princeton University about dark matter.

5. "Once Upon an Arc" by **Stella Belonwu**, Rebecca Fang, Deanna Necula, and Ben Mansky. January 21, 2020. https://carrytheoneradio.com/episodes/once-upon-an-arc

   An interview with Dr. Mercedes Paredes, an Assistant professor of neurology at UCSF about developmental changes in the brain in the period before and just after birth.

6. "Beleaf What You Wanna Beleaf" by **Stella Belonwu** and Seesha Takagishi. April 20, 2020. https://carrytheoneradio.com/episodes/2020/4/20/beleaf-what-you-wanna-beleaf

   Interview with two cannabis researchers Dr. Steven Laviolette and Dr. Salomeh Keyhani about the constituents of cannabis, and investigations into its relation to neuropsychiatric disorders, the funding landscape for research, the cannabis industry, public perception, and cardiovascular health effects

7. "Young Scientist Spotlight 4: Jhia Jackson" by **Stella Belonwu**. May 18, 2020. https://carrytheoneradio.com/episodes/spotlight-4

   Interview with Jhia Jackson, a sociology PhD candidate at UCSF about her educational journey, research on pediatric palliative and hospice care, her career as a professional dancer, being black at UCSF, cats, and ways to stay sane in grad school.

8. "Young Scientist Spotlight 11: Oluwasegun Akinniyi" by **Stella Belonwu**. November 30, 2020. https://carrytheoneradio.com/episodes/2020/11/30/young-scientist-spotlight-11-oluwasegun-akinniyi

   Spoke to Oluwasegun Akiniyi, a bioengineering master's student at Obafemi Awolowo University in Ile Ife, Osun State, Nigeria who is developing rehabilitation devices for stroke victims about his research and career interests.

9. "Art is Science is Art (Part 1): The Process" by **Stella Belonwu**, Celia Ford, and Devika Nair. January 19, 2021. https://carrytheoneradio.com/episodes/2021/1/19/art-is-science-is-art-part-1-the-process

   Interview with choreographer-slash-educator Suba Subramaniam and computational-biologist-slash-generative-artist Dr. Alex Naka about how they each blended science and

art through their own career journeys, how the methodology of science can be used as an artistic tool, and how the creative process drives scientific curiosity.

10. "Art is Science is Art (Part 2): The Impact" by **Stella Belonwu**, Celia Ford, and Devika Nair. February 16, 2021. https://carrytheoneradio.com/episodes/2021/2/16/art-is-science-is-art-part-2-the-impact

    Interview with visual artists and science communicators Kelly Montgomery and Sophie Wang about how they've used art to communicate big ideas, and how scientists can approach making knowledge more accessible.

11. "Global Health in the Time of COVID: Ramses Escobado, Jess Celentano, and Dr. Mike Reid" by **Stella Belonwu** and Maggie Colton. March 16, 2021. https://carrytheoneradio.com/episodes/2021/3/16/global-health-in-the-time-of-covid-ramses-escobado-jess-celentano-and-dr-mike-reid

    Episode about contact tracing efforts during the COVID-19 pandemic.

12. "Towards a Sustainable Earth" by Rachel Rock and **Stella Belonwu**. April 12, 2021. https://carrytheoneradio.com/episodes/2021/4/11/towards-a-sustainable-earth

    Interview with Dr. Sheri Weiser, a physician-scientist at UCSF with a long history of researching food insecurity and climate justice about her path to sustainability activism and some initiatives she has led in the UC system to tackle climate change with a key focus on environmental justice and equity.

13. "Psychedelics Down to a Tea" by **Stella Belonwu**, Cindy Liu, and Seesha Takagishi, May 24, 2021. https://carrytheoneradio.com/episodes/2021/5/24/psychedelics-tea

    Interview with Dr. Victoria Hale about the re-introduction of psychedelics for use in treating neuropsychiatric illnesses, her ayahuasca tea company, and social and political factors surrounding these changes.

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:

Stella Belonwu

E0550F40B7044F3...   Author Signature

5/31/2021

Date

139