

UCSF

UC San Francisco Electronic Theses and Dissertations

Title

Heterogeneities in Alzheimer's Disease and Related Dementias: Elucidating the Genetic and Transcriptomic Underpinnings

Permalink

<https://escholarship.org/uc/item/3n34v3gr>

Author

Jonson, Caroline

Publication Date

2024

Peer reviewed|Thesis/dissertation

Heterogeneities in Alzheimer's Disease and Related Dementias: Elucidating the Genetic and Transcriptomic Underpinnings

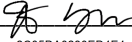
by
Caroline Jonson

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

in
Pharmaceutical Sciences and Pharmacogenomics

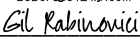
in the
GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:

2C05DA0230ED4E4... Jennifer Yokoyama
Chair

DocuSigned by:

400... Mike Nalls

DocuSigned by:

8C530F80377E4D7... Gil Rabinovici

Committee Members

Copyright 2024
by
Caroline Jonson

Contributions

The research in the following thesis chapters were performed under the guidance of Dr. Jennifer Yokoyama. Additional guidance and contributions were provided by Dr. Gil Rabinovici, Dr. Mike A Nalls, Dr. Daniel Sirkis, Dr. Dena Dubal, Hampton Leonard, Kristin Levine, Dr. Shea Andrews, and other members of the Yokoyama lab.

Chapter 1 is a reprint of the material here:

Davis EJ, **Solsberg CW**, White CC, Miñones-Moyano E, Sirota M, Chibnik L, Bennett DA, De Jager PL, Yokoyama JS, Dubal DB. Sex-Specific Association of the X Chromosome With Cognitive Change and Tau Pathology in Aging and Alzheimer Disease. *JAMA Neurol.* 2021 Oct 1;78(10):1249-1254. doi: 10.1001/jamaneurol.2021.2806. PMID: 34424272; PMCID: PMC8383157.

Chapter 2 Paper 1 is a reprint of the material here:

Lake, J., **Warly Solsberg, C.**, Kim, J.J. et al. Multi-ancestry meta-analysis and fine-mapping in Alzheimer's disease. *Mol Psychiatry* **28**, 3121–3132 (2023). <https://doi.org/10.1038/s41380-023-02089-w>

Chapter 2 Paper 2 is a reprint of the preprint material here:

Jonson C, Levine KS, Lake J, Hertslet L, Jones L, Patel D, Kim J, Bandres-Ciga S, Terry N, Mata IF, Blauwendraat C, Singleton AB, Nalls MA, Yokoyama JS, Leonard HL. Assessing the lack of diversity in genetics research across neurodegenerative diseases: a systematic review of the GWAS Catalog and literature. medRxiv [Preprint]. 2024 Jan 9:2024.01.08.24301007. doi: 10.1101/2024.01.08.24301007. PMID: 38260595; PMCID: PMC10802650.

Chapter 3 Paper 1 is a reprint of the material here:

Sirkis DW, **Warly Solsberg C**, Johnson TP, Bonham LW, Sturm VE, Lee SE, Rankin KP, Rosen HJ, Boxer AL, Seeley WW, Miller BL, Geier EG, Yokoyama JS. Single-cell RNA-seq reveals alterations in peripheral CX3CR1 and nonclassical monocytes in familial tauopathy. Genome Med. 2023 Jul 18;15(1):53. doi: 10.1186/s13073-023-01205-3. PMID: 37464408; PMCID: PMC10354988.

Chapter 3 Paper 2 is a reprint of the preprint material here:

Sirkis DW, **Solsberg CW**, Johnson TP, Bonham LW, Oddi AP, Geier EG, Miller BL, Rabinovici GD, Yokoyama JS. Expansion of highly interferon-responsive T cells in early-onset Alzheimer's disease. bioRxiv [Preprint]. 2024 Jan 10:2023.09.26.559634. doi: 10.1101/2023.09.26.559634. PMID: 37823036; PMCID: PMC10563505.

Heterogeneities in Alzheimer’s Disease and Related Dementias: Elucidating the Genetic and Transcriptomic Underpinnings

Caroline Jonson

Abstract

Despite the tremendous burden of Alzheimer’s disease and related dementias (ARD) on patients, their families, and the public health system, there are still no effective treatments for the disease. Seeking a better understanding of important disease-subgroup-specific associations in a highly heterogeneous patient population, like those diagnosed with ARD, can help facilitate the development of new therapeutics by enhancing clinical trial design. This thesis focuses on elucidating the genetic and transcriptomic underpinnings of the heterogeneities in ARD and where these subgroups mechanistically converge.

Chapter 1 focuses on features driving sex differences in AD progression to help reveal what protects some individuals and what makes others more vulnerable, ultimately informing the development of personalized therapeutics that benefit both sexes.

Chapter 2 presents a meta-analysis of Alzheimer's disease genetics across multiple ancestries and highlights the current lack of diversity in neurodegenerative disease genetic research, suggesting a need for more inclusive studies.

Chapter 3 links peripheral immune changes to familial tauopathy and early-onset Alzheimer's, showing altered monocytes in tauopathies and increased interferon-responsive T cells in EOAD. This chapter highlights the importance of the immune system in ARD.

TABLE OF CONTENTS

GLOSS TO CHAPTER 1.....	1
CHAPTER 1	3
1.1 Abstract	4
1.2 Introduction	6
1.3 Methods	7
1.4 Results	7
1.5 Discussion	9
1.6 Conclusions	11
1.7 Figures	12
1.8 Tables	13
1.9 Supplemental Online Content	15
1.10 Article Information.....	15
1.11 References	17
GLOSS TO CHAPTER 2.....	19
CHAPTER 2 PAPER 1	21
2.1.1 Abstract	22
2.1.2 Introduction	23
2.1.3 Discussion	31
2.1.4 Online Methods	39
2.1.5 Figures	47
2.1.6 Tables	68
2.1.7 Supplemental Online Content, Data and Code Availability.....	70
2.1.8 Article Information.....	70
2.1.9 References	74

CHAPTER 2 PAPER 2	85
2.2.1 Abstract	86
2.2.2 Introduction	88
2.2.3 Methods.....	88
2.2.4 Results	90
2.2.5 Discussion	101
2.2.6 Conclusion.....	106
2.2.7 Figures and Tables.....	107
2.2.8 Supplemental Online Content	113
2.2.9 References	114
GLOSS TO CHAPTER 3.....	123
CHAPTER 3 PAPER 1	125
3.1.1 Abstract	126
3.1.2 Introduction	128
3.1.3 Methods.....	130
3.1.4 Results	138
3.1.5 Discussion	146
3.1.6 Conclusions	153
3.1.7 Article Information.....	154
3.1.8 Figures and Tables.....	156
3.1.8 Supplemental Online Content, Data and Code Availability.....	171
3.1.9 References	172
CHAPTER 3 PAPER 2	186
3.2.1 Abstract	187
3.2.2 Background	187
3.2.3 Methods.....	189

3.2.4 RESULTS	191
3.2.5 Discussion	195
3.2.6 Article Information	197
3.2.7 Figures and Tables	199
3.2.8 Supplemental Online Content	209
3.2.9 References	210

List of Figures

Figure 1.1 Association of X Chromosome Genes with Cognitive Change in Aging and AD.....	12
Figure 2.1.1 Outline of multi-ancestry meta-analysis.....	47
Figure 2.1.2 Summary of multi-ancestry meta-analysis.....	48
Figure 2.1.3 Graphical summary of heterogeneity at ADD genetic risk loci.....	49
Figure 2.1.4 Graphical summary of genetic risk scores.....	50
Figure 2.1.S1 Quantile-quantile plots and corresponding genomic inflation estimates.....	51
Figure 2.1.S2 MR - MEGA $-\log_{10}$ P-values plotted against random effects $-\log_{10}$ P-values.....	52
Figure 2.1.S3 LocusCompare plots showing colocalization at the <i>TRANK1</i> locus.....	53
Figure 2.1.S4 LocusCompare plots showing colocalization at the <i>VWA5B2</i> locus.....	54
Figure 2.1.S5 LocusZoom and forest plots for fine-mapped SNPs.....	55
Figure 2.1.S6 LocusZoom and forest plots for fine-mapped SNPs.....	56
Figure 2.1.S7 LocusCompare plots for a) <i>GRN</i> and b) <i>CTSB</i> between AD and Parkinson’s disease (PD).....	57
Figure 2.1.S8 LocusZoom and forest plots for loci that were fine-mapped to a missense variant with PP ≥ 0.3	58
Figure 2.1.S9 LocusZoom plots showing different regional architecture at the <i>SORL1</i> locus.....	59
Figure 2.1.S10 Forest plots for loci with significant heterogeneity outside of the <i>APOE</i> region.....	60
Figure 2.1.S11 Forest plots for a) <i>APOE</i> -rs429358 and b) <i>APOE</i> -rs7412.....	61
Figure 2.1.S12 a) Manhattan plot for MR - MEGA meta-analysis.....	62
Figure 2.1.S13 LocusZoom and Beta-beta plots.....	63
Figure 2.1.S14 LocusCompare at the TRANK 1 locus for BDI.....	64
Figure 2.1.S15 LocusCompare at the TRANK 1 locus for SZ.....	65
Figure 2.1.S16 Power calculation for <i>VWA5B2</i> -rs9837978 in the GARD cohort using GAS Power Calculator.....	66
Figure 2.1.S17 The first 2 ancestral principal components (PCs) created and used by MR - MEGA.....	67
Figure 2.1.1 PRISMA flow diagram for systematic review.....	107
Figure 2.2.2 Number of studies over time from 2012 to 2022.....	108
Figure 2.2.3 Cumulative count of discovered SNPs from 2012 through 2022.....	109
Figure 3.1.1 Single-cell RNA-seq reveals reductions in NCM in <i>MAPT</i> pathogenic variant carriers.....	156
Figure 3.1.2 Differential expression in <i>MAPT</i> variant carriers by cell cluster.....	157
Figure 3.1.3 STRING interactions in NCM and NK cell differentially expressed genes.....	158

Figure 3.1.4 CX3CR1 expression is reduced in peripheral myeloid and lymphoid cells in familial tauopathy.....	159
Figure 3.1.5 Confirmation of reduced <i>CX3CR1</i> expression in <i>MAPT</i> variant carrier PBMCs via ddPCR.....	160
Figure 3.1.6 Analysis of nonclassical monocyte marker genes in <i>MAPT</i> variant carriers	161
Figure 3.1.7 Analysis of <i>TMEM176A/B</i> in <i>MAPT</i> pathogenic variant carriers	162
Figure 3.1.8 Potential dysregulation of <i>C3AR1</i> in <i>MAPT</i> pathogenic variant carriers.....	163
Figure 3.1.9 Validation of the Reduction in Nonclassical Monocytes via Flow Cytometry	164
Figure 3.1.S1 Annotation of cell types from cluster markers	165
Figure 3.1.S2 Expression of FTD-associated genes in PBMCs.....	166
Figure 3.1.S3 Nonclassical monocyte abundance relative to other myeloid populations.....	167
Figure 3.1.S4 Mitochondrial genome mapping percentages and mitochondrial DEGs in selected clusters	168
Figure 3.1.S5 Additional droplet digital PCR analysis.....	169
Figure 3.1.S6 Potential dysregulation of <i>C3AR1</i> in <i>MAPT</i> pathogenic variant carriers.....	170
Figure 3.2.1 Expansion of ISAG ^{hi} T cells in EOAD characterized by scRNA-seq.....	199
Figure 3.2.2 ISAG ^{hi} T-cell marker gene expression is increased in CD4 T cells in EOAD.....	200
Figure 3.2.3 Heightened IFN response signatures in CSF T cells in MCI and AD.....	201
Figure 3.2.S1 Analysis of additional variables' relationships with ISAG ^{hi} T cells	202
Figure 3.2.S2 Reanalysis after integration of additional cognitively normal control samples	203
Figure 3.2.S3 Analysis of changes in cell-type abundance in familial tauopathy and EOAD	204
Figure 3.2.S4 Analysis of proliferating NK cells and NK cell IFN signaling in EOAD.....	205
Figure 3.2.S5 Differential expression analysis in additional PBMC types in EOAD	206
Figure 3.2.S6 ddPCR validation of CD4 T-cell isolation	207
Figure 3.2.S7 Dysregulation of IFN response genes in a mouse model of familial EOAD	208

List of Tables

Table 1.1 Demographic Information.....	13
Table 1.2 Gene expression association summary.....	14
Table 2.1.1 Summary of novel loci.....	68
Table 2.1.2 Fine-mapping results.....	69
Table 2.2.1 Largest GWAS sample size by NDD and ancestry for single and multi ancestry studies.....	110
Table 2.2.2 Genome-wide significant novel loci nominated in non-European populations.....	113
Table 3.1.3 Demographic characteristics of cohort.....	171
Table 3.2.1 Demographic and experimental information for samples used in scRNA-seq and ddPCR studies.....	209

List of abbreviations

AD, Alzheimer's disease

ADD, Alzheimer's disease and related dementias

ANC, ancestry

RESID, residual

AUC, area under the curve

BD, Bipolar disorder

RD, related dementias

AFR, African American

ALS, amyotrophic lateral sclerosis

ALSP, adult-onset leukoencephalopathy with axonal spheroids and pigmented glia

bvFTD, behavioral variant frontotemporal dementia

CAT, Center for Advanced Technology

CAR HISP, Caribbean Hispanic.

cDC, conventional dendritic cell

CDR-SB, Clinical Dementia Rating scale Sum of Boxes

CNS, central nervous system

CSF, cerebrospinal fluid

dd, droplet digital

DEG, differentially expressed gene

DLPFC Dorsolateral prefrontal cortex

EUR, European

EUR (FIN), Finnish European

eQTL, expression quantitative trait locus

FDR, false discovery rate

FE, Fixed effect

FTD, frontotemporal dementia

FTLD, frontotemporal lobar degeneration

IHG, Institute for Human Genetics

LD, linkage disequilibrium

LFC, log₂ fold-change

LPS, lipopolysaccharide

MAC, Memory and Aging Center

MAF, minor allele frequency

MAP, Rush Memory and Aging Project

MCL, Markov cluster algorithm

MR, MR - MEGA

NC, nonclassical

NDD, neurodegenerative diseases

NK, natural killer

PBMCs, peripheral blood mononuclear cells

PCA, principal component analysis

PD, Parkinson's disease

PET, positron emission tomography

PP, posterior probability

QC, quality control

RE, Random effects

RIN, RNA integrity number

ROS, Religious orders study

RT, reverse transcription

scRNA-seq, single-cell RNA sequencing

SD, standard deviation

SMR, summary-based Mendelian Randomization

SNP, single nucleotide polymorphism

SZ, Schizophrenia

UCSF, University of California, San Francisco

UMAP, uniform manifold approximation and projection

Gloss to Chapter 1.

This chapter marks the beginning of my thesis research as my first project I took on after joining the Yokoyama lab. I had spent about a year conducting a sex stratified genome wide association study and found limited autosomal genetic differences between the sexes. Jen, and other Yokoyama lab members, had previously worked with Dena Dubal on some sex chromosome studies and given biological sex is defined by the possession of various sex chromosome complements (most commonly XY or XX), we felt studying the impact of sex chromosome features on sex differences in Alzheimer's disease would be a reasonable shift for me to take.

Emily Davis, in the Dubal lab, had already downloaded an RNA-sequencing data set to begin this study, but Dena had reached out for extra hands on this project with perfect timing for me to jump in and help. While I had helped with analyses and manuscript writing during my undergraduate research, this was my first true experience with a scientific collaboration like this one. We ran our analyses by various experts in the field and I learned how important it is to me to do research as a team. Having different perspectives helps mitigate using incorrect analyses, having visualization errors, and sometimes entirely missing an interesting pattern in the data. After this project I felt fully committed to always being available as a second pair of eyes or a mentor to pass this experience forward.

This project was not the project that provided me the most opportunity for technical growth, but it jumpstarted my thesis and empowered me embrace negative results as a transition towards a different way of looking at things. Conceptually it makes sense to me, now, that from an inheritance perspective, looking at purely genetic difference on the autosomes may not be the best way to capture sex difference, but without taking the path I took, I may not have had this insight. Studying sex differences in AD opened many doors for me including the opportunity to present at an international conference in Mexico City and the confirmation that looking at disease in an individualized way was important to me. If diseases impact people differently then why were so many scientists just throwing data at algorithms instead of pausing to consider what differences in humans lead to differences in disease? After this project, I knew I wanted to focus on this general theme.

Chapter 1

Sex-Specific Association of the X Chromosome with Cognitive Change and Tau Pathology in Aging and Alzheimer Disease

Contributing Authors: Emily J. Davis, PhD; **Caroline Warly Solsberg**, BS; Charles C. White, PhD; Elena Miñones-Moyano, PhD; Marina Sirota, PhD; Lori Chibnik, PhD; David A. Bennett, MD; Philip L. De Jager, MD, PhD; Jennifer S. Yokoyama, PhD; Dena B. Dubal, MD, PhD**

**Correspondence

1.1 Abstract

Importance

The X chromosome represents 5% of the human genome in women and men, and its influence on cognitive aging and Alzheimer disease (AD) is largely unknown.

Objective

To determine whether the X chromosome is associated with sex-specific cognitive change and tau pathology in aging and AD.

Design, Setting, Participants

This study examined differential gene expression profiling of the X chromosome from an RNA sequencing data set of the dorsolateral prefrontal cortex obtained from autopsied, elderly individuals enrolled in the Religious Orders Study and Rush Memory and Aging Project joint cohorts. Samples were collected from the cohort study with enrollment from 1994 to 2017. Data were last analyzed in May 2021.

Main Outcomes and Measures

The main analysis examined whether X chromosome gene expression measured by RNA sequencing of the dorsolateral prefrontal cortex was associated with cognitive change during aging and AD, independent of AD pathology and at the transcriptome-wide level in women and men. Whether X chromosome gene expression was associated with neurofibrillary tangle burden, a measure of tau pathology that influences cognition, in women and men was also explored.

Results

Samples for RNA sequencing of the dorsolateral prefrontal cortex were obtained from 508 individuals (mean [SD] age at death, 88.4 [6.6] years; 315 [62.0%] were female; 197 [38.8%] had clinical diagnosis of AD at death; 293 [58.2%] had pathological diagnosis of AD at death) enrolled

in the Religious Orders Study and Rush Memory and Aging Project joint cohorts and were followed up annually for a mean (SD) of 6.3 (3.9) years. X chromosome gene expression (29 genes), adjusted for age at death, education, and AD pathology, was significantly associated with cognitive change at the genome-wide level in women but not men. In the majority of identified X genes (19 genes), increased expression was associated with slower cognitive decline in women. In contrast with cognition, X chromosome gene expression (3 genes), adjusted for age at death and education, was associated with neuropathological tau burden at the genome-wide level in men but not women.

Conclusions and Relevance

In this study, the X chromosome was associated with cognitive trajectories and neuropathological tau burden in aging and AD in a sex-specific manner. This is important because specific X chromosome factors could contribute risk or resilience to biological pathways of aging and AD in women, men, or both.

1.2 Introduction

The X chromosome represents 5% of the genome in women and men and is understudied in aging and Alzheimer disease (AD). In the brain, more genes are expressed from the X chromosome than from any other single autosome [1]; however, analytic challenges posed by X hemizyosity in male individuals, random X inactivation and baseline X escape in female individuals, shared sequences between the X and Y, and limited representation of the X in genome-wide association studies, have largely led to its exclusion in studies [2]. Despite historical constraints, tool kits are expanding, and varied sequencing approaches offer complimentary opportunities to investigate the X with high fidelity in brain aging and neurodegenerative disease. Any advances gained in dedicated study of the X are particularly important given its high density of neural genes, potential contribution to disease-relevant biology within and between each sex, and history of meaningful discoveries in other fields of medicine [3,4]. With this in mind, we investigated an RNA sequencing (RNA-seq) data set from the well-established Religious Orders Study and Rush Memory and Aging Project joint cohorts to measure transcriptional levels of X gene expression in the dorsolateral prefrontal cortex, a cortical hub of multiple cognitive circuits, targeted by aging and AD [5]. Because X gene expression is imbalanced between the sexes, we performed separate analyses of women and men. In our main analysis, we examined whether X expression is associated with cognitive change during aging and AD, independent of AD pathology, in women and men. We also explored whether X expression is associated with neurofibrillary tangle (NFT) burden, a major component of AD pathology linked with cognitive decline in women and men.

1.3 Methods

We performed linear regressions of data derived from RNA-seq of dorsolateral prefrontal cortex with longitudinal change in global cognition and with NFT burden (assessed over 8 regions) in individuals from the Religious Orders Study and Rush Memory and Aging Project joint cohorts. Participants were without known dementia at enrollment (1994-2017) and were followed up longitudinally until death. Institutional review board approval was obtained from Rush University, and all participants provided written informed consent, agreed to brain donation, and signed a repository consent allowing their data to be repurposed. RNA-seq methods have been described in detail [5]. Of 13 822 coding genes detected genome-wide, 488 were from the X chromosome. Neuropathological examination and antemortem clinical and neuropsychological profiling were performed [5,6]. Global cognitive function was derived for each individual from the annual neuropsychological evaluation, comprising 17 different tests that were collapsed to form rates of cognitive decline, controlling for age and years of education [5]. Cognitive decline was regressed against messenger RNA expression and covaried by extent of AD pathology (NFT and neuritic plaque scores); the association was defined as β . Brain NFT burden was regressed against messenger RNA expression, accounting for age and education; the association was defined as β . Additional methods are provided in the eMethods in the Supplement. Significance was established genome-wide at a false discovery rate-adjusted P value of less than .05. Analysis of women was performed separately from men owing to the biologic imbalance of X gene expression between the sexes. Analyses took place in May 2021.

1.4 Results

Demographics of the samples from the Religious Orders Study and Rush Memory and Aging Project joint cohorts that underwent RNA-seq are shown in **Table 1.1** and eTable 1 in the

Supplement, with 508 individuals followed up longitudinally for a mean (SD) of 6.3 (3.9) years. Of these, 315 (62.0%) were female, 197 (38.8%) carried a clinical diagnosis of AD, and 296 (58.2%) carried a pathological diagnosis of AD. Most individuals (499 [98.2%]) self-reported as non-Hispanic White (eMethods in Supplement). Individuals with no cognitive impairment (166 [32.7%]), mild cognitive impairment (124 [24.4%]), clinical AD (173 [34.1%]), mixed mild cognitive impairment (9 [1.8%]), mixed AD (24 [4.7%]), and other dementias (12 [2.4%]) did not differ by sex (eTable 1 in the Supplement).

In women, select X chromosome genes were significantly associated with cognitive change at the genome-wide level (29 genes) (**Figure 1.1 A and B, Table 1.2**, and eTable 2 in the Supplement), adjusted for age, education, and AD pathological burden. Of these, 19 genes (65.5%) showed a positive β score, indicating increased messenger RNA expression associated with slower cognitive decline. In men, X genes were not significantly associated with cognitive change (**Table 1.2** and eTable 3 in the Supplement), despite similar cognitive decline to women ($\beta = -0.18$, $P = .86$; eTable 4 in the Supplement). While lower numbers of men contributed to decreased statistical power, subsampling of women to a male-equivalent-sized cohort continued to show significance of some genes, indicating female specificity of X chromosome–cognition associations. Nonetheless, β scores between women and men showed a strong statistical correlation revealing similar magnitude and direction of X expression with cognitive change between the sexes (**Figure 1.1, C**).

In contrast with cognition, X chromosome gene expression was associated with NFT burden at the genome-wide level (3 genes) in men (**Table 1.2**; eTable 5 in the Supplement) but not women (**Table 1.2** and eTable 6 in the Supplement). This is despite the lower NFT burden in men ($\beta = -0.06$, $P = .07$; eTable 4 in the Supplement) compared with women.

1.5 Discussion

Significant associations of the X chromosome with cognitive change and tau pathology in aging and AD were sex specific. X chromosome gene expression assessed by RNA-seq in the dorsolateral prefrontal cortex was associated with cognitive change in women but not men, independent of AD pathology. In contrast with cognition, X gene expression was associated with neuropathological tau burden in men but not women.

Sex-specific findings of X gene expression in aging and AD were observed at the genome-wide level, including statistical correction for all autosomal and X genes detected. Thus, our results represent strong biological signals comparable with studies reporting autosomal gene associations. Sex stratification likely increased accuracy and resolution of findings because sex-specific biology governs X expression.

For the majority of identified X genes, higher levels were associated with slower cognitive decline in women. Among these, *GRIA3*, *GPRASP2*, and *GRIPAP1* (or *GRASP1*) code for proteins critical to mechanisms of synaptic transmission and plasticity, substrates of cognition. It is possible that women with a higher X dose from baseline escape or reactivation of the silent X showed resilience and better cognitive outcomes, compared with women with a lower X dose. Female-specific X biology, including harboring a second X chromosome, could also contribute to sex differences favoring female individuals [7]. This includes female longevity in AD and female resilience to higher tau burden [7-9]. Of note, more women have AD in large part owing to their longevity with the disease, along with survival to advanced ages when risk and incidence is highest [10]. Causal biological studies of X factors are needed for a deeper understanding for any of these putative roles.

Our observation that X chromosome gene expression, like *UBL4A*, which encodes a protein folding factor, is associated with neuropathological tau burden in men but not women could represent male-specific X biology. Emerging sex differences in tau observed in human populations, along with increased tau-induced gene expression in male mice, support this possibility [8-9,11]. Male-specific X biology includes hemizyosity of the X and maternal X inheritance, sources of genetic and epigenetic sex difference.

The spatial landscape of significant associations with cognitive change revealed transcriptional hot spots of genes clustered proximally on the X chromosome (**Figure 1.1, A**), suggesting common epigenetic regulators. Among these hot spots are genes linked to cognitive preservation and longevity protein families, including *MED12* and *FOXO4*. Whether they could contribute resilience or risk in aging and AD remains unknown.

Recent databases increasingly cover the X chromosome with high fidelity using varied informatic approaches, from expanded genome-wide association studies with X genetic variants to RNA-seq for direct gene expression levels, enabling proper X investigation [12-14]. Two studies using single-cell RNA-seq broadly identified genes linked with AD phenotypes and detected X expression [13,14]. One gene, *MIDI1P1*, also emerged in our study. Its putative role modulating a phosphatase dysregulated in tau biology highlights how a deeper dive into X factors might reveal important pathways.

1.5.1 Limitations

Limitations and caveats of our work include study of predominantly non-Hispanic White individuals within the United States, focus on 1 affected brain region, and lack of cell-type specificity of gene expression changes. It remains to be determined how broadly our findings extend and if X associations could differ with aging vs AD, not separated in this study.

1.6 Conclusions

A disproportionate density of factors influencing neural function reside on the X chromosome and their roles in aging, AD, and other neurodegenerative diseases require identification and investigation in both sexes [15]. This is important because X factors could contribute understanding of disease-relevant neurobiology along with sex differences and sex specificity of biomarkers, disease courses, and eventually pathways for personalized treatments against pathological aging and AD for women and men.

1.7 Figures

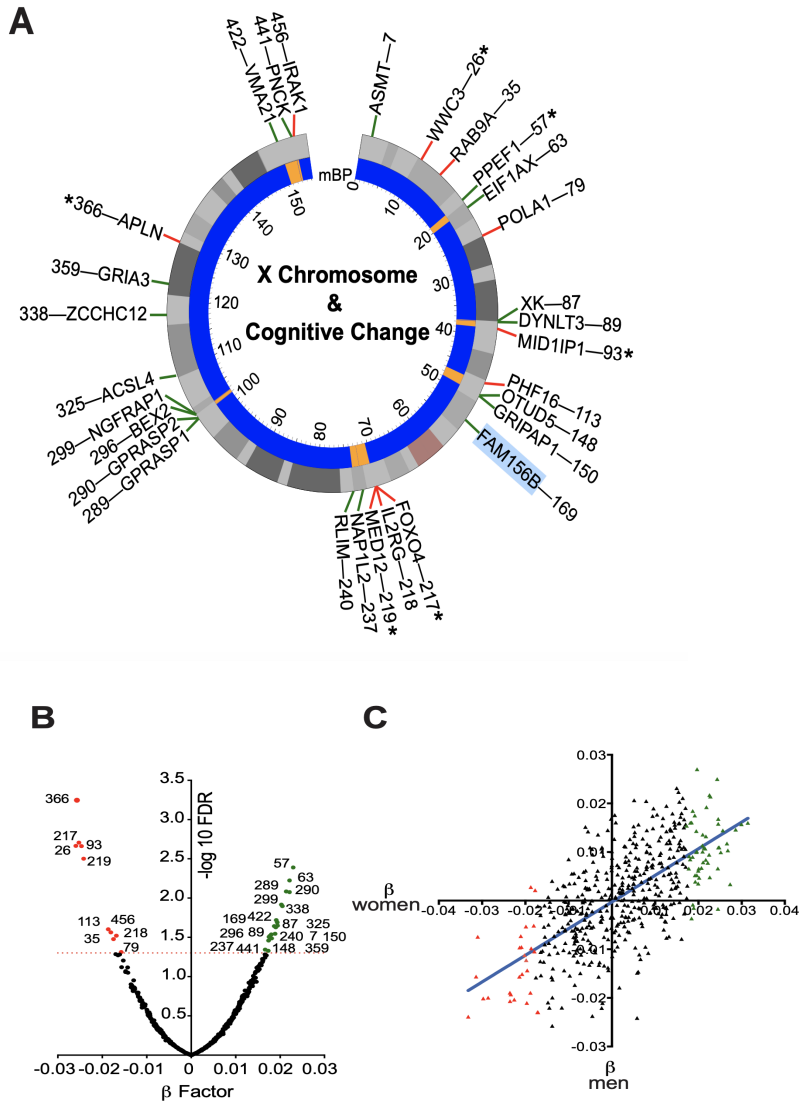


Figure 1.1 Association of X Chromosome Genes with Cognitive Change in Aging and AD
 A, Circos plot of the human X chromosome with coding genes that passed RNA sequencing threshold of significance after genome-wide correction in women, numbered consecutively by location. The inner dark blue band maps clusters of genes in yellow. B, Volcano plot shows β for each gene along with level of statistical significance. Numbering refers to location on X chromosome as depicted in panel A. C, Linear regression analysis of β scores are indicated between men and women ($P < .001$; $R^2 = 0.33$). FDR indicates false discovery rate; mBP, mega base-pairs.

1.8 Tables

Table 1.1 Demographic Information
ROSMAP sample demographics used in RNA seq of cortex

Characteristic	ROS (n = 278)	MAP (n = 230)	ROS/MAP (N = 508)
Age, mean (SEM), y			
At enrollment	78.9 (7.1)	84.3 (5.8)	81.3 (7.0)
At death	87.6 (7.2)	89.3 (5.8)	88.4 (6.6)
Education, mean (SEM), y	18.0 (3.2)	14.7 (2.7)	16.5 (3.5)
Female, No. (%)	172 (61.8)	143 (62.2)	315 (62.0)
Male, No. (%)	106 (38.1)	87 (37.8)	193 (38.0)
Clinical diagnosis of AD at death, No. (%)	107 (38.5)	90 (39.1)	197 (38.8)
Pathological diagnosis of AD, No. (%)	162 (58.2)	134 (58.2)	296 (58.2)

Table 1.2 Gene expression association summary
 Association of X Chromosome Genes Expression with Cognition and Neurofibrillary Tangle Burden in a Sex specific Manner

Characteristic	X-linked gene	Women		Men	
		β	P value	β	P value
Cognition	APLN	-0.0257	<.001 ^b	-0.0141	.45
	FOXO4	-0.0253	.002 ^b	-0.0166	.32
	WWC3	-0.0259	.002 ^b	-0.0039	.86
	MID1IP1	-0.0247	.002 ^b	-0.0144	.37
	MED12	-0.0242	.003 ^b	-0.0143	.37
	PPEF1	0.0230	.004 ^b	0.0157	.35
	EIF1AX	0.0221	.006 ^b	0.0112	.54
	GPRASP1	0.0214	.008 ^b	0.0124	.48
	GPRASP2	0.0221	.008 ^b	0.0131	.41
	NGFRAP1	0.0203	.01 ^b	0.0067	.75
	ZCCHC12	0.0205	.01 ^b	0.0070	.74
	FAM156B	0.0191	.02 ^b	0.0034	.90
	XK	0.0194	.02 ^b	0.0096	.61
	ACSL4	0.0194	.02 ^b	0.0033	.90
	VMA21	0.0187	.02 ^b	0.0206	.21
	RLIM	0.0190	.02 ^b	-0.0009	.98
	PHF16	-0.0187	.03 ^b	-0.0096	.64
	IRAK1	-0.0181	.03 ^b	-0.0100	.60
	GRIPAP1	0.0189	.03 ^b	0.0038	.88
	BEX2	0.0180	.03 ^b	0.0119	.51
	IL2RG	-0.0168	.03 ^b	-0.0105	.62
	ASMTL	0.0177	.03 ^b	0.0157	.36
	NAP1L2	0.0175	.03 ^b	0.0088	.66
	OTUD5	0.0181	.03 ^b	0.0083	.65
	RAB9A	-0.0175	.03 ^b	-0.0141	.41
	DYNLT3	0.0174	.03 ^b	0.0126	.49
	PNCK	0.0166	.045 ^b	0.0058	.80
	GRIA3	0.0173	.047 ^b	0.0098	.60
	POLA1	-0.0157	.049 ^b	-0.0137	.47
	Neurofibrillary tangles	EMD	-0.0325	.57	-0.0982
UBL4A		0.0206	.75	0.0975	.03 ^b
PHF16		0.0343	.57	0.0901	.03 ^b

^a Models for each sex include covariates for age at death and educational attainment.

^b Significant false discovery rate-adjusted P values for genome-wide correction are indicated.

1.9 Supplemental Online Content

All supplemental online content can be downloaded directly from the published material at JAMA Neurology.

1.10 Article Information

Corresponding Author: Dena B. Dubal, MD, PhD, Department of Neurology, University of California, San Francisco, 675 Nelson Rising Ln, 212B, San Francisco, CA 94158 (dena.dubal@ucsf.edu).

Accepted for Publication: June 25, 2021.

Published Online: August 23, 2021. doi:10.1001/jamaneurol.2021.2806

Open Access: This is an open access article distributed under the terms of the CC-BY License. ©

2021 Davis EJ et al. *JAMA Neurology*

1.10.1 Conflict of Interest Disclosures

Dr Bennett reported grants from the National Institutes of Health during the conduct of the study. Dr De Jager reported receiving a sponsored research agreement from Biogen and Roche on a different Alzheimer disease–related theme. Dr Yokoyama reported grants from the National Institute on Aging during the conduct of the study; grants from National Institute on Aging, Rainwater Charitable Foundation, Transposon Therapeutics, Alector, and the US Department of Defense; and other support from the Mary Oakley Foundation and French Foundation outside the submitted work. Dr Dubal reported grants from National Institutes of Health, American Federation for Aging Research, Simons Foundation, and Philanthropy; other support from Unity Biotechnology during the conduct of the study; and had a patent for methods for improving cognition issued filed by the UC Regents and issued. No other disclosures were reported.

1.10.2 Funding/Support

This work is supported by the National Institutes of Health (grants NS092918 and AG068325 [Dr Dubal]; P30AG10161, R01AG15819, R01AG17917 [Dr Bennett]; U01AG61356 [Dr Bennett]; U01AG46152 [Drs De Jager and Bennett]; R01AG36836 [Dr De Jager]; R01AG060393 [Dr Sirota]; R01 AG062588 [Dr Yokoyama]), the Coulter-Weeks Foundation (Dr Dubal), the Simons Foundation (Dr Dubal), Bakar Family Foundation (Dr Dubal), and the American Federation for Aging Research (Dr Dubal).

Role of the Funder/Sponsor: The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

1.10.3 Disclaimer

Dr Dubal is Associate Editor of *JAMA Neurology* but was not involved in any of the decisions regarding review of the manuscript or its acceptance.

1.11 References

1. Nguyen DK, Disteche CM. Dosage compensation of the active X chromosome in mammals. *Nat Genet.* 2006;38(1):47-53. doi:10.1038/ng1705
2. Wise AL, Gyi L, Manolio TA. eXclusion: toward integrating the X chromosome in genome-wide association analyses. *Am J Hum Genet.* 2013;92(5):643-647. doi:10.1016/j.ajhg.2013.03.017
3. Sidorenko J, Kassam I, Kemper KE, et al. The effect of X-linked dosage compensation on complex trait variation. *Nat Commun.* 2019;10(1):3009. doi:10.1038/s41467-019-10598-y
4. Natarajan P, Pampana A, Graham SE, et al; NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium; FinnGen. Chromosome Xq23 is associated with lower atherogenic lipid concentrations and favorable cardiometabolic indices. *Nat Commun.* 2021;12(1):2182. doi:10.1038/s41467-021-22339-1
5. Mostafavi S, Gaiteri C, Sullivan SE, et al. A molecular network of the aging human brain provides insights into the pathology and cognitive decline of Alzheimer's disease. *Nat Neurosci.* 2018;21(6):811-819. doi:10.1038/s41593-018-0154-9
6. Bennett DA, Buchman AS, Boyle PA, Barnes LL, Wilson RS, Schneider JA. Religious Orders Study and rush memory and aging project. *J Alzheimers Dis.* 2018;64(s1):S161-S189. doi:10.3233/JAD-179939
7. Davis EJ, Broestl L, Abdulai-Saiku S, et al. A second X chromosome contributes to resilience in a mouse model of Alzheimer's disease. *Sci Transl Med.* 2020;12(558):eaaz5677. doi:10.1126/scitranslmed.aaz5677
8. Ossenkoppele R, Lyoo CH, Jester-Broms J, et al. Assessment of demographic, genetic, and

imaging variables associated with brain resilience and cognitive resilience to pathological tau in patients with Alzheimer disease. *JAMA Neurol.* 2020;77(5):632-642. doi:10.1001/jamaneurol.2019.5154

9. Dubal DB. Sex Difference in Alzheimer's disease: an updated, balanced and emerging perspective on differing vulnerabilities. In: Lanzberger R, Kranz GS, Savic I, eds. *Sex Differences in Neurology and Psychiatry*. Elsevier; 2020. doi:10.1016/B978-0-444-64123-6.00018-7
10. Shaw C, Hayes-Larson E, Glymour MM, et al. Evaluation of selective survival and sex/gender differences in dementia incidence using a simulation model. *JAMA Netw Open.* 2021;4(3):e211001. doi:10.1001/jamanetworkopen.2021.1001
11. Kodama L, Guzman E, Etchegaray JI, et al. Microglial microRNAs mediate sex-specific responses to tau pathology. *Nat Neurosci.* 2020;23(2):167-171. doi:10.1038/s41593-019-0560-7
12. Smith SM, Douaud G, Chen W, et al. An expanded set of genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nat Neurosci.* 2021;24(5):737-745. doi:10.1038/s41593-021-00826-4
13. Grubman A, Chew G, Ouyang JF, et al. A single-cell atlas of entorhinal cortex from individuals with Alzheimer's disease reveals cell-type-specific gene expression regulation. *Nat Neurosci.* 2019;22(12):2087-2097. doi:10.1038/s41593-019-0539-4
14. Mathys H, Davila-Velderrain J, Peng Z, et al. Single-cell transcriptomic analysis of Alzheimer's disease. *Nature.* 2019;570(7761):332-337. doi:10.1038/s41586-019-1195-2
15. Skuse DH. X-linked genes and mental functioning. *Hum Mol Genet.* 2005;14(Spec No 1):R27-R32. doi:10.1093/hmg/ddi112

Gloss to Chapter 2

One could spend their entire career studying the heterogeneities ADRD. I had the amazing opportunity to dive deeply into two of the most significant contributors: sex, described in chapter 1 and ancestry, described in this chapter. While working on this chapter, Jen connected me with Mike Nalls and his team at the NIH. I had been working on my own genome wide association study meta-analysis when Jen heard Mike's team was working on a similar project and we decided it would be another great opportunity for collaboration. Collaborating with members of Data Tecnica and the Center for Alzheimer's and Related Dementias, including Julie Lake and Hampton Leonard, on this project allowed me to share ideas and learn from leading experts in this field.

Together, we conducted a multi-ancestry meta-analysis followed by fine-mapping, polygenic risk score assessments, and other functional inferences. We attempted numerous methods for fine-mapping and ended up publishing only using the MR - MEGA methods due to inconsistencies between the other methods we attempted like PAINTOR. During this process I learned a lot more about efficient scripting and efficiently implementing functions. Despite conducting the largest multi-ancestry meta-analysis on Alzheimer's to date, we were surprised that we were still unable to achieve a cohort with greater than 10% non-European samples included. This drove us to conduct a systematic review to take a deeper dive into understanding the extent of non-European representation across neurodegenerative disease genome wide association studies.

To conduct the review, we worked closely with the NIH library and discovered how difficult it can be to perform a comprehensive review. Many datasets that came up in our GWAS catalog search did not come up in our Covidence search or in our pubmed search. Despite our efforts to largely automate the review process by using Covidence, the process ended up being very iterative and required a lot of time for us to get our final set of studies. During this process, I learned a lot

about documentation, the importance of data verification, and that many papers are published despite having questionable study design or conclusions. This was an invaluable lesson because when I started graduate school, I had only worked on one other publication and I was working with subject matter experts, so I assumed that if something is published, it means it's rigorous and can be trusted. Now, I understand how to distinguish between reliable and unreliable papers and can determine whether their methods are sensible. This review paper highlighted the importance of representation in NDD research and taught me a lot about proper study design. Additionally, this deep dive into genetic research opened an opportunity for me to be a course instructor of human genetics for UCSF genetic counseling students.

Chapter 2 Paper 1

Multi-ancestry meta-analysis and fine-mapping in Alzheimer's Disease

Contributing Authors: Julie Lake*, Caroline Warly Solsberg*, Jonggeol Jeffrey Kim, Juliana Acosta-Uribe, Mary B. Makarios, Zizheng Li, Kristin Levine, Peter Heutink, Chelsea Alvarado, Dan Vitale, Sarang Kang, Jungsoo Gim, Kun Ho Lee, Stefanie D. Pina-Escudero, Luigi Ferrucci, Andrew B. Singleton, Cornelis Blauwendraat, Mike A. Nalls, Jennifer S. Yokoyama**, and Hampton L. Leonard**

*Equal contribution

**Correspondence

2.1.1 Abstract

Genome-wide association studies (GWAS) of Alzheimer's disease are predominantly carried out in European ancestry individuals despite the known variation in genetic architecture and disease prevalence across global populations. We leveraged published GWAS summary statistics from European, East Asian, and African American populations, and an additional GWAS from a Caribbean Hispanic population using previously reported genotype data to perform the largest multi-ancestry GWAS meta-analysis of Alzheimer's disease and related dementias to date. This method allowed us to identify two independent novel disease-associated loci on chromosome 3. We also leveraged diverse haplotype structures to fine-map nine loci with a posterior probability > 0.8 and globally assessed the heterogeneity of known risk factors across populations. Additionally, we compared the generalizability of multi-ancestry- and single-ancestry-derived polygenic risk scores in a three-way admixed Colombian population. Our findings highlight the importance of multi-ancestry representation in uncovering and understanding putative factors that contribute to risk of Alzheimer's disease and related dementias.

2.1.2 Introduction

Alzheimer’s disease (AD) is a complex genetic disorder with a range of deleterious variants across multiple genes attributed to both early and late-onset forms of sporadic AD [1]. The strongest genetic risk factor for late-onset AD is *APOE-e4*, yet it has been estimated that there may be anywhere from 100 to 11,000 variants that also contribute to risk of late-onset AD [2, 3]. Large-scale genome-wide association studies (GWAS) in European ancestry populations have identified over 75 loci that are associated with AD and related dementias (ADD) [4]. However, genetic research in ADD that focuses solely on European populations limits additional discoveries afforded by studying diverse cohorts. Including non-European populations in genetic research provides new opportunities to uncover ancestry-specific risk variants and loci, increase statistical discovery power, improve fine-mapping resolution to identify putative causal variants, and identify loci with heterogeneous effects across ancestry groups [5–7].

Implementing existing ancestry-aware or heterogeneity penalizing meta-regression approaches have proven powerful at deconvoluting the genetic architecture of other phenotypes across populations [8–18]. We leveraged such techniques, layering existing diverse data on top of more extensive European-derived data, to facilitate discovery of novel ADD risk loci. Here we report the results of our multi-ancestry genome-wide meta-analysis of the largest publicly available ADD GWAS from individuals of European, East Asian, and African American ancestry, and an additional GWAS of Caribbean Hispanic individuals using previously reported genotype data since those summary statistics were not available. Using a meta-regression approach implemented in MR - MEGA, we demonstrate improved fine-mapping at several known ADD loci and estimate the extent to which heterogeneity at these loci is attributable to genetic ancestry. This study highlights the utility of multi-ancestry analyses to further our understanding of disease biology and

reduce health disparities in research by nominating novel loci and characterizing genetic differences across populations.

2.1.2.1 Data included in this study

Our multi-ancestry meta-analysis included a total of 54,233 AD cases, 46,828 proxy AD and related dementia (proxy-ADD) cases, and 543,127 controls (**Figure 2.1.1; Table 2.1.S1**). Detailed information about the existing GWAS summary statistics used in this report are described elsewhere [4, 6, 19]. In brief, the most recent publicly available ADD GWAS includes 39,106 clinically diagnosed AD cases, 46,828 proxy-ADD cases (defined as having a parent with AD/dementia) and 401,577 controls of European ancestry [4]. FinnGen data from Release 6 includes 7,329 AD cases and 131,102 controls free of any neurological disorder. We also included the largest publicly available AD GWAS of African American (2,748 cases and 5,222 controls) [6] and East Asian (3,962 cases and 4,074 controls) [19] populations and an additional GWAS including 1,095 cases and 1,179 controls of Caribbean Hispanic ancestry. Select SNPs from the Gwangju Alzheimer's & Related Dementias (GARD) East Asian cohort (1,119 cases and 1,172 controls) were used to assess East Asian risk at our novel loci post-hoc since these SNPs were not present in the discovery East Asian dataset from Shigemizu et al. used in our meta-GWAS [20]. In this study, we considered significant variants as passing the standard p-value threshold of 5×10^{-8} , consistent with most GWAS meta-analyses and used previously in other multi-ancestry studies [21–23]. Our analysis included only variants that passed quality control and with a minor allele frequency $> 1\%$ in a minimum of three datasets to accurately quantify heterogeneity, effectively reducing the number of potential haplotypes and tests.

2.1.2.2 Meta-GWAS

Association summary statistics from all five datasets, representing four super populations, were aggregated via fixed and random effects models implemented in PLINK v1.9 [24] and a multi-ancestry meta-regression implemented in MR - MEGA [25] (**Table 2.1.S1**). A fixed effect analysis was conducted in conjunction with random effects in PLINK as it is the standard choice for many GWAS meta-analyses. Since the fixed effect analysis did not identify any additional novel loci, we focused on the random effects and MR - MEGA results as these methods are generally more appropriate for multi-ancestry studies. In particular, MR - MEGA was specifically designed for multi-ancestry meta-analyses and random effects models penalize heterogeneity in their construction of effect estimates, allowing these estimates to be more generalizable across global populations. We did not observe any genomic inflation in these analyses after excluding rare variants ($MAF < 1\%$ per study) and correcting for case-control imbalance (**Table S1; Figure 2.1.S1**). Chromosome 19 was also excluded from genomic inflation estimates to avoid bias from the *APOE* region. Association results from the random effects and MR - MEGA meta-analyses were moderately concordant for SNPs without heterogeneity ($I^2=0$, $R^2=0.6$). Our study also demonstrated that MR - MEGA is advantageous for SNPs with heterogeneous allelic effects (**Figure 2.1.S2**). Results from all meta-analyses (MR - MEGA, random effects, and fixed effect) can be found in **Tables S2a-c**.

A total of 68 loci reached genome-wide significance ($P < 5 \times 10^{-8}$) in the fixed effect, random effects or MR - MEGA meta-analyses. While most of these loci overlapped all analyses, 1 was only significant using MR - MEGA (*JAZF1*), 1 using random effects (*KANSL1*), and 5 using the fixed effect model (*ADAMTS1*, *MAF*, *PLEKHAI*, *TSPOAP1*, and *UMADI*) (**Tables 2.1.S2a-c**). 66 of these loci overlapped previously established genomic regions associated with ADD (see

“Defining associated loci” in the **Online Methods**). Our analysis may have been underpowered to detect the remaining loci without the replication summary statistics from Bellenguez et. al. as only the discovery phase statistics from that study were available.

We additionally identified two independent ADD risk loci on chromosome 3 near *TRANK1* (rs9867455; $P_{RE}=3.49 \times 10^{-8}$, $\beta_{RE}=-0.0424$, $I^2=0$) and *VWA5B2* (rs9837978; $P_{RE}=3.75 \times 10^{-8}$, $\beta_{RE}=-0.0526$, $I^2=0$) that are outside of the maximal linkage disequilibrium (LD) boundary for any known AD risk loci (**Figure 2.1.2**). These two loci were identified using both the fixed and random effects models. These loci also showed similar P-values using MR - MEGA but did not reach genome-wide significance using this method (**Table 2.1.1**). The association signals are primarily driven by the European-focused study by Bellenguez et al., where they were also sub-significant (P-values of 6.95×10^{-7} for *TRANK1* and 1.17×10^{-6} for *VWA5B2*, respectively). We did not identify any additional novel loci using MR - MEGA (**Figure 2.1.S12a**).

Since the lead SNPs for these potential novel loci were absent in the East Asian dataset from Shigemizu et al. used for initial discovery [19], we attempted to test these SNPs in an independent East Asian ancestry dataset [20]. Genome-wide data for this cohort was not available to include in the meta-analysis at the time of publication. We observed an association at $P < 0.05$ at *VWA5B2*-rs9837978 ($P=0.048$, $\beta=0.204$) in the GARD cohort (n=2,291). The direction of effect was not consistent with the direction seen in the other populations included in the discovery GWAS, suggesting the effect of this locus may be heterogeneous across populations, but more extensive testing in East Asian populations must be performed for confirmation. We were unable to test the association at *TRANK1*-rs9867455 in the GARD replication cohort since this SNP was not included in their GWAS and an LD proxy SNP was not available.

2.1.2.3 Gene prioritization for novel loci

Using public expression quantitative trait locus (eQTL) evidence from Open Targets [26] and multi-ancestry brain eQTL summary data [27], we assessed whether *TRANK1*-rs9867455 and *VWA5B2*-rs9837978 are associated with the expression of nearby genes. Open Targets reported rs9867455 as a significant eQTL ($P < 1 \times 10^{-6}$) for *LRRFIP2*, *ITGA9*, *GOLGA4*, *MLH1*, and *TRANK1* across blood or other tissues. *LRRFIP2*, *GOLGA4* and *TRANK1* were also nominated in the multi-ancestry brain eQTL data. Open Targets reported rs9837978 as a significant eQTL for *AP2M1*, *ABCF3*, *VWA5B2*, *ALG3*, *ABCC5*, *DVL3*, and *CLCN2*. *AP2M1*, as well as two additional genes (*EIF2B5* and *ECE2*) were nominated in the multi-ancestry brain eQTL data.

To prioritize susceptibility genes with expression effects on ADD risk, we performed summary-based Mendelian Randomization (SMR) to infer whether expression of the eQTL-nominated genes is causal for AD. More details regarding the purpose and methods used to perform SMR can be found in the “Functional inferences” section of the Online Methods. At the *TRANK1*-rs9867455 locus, *TRANK1*, *LRRFIP2*, *GOLGA4*, and *ITGA9* were significant in our SMR results for affecting AD risk via expression across multiple tissue types. The strongest associations in cortex tissue were seen with *TRANK1* and *LRRFIP2*. The GWAS signal at the *TRANK1*-rs9867455 locus colocalized most strongly with *TRANK1* expression in cortex tissue ($R^2=0.52$; **Figure 2.1.S3**). At the *VWA5B2*-rs9837978 locus, *VWA5B2*, *AP2M1*, *ABCF3*, *ALG3*, *EIF2B5*, *DVL3*, *CLCN2*, *ABCC5* were significant in our SMR results. The strongest associations in brain tissues were seen in *ABCF3*, *ALG3* and *EIF2B5*, although colocalization between the GWAS signal and these eQTLs were not very strong ($R^2 < 0.5$; **Figure 2.1.S4**). For more details on directionality of these associations, see **Table S3**.

2.1.2.4 Fine-mapping

A total of nine loci outside of the *APOE*, *MAPT*, and major histocompatibility complex (MHC) regions were fine-mapped to a credible set of ≤ 2 SNPs with a combined posterior probability (PP) of 99% (**Table 2.1.2**; **Figures 2.1.S5-6**). The *MHC* and *MAPT* regions were excluded from fine-mapping due to a complex haplotype structure across populations [28, 29] and known haplotype inversions [30], respectively. Five of these loci were previously fine-mapped with $PP > 0.8$ in large GWAS of European populations (**Figure 2.1.S5**; *BINI*-rs6733839; *INPP5D*-rs10933431; *ECHDC3*-rs7912495; *APHIB*-rs117618017; *ABCA7*-rs12151021) [31, 32]. Four additional ADD loci with 1-2 variants in their 99% credible sets have not been previously fine-mapped (**Figure 2.1.S6a-b,d-e**; *RHOH*-rs2245466; *CTSB*-rs1065712; *FAM157C/PRDM7*-rs56407236; *GRN*-rs5848). Interestingly, *GRN* is also a candidate risk locus for frontotemporal dementia (FTD) [33] and Parkinson's disease (PD) [34, 35]. More than 70 pathogenic variants in *GRN* have been linked to familial FTD [33]. While *GRN* mutations do exist in sporadic forms of FTD, we were unable to find publicly available FTD GWAS summary statistics with a significant association at this locus. Using publicly available PD GWAS summary statistics [34], we compared the regional genetic correlations between PD and ADD at both *GRN* and *CTSB*, another candidate risk locus for PD. LocusCompare plots show low genetic correlation between ADD and PD at both loci (*GRN*, $R^2=0.36$; *CTSB*, $R^2=0.0048$), which may indicate that distinct causal variants drive the associations (**Figure 2.1.S7**). One additional locus with a credible set size > 2 was fine-mapped to a single SNP with $PP > 0.8$ (**Figure 2.1.S6c**; *SLC24A4*-rs12590654, $PP=0.91$, $n=32$ in 99% credible set). In addition, two SNPs with a $PP \geq 0.3$ were annotated as missense variants (**Figure 2.1.S8**; *MS4A6A*-rs7232, $PP=0.54$, $n=4$ in 99% credible set; *SHARPIN*-rs34674752, $PP=0.30$, $n=5$ in 99% credible set). Notably, our fine-mapping analysis did not replicate *SORL1*-

rs11218343, which has been previously fine-mapped with a PP > 0.999 in two large European studies [31, 32], likely due to a different regional architecture in the East Asian population as has been previously reported (**Figure 2.1.S9**) [36]. All 99% credible sets are provided in **Table 2.1.S4**.

2.1.2.5 Heterogeneity analysis

We observed significant heterogeneity ($I^2 > 30\%$) at 19 of the 48 loci that reached genome-wide significance ($P < 5 \times 10^{-8}$) in MR - MEGA (**Table 2.1.S2a**), which does not include the additional loci (including 2 novel) identified by the fixed and random effects analyses (**Table 2.1.S2b-c**). Several factors can account for the observed heterogeneity, such as differences in study design, geographical region, and diagnostic accuracy. We estimated the proportion of heterogeneity that is attributable to genetic ancestry using MR - MEGA (see “Assessment of allelic effect heterogeneity” in the Online Methods) and observed that at least 50% of the heterogeneity was attributable to genetic ancestry at 10 of these loci (**Figure 2.1.3, Figure 2.1.S10**). We also assessed heterogeneity at lead SNPs from the most recent European GWAS [4] and found that 37% of the lead SNPs tested presented significant heterogeneity ($I^2 > 30\%$), of which 48% were primarily attributable to ancestry (**Table 2.1.S5**). Five of the fine-mapped SNPs also showed significant heterogeneity ($I^2 > 30\%$), of which only *SLC24A4* showed heterogeneity that was primarily attributable to genetic ancestry (**Table 2.1.2**).

The genetic polymorphisms rs7412 and rs429358 that form the *APOE* e2/e3/e4 alleles presented very different allelic heterogeneity. Consistent with previous studies [7, 37], we observed an attenuated signal at *APOE*-rs429358, which determines the *APOE*-e4 allele, among the cohorts of African Americans and Caribbean Hispanics (**Figure 2.1.S11a**). *APOE*-rs429358 had the highest heterogeneity ($I^2=96.54$) of the SNPs tested with ~42% attributable to genetic ancestry, although these polymorphisms were not available to test in the most recent European-ancestry AD GWAS

[4]. In contrast, *APOE*-rs7412, which determines the *APOE*-e2 allele, did not present any heterogeneity (**Figure 2.1.S11b**; $I^2=0$). Complementary to standard GWAS association tests, we also generated P-values representing heterogeneity of effect estimates attributable to genetic ancestry in the multi-ancestry meta-regression and observed the strongest signal near *APOE* (**Figure 2.1.S12**). Additionally, we observed strong evidence of ancestry-related heterogeneity ($P_{\text{HET}} < 1e-6$) near *SORL1*, as well as *PAPOLG*, *AC026202.5*, and *snoU13* which did not meet genome-wide significance in the association results. LocusZoom and beta-beta plots of these loci suggest that non-European populations primarily drive these association signals, and there are likely discordant effects across populations (**Figure 2.1.S13**).

2.1.2.6 Polygenic risk scoring

We tested the performance of the multi-ancestry fixed and random effects models and each of the GWAS from single ancestral populations in a Colombian cohort of AD cases (n=281) and neurologically normal controls (n=87). This cohort is an admixture of three ancestral populations, with European substructure making up the highest proportion of global ancestry (mean of 64%, SD=15%), followed by Indigenous American (mean of 27%, SD=11%), and African (mean of 9%, SD=11%). Colombian samples were used to test PRS applicability as they were a population not represented in the meta-analyses. While the Caribbean Hispanic cohort included in the meta-analyses is also three-way admixed, this cohort likely has a lower Indigenous American proportion and higher African ancestry proportion than the Colombian cohort [38]. Single-ancestry PRS performed worse than multi-ancestry random-effects-derived PRS in terms of area under a receiver operating characteristic curve (AUC). We observed maximal AUCs of 79% and 68% including and then excluding *APOE* variants in this population, and 75% and 63% for the European-derived PRS (**Figure 2.1.4**; **Table S6**). Non-European AUCs tended to improve with increasing sample

size (**Figure 2.1.4**), suggesting that the composite score, combining ancestry-specific PRS by population weights, may have performed as well or better than the random-effects-derived PRS if the component GWASs from under-represented populations were better powered.

2.1.3 Discussion

We performed a large, genome-wide meta-analysis of ADD across five datasets, representing four super-populations. By leveraging data from multiple ancestry groups, we replicated 66 known ADD loci and identified two novel risk loci on chromosome 3 (**Tables S2a-c**). These novel loci reached genome-wide significance using the fixed and random effects models but were sub-significant using MR - MEGA as this model is not as well-powered to detect loci with homogenous allelic effects [25].

The first novel locus identified in this study is near *TRANK1*, which encodes tetratricopeptide repeat and ankyrin repeat containing 1. *TRANK1* is associated with DNA- and ATP-binding and DNA repair and is highly expressed in brain tissue [39, 40]. Previously, *TRANK1* has been cited as a robust risk locus for both schizophrenia (SZ) [41,42,43] and bipolar disorder (BD) [39, 44, 45], although subtype analyses suggest that this signal is primarily driven by the most heritable subtype, BD I, which is genetically correlated with SZ [46]. Notably, BD has also been shown to increase risk for AD, with the two sharing significant genetic overlap [47]. LocusCompare plots show a modest correlation between our random effects meta-analysis of ADD with BD I ($R^2=0.46$) and SZ ($R^2=0.23$) GWAS at this locus (**Figure 2.1.S14-15**). We also observed moderate LD ($R^2=0.44$, 1000 Genomes EUR) between the lead SNP identified in our study (rs9867455) and the lead SNP identified in the BD I GWAS (rs9834970), indicating potential cross-disease overlap for this locus.

Previous SMR analysis of the *TRANK1* region in BD suggested that both *TRANK1* and *GOLGA4* may be susceptibility genes [48]. *TRANK1* expression is decreased in both BD and AD, and decreased expression of *TRANK1* was found to alter the expression of genes related to neuronal development and differentiation [47]. Altered neurogenesis has been implicated in human AD brains and AD rodent models [48]. Previous studies have suggested that *TRANK1* may be involved in blood brain barrier permeability changes and neuroinflammation, both of which may be relevant to neurodegeneration [48, 49]. Interestingly, Kunkle et al. nominated *TRANK1* through a gene-based analysis conducted in an African American population, but not through the single SNP association testing included in this study [6]. A combination of eQTL and SMR nominated *TRANK1*, *LRRFIP2*, *GOLGA4*, and *ITGA9* as potential genes underlying this SNP association, with the strongest associations in cortex tissue seen with *TRANK1* and *LRRFIP2* (**Table S3**). *LRRFIP2* encodes LRR binding FLII interacting protein 2, which regulates Toll-like receptor 4 (TLR4) and can downregulate the *NLRP3* inflammasome. TLR4 can induce microglial amyloid- β clearance in the brain in early stages of AD but can later induce an inflammatory response, suggesting that disruptions to *LRRFIP2* may affect AD pathology in patients [50].

The second locus is nearest to *VWA5B2*, which encodes von Willebrand factor A domain-containing protein 5B1. Von Willebrand factor (VWF) is a glycoprotein that facilitates blood clotting at areas of injury. High VWF is associated with short-term risk of dementia, possibly due to the increased risk of blood clots restricting blood flow in the brain [51]. Interestingly, *VWA5B2* was found to be downregulated in AD patients, and other variants in *VWA5B2* have been linked to decreased mean hemoglobin concentration [12, 52]. Low hemoglobin, or anemia, has been linked to an increase in risk for AD [53]. This information seems to further implicate the involvement of the vasculature system with AD, complementing previous studies such as those investigating

traumatic brain injury [54]. Whether *VWA5B2* has biological implications on risk for AD needs to be further investigated. The lead variant rs9837978 does not lie within any of the nearby genes at this locus, but eQTL and SMR evidence for this variant nominated eight nearby genes including *VWA5B2* (**Table S3**). The strongest SMR association in brain tissue is seen with *ABCF3*. The *ABCF3* protein is a member of the ATP-binding cassette (ABC) superfamily, all of which transport a variety of substrates across intra- and extracellular barriers [55]. Members of the ABC A subfamily, such as *ABCA7* and *ABCA1*, have previously been nominated as AD risk genes [4]. *ABCF3* is a unique family member in that it lacks a transmembrane domain but has been nominated as a candidate of TLR signaling, similar to *LRRFIP2* [56]. In addition to inducing inflammatory responses, TLRs can affect microglial activity, synaptic plasticity, and tau phosphorylation, providing additional evidence to their potential importance in AD pathology [57]. Additionally, downregulation of *ABCF3* has been associated with an increase in viral load after infection by a flavivirus, specifically the West Nile virus which has been linked to long-term neurological problems and dementia [58, 59]. It's plausible that in the presence of viral infection, changes in *ABCF3* expression may affect immune response and inflammation, two processes that play a role in the amyloid cascade hypothesis [60].

Future studies will be required to further disentangle the potential roles of the nominated genes in the context of ADD risk. Although genome-wide summary statistics were unavailable for replication, we attempted to replicate the lead SNP *VWA5B2*-rs9837978 in a small East Asian cohort [61]. However, this SNP demonstrated an opposite direction of effect and the replication dataset was underpowered to detect an association (**Figure 2.1.S16**). Assuming a disease prevalence of 2%, MAF of 0.1642 (gnomAD v3.2.1 East Asian), and a nominal significance threshold of 0.05, we had ~80% power to detect genotype relative risks ≥ 1.245 but our odds ratio

for this allele, which is generally an overestimate of risk [62], was 1.05. Replication in larger and more diverse cohorts is warranted in future studies. Further, the disparity seen at points between the results on Open Targets, which consists of largely European data, and the multi-ancestry eQTL results for nominated genes also highlights the need for more multi-modal reference data including diverse ancestries. However, it is also possible that there could be different mechanisms underlying disease risk conferred by the implicated loci across different populations.

Our study highlights the utility of multi-ancestry datasets at uncovering putative mechanisms that contribute to ADD. Fine-mapping at several known ADD loci was better resolved using the multi-ancestry meta-regression compared to previous efforts in European populations. For example, fine-mapping near *RHOH*, *CTSB* and *FAM157C/PRDM7* nominated variants that are located in untranslated regions that were not well-resolved in European-focused studies. Variants in the 3'UTR region can impact translation or protein stability, and transcription binding can be impacted by variants in the 5'UTR region. Additionally, *GRN*-rs5848 is associated with circulating progranulin levels and decreased *GRN* expression has been implicated in several neurodegenerative diseases, including AD and FTD [35, 63, 64]. In contrast to previous studies in European populations, the *SORL1* locus was not resolved to a single putative causal SNP. Lead SNPs in both the European (*SORL1*-rs11218343) and East Asian (*SORL1*-rs117807585) GWAS are more common among East Asians compared to all other populations in the Genome Aggregation Database v2.1.1 (rs11218343: $AF_{EAS}=0.30$, $AF_{EUR}=0.039$; rs117807585: $AF_{EAS}=0.22$, $AF_{EUR}=0.020$). It is possible that alternative fine-mapping approaches that allow for multiple causal variants per locus will provide greater insight into the *SORL1* locus.

At the *MS4A* gene cluster, multi-ancestry fine-mapping resolved the signal to a credible set of five variants, with a common missense variant (rs7232, PP=0.54) and an intergenic variant nearest

MS4A4A (rs1582763, PP=0.45) that are in moderate LD ($R^2=0.55$, 1000 Genomes all populations) having the highest probability of causality (**Figure 2.1.S8**). *MS4A4A* and/or *MS4A6A* modulate soluble TREM2 (sTREM2) in cerebrospinal fluid (CSF), which is correlated with AD progression. Previous studies have shown that rs7232 is associated with *MS4A6A* gene expression and CSF sTREM2 [65, 66], while rs1582763 is a cis-eQTL for *MS4A4A* and *MS4A6A* [67]. Conditional analysis of CSF sTREM2 levels in this region have pointed to two independent signals represented by rs1582763 and rs6591561 (*MS4A4A* p.M159V) [67]. Therefore, a fine-mapping approach that allows for multiple causal variants may be more appropriate for this region.

In addition to highlighting genetic risk factors that are shared across populations, our results also highlight ADD loci with significant heterogeneity that may reflect variation in effect sizes, allele frequencies or interaction(s) with environmental risk factors that vary by ancestral group. For example, we observed the strongest evidence of heterogeneity at *APOE*-rs429358. Around 42% of the heterogeneity at this allele was attributable to genetic ancestry, while the remaining heterogeneity may reflect other sources of variation such as imputation accuracy since this allele is rarely assayed successfully on genotyping arrays. At *JAZF1*-rs67250450 and *CLU*-rs1532276, we observed the strongest evidence of ancestry-related heterogeneity, both of which are most common among individuals of East Asian ancestry and showed the strongest effects in this population (**Figure 2.1.3**). We also observed significant ancestry-related heterogeneity at *SORL1* and *TREM2*, which have been previously shown to harbor population-specific risk variants [68, 69]. Given that our analyses focused on common variation, the effects of rare heterogeneous variants (e.g. *ABCA7*-rs115550680, which has comparable effects to *APOE*-rs429358 among African Americans [70]) may not have been fully captured.

While this study marks progress towards assessing genetic risk of ADD across multiple populations, we acknowledge several limitations. First, we recognize that the magnitude of clinical and pathological diversity among ADD cases is extensive. The diagnostic inaccuracy rate of AD likely differs across studies and populations [71, 72] both early-onset forms of AD and other pathologies such as frontotemporal and Lewy body dementias. This clinicopathological heterogeneity is further exacerbated by the inclusion of proxy-ADD cases in the Bellenguez et al. study, a study which comprised the majority of European samples included in our analysis. While proxy-ADD cases may introduce more variability than a clinical or pathological diagnosis of AD, prior studies have demonstrated strong genetic correlation between proxy-ADD and AD ($r_g=0.81$), further supporting the use of this data [73]. Additionally, despite the phenotypic heterogeneity in the Bellenguez et al. study, the utility of ADD GWAS are supported by genes with well-defined involvement in relevant molecular pathways. For example, the *APP* locus was first identified at genome-wide significance in Bellenguez et al. ($P=1.02 \times 10^{-9}$) and we found a stronger level of significance in our random effects meta-analysis ($P=8.1 \times 10^{-12}$, $I^2=0.79$). This locus is likely driven by underlying AD pathology due to its role in the formation of amyloid- β and its previous implications in both LOAD and EOAD [74, 75]. Although disease subtype-specific conclusions that can be made from our analysis are limited by the diagnostic criteria of the included GWAS, similar analyses can be applied as larger datasets with high phenotypic specificity become available.

Additionally, although MR - MEGA is a useful tool for fine-mapping and ancestral heterogeneity estimation, the software requirements of population overlap ($K > 3$) often result in reduced variant sets after study level quality control. This can bias fine-mapping results as we reduce the potential resolution on local haplotypes, and usually necessitates the inclusion of at least one of the larger

European-focused studies. In our case, previous European and Finnish studies served as the backbone of our meta-GWAS. We did not replicate previous fine-mapping at *NCK2*, *TREM2* and *RNF223* from European-focused studies since study level quality control included filtering for common (MAF > 1%), biallelic SNVs due to potentially poor imputation and general low power for rare variants across ancestral groups. We acknowledge that variants with a minor allele frequency < 1% in one or more populations, as well as indels and structural variants, may contribute to the observed associations. While less stringent QC may have allowed us to detect more variants, we used a more conservative approach to accommodate the smaller sample size of non-European GWAS and the need for MR - MEGA SNPs to be present in at least four studies. Future work should include imputation using diverse reference panels from long read sequence data specific to ADD to improve genomic coverage and provide insights into structural variation that may be population specific.

In addition, the number of axes of genetic variation (T) in MR - MEGA is restricted to $T \leq K-2$, where K is the number of input GWAS. The East Asian GWAS from Shigemizu et al. used in our meta-GWAS tested less than half as many SNPs as the others (**Table S1**), limiting the meta-regression to a single axis of genetic variation (PC0) at SNPs that overlap the remaining GWAS (K=4). Including a larger number of input GWAS from underrepresented populations will likely improve the heterogeneity estimates outlined in this study and are worth pursuing when such data become available.

While our study is inclusive, due to data availability and the European-dominated nature of genetic research, European-ancestry individuals make up approximately 85% of cases and the discovery efforts here maintain a baseline level of Eurocentric bias. Despite this bias, our random-effects-derived PRS including *APOE* variation achieved a higher maximal AUC of 79% in an independent

admixed Colombian cohort compared to 75% achieved by the European-based PRS. Additionally, while our novel method of creating a composite PRS model that leverages admixture percentages is a potentially promising approach for assessing ADD risk across ancestrally heterogeneous and/or admixed cohorts, its performance relies on sufficient sample sizes and global genetic representation. As larger scale GWAS for multiple continental “super populations” continue to become available, we believe this method of tuning PRS to an individual’s genetic admixture could have utility in a precision medicine context. Reducing the Eurocentric bias in AD genetics research will require the harmonization and refining of diagnosis in non-European research sites that serve communities with unique cultural and logistic concerns for participation in research. Overall, our study provides a critical framework for future ADD meta-analyses. It is our hope to improve representation in ADD genetic studies in the future, increasing the balance between European and well-powered non-European cohorts.

2.1.4 Online Methods

2.1.4.1 Existing GWAS studies

Summary statistics from Bellenguez et al. 2022 were accessed through the National Human Genome Research Institute-European Bioinformatics Institute GWAS catalog under accession number GCST90027158 in May 2022. Summary statistics from FinnGen Release 6 were accessed online in April of 2022. Summary statistics from Kunkle et al. 2021 were accessed through NIAGADS under accession number NG00100 in April of 2022. Summary statistics from Shigemizu et al. 2021 were accessed through the National Bioscience Database Center (NBDC) at the Japan Science and Technology Agency (JST) through accession number hum0237.v1.gwas.v1 in April of 2022. All summary statistics were aligned to GRCh37 and cleaned to remove indels, multi-allelics and rare variants ($MAF < 1\%$) prior to multi-ancestry analysis.

2.1.4.2 Caribbean Hispanic GWAS

Data from the Columbia University Study of Caribbean Hispanics and Late Onset Alzheimer's disease were accessed via application to dbGaP accession number phs000496.v1.p1 in April of 2022. Samples were filtered to keep unrelated individuals without missing values for AD affection status, age, study category, education, and a missing call rate < 0.02 . Principal component analysis was performed on a combined dataset of study subjects and HapMap was used as a reference to identify potential outliers. Controls with a family history of dementia were removed to ensure that potential proxy-ADD cases were not present in the control group. Variant QC included exclusion filters for monomorphic SNPs, variants with $MAF < 1\%$, missingness rates $> 2\%$, sex differences in allelic frequency ≥ 0.2 and heterozygosity > 0.3 , duplicate SNPs, Hardy–Weinberg Equilibrium (HWE) P -value $< 1 \times 10^{-4}$, and > 1 discordant calls or Mendelian errors. All variants with a

significant frequency mismatch ($\chi^2 > 300$) with the TOPMed reference panel were removed prior to imputation.

As increasing age is the most significant risk factor for ADD, age-matching is commonly used to control for differences in age distributions between cases and controls. While we did not perform case-control age matching to maximize sample size, the distribution of ages between cases and controls largely overlaps and we include age as a covariate in our GWAS analysis. A demographic summary table detailing age, *APOE*-e4 status, sex for the Caribbean Hispanic cohort is provided in **Table S7**.

Using PLINK v1.9 [24], we evaluated the association between AD and imputed genotypes via logistic regression on allele dosages with imputation quality > 0.3 , adjusting for sex, age (age at disease onset for cases, age at last evaluation for controls), education, study category, and the first 10 principal components (PCs). Study category denotes subcategories within the Caribbean Hispanic dataset (individuals are from the United States, Puerto Rico and the Dominican Republic) and is included to account for potential batch effects.

2.1.4.3 Meta-analysis and fine-mapping

Meta-analysis

Three models were used to conduct multi-ancestry meta-analyses. Fixed effect and random effects models were performed using PLINK v1.9, while a separate analysis was performed using MR - MEGA v0.2 [25]. PLINK v1.9 was preferred over METAL due to its capacity to perform fixed and random effects analyses in parallel. A random effects model provides a more conservative framework which allows each study to have unique effects, as can be expected in different populations. MR - MEGA was also employed since it is well-powered to detect associations at loci with allelic heterogeneity. MR - MEGA models allelic effects as a function of axes of genetic variation that are derived from the input GWAS summary statistics. This method can result in reduced variant sets since it requires that variants have sufficient overlap between the input datasets ($K > 3$), where K is the number of input GWAS, in contrast to both the fixed and random effects models implemented in PLINK v1.9 which were limited to $K > 2$ to accurately quantify heterogeneity.

The European and Finnish European GWAS were included separately in all multi-ancestry meta-analyses to account for finer-scale differences in allele frequencies. To determine the optimal number of PCs needed to distinguish cis- and multi-ancestry ADD summary statistics using MR - MEGA, we visually inspected pairwise PC plots generated using all five GWAS referenced in **Table S1**. We observed adequate separation between the Caribbean Hispanic, European, African American, and East Asian GWAS using the first two meta-regression PCs (**Figure 2.1.S17**). To increase the variant set, we also ran MR - MEGA separately for each combination of four input GWAS. A single axis of genetic variation ($T=1$) was included in this

analysis since this is the maximum allowable given the constraints of the model ($T \leq K-2$). Summary statistics were aggregated to maximize the effective sample size for each variant.

Defining associated loci

FUMA was used to find maximal LD blocks around loci that reached $P < 5 \times 10^{-8}$ in the specified multi-ancestry meta-analysis. LD blocks of independent significant SNPs ($R^2 > 0.3$, 1000 Genomes all populations) were merged into a single genomic locus if the distance between LD blocks was less than 250kb. These loci were compared to the previous GWAS by Bellenguez et al. [4] and Open Targets to assess whether these regions were known to be associated with ADD. These genomic intervals were also used as inputs for fine-mapping as described below.

Fine-mapping

Fine-mapping was performed using approximate Bayes' factors in favor of association from the meta-regression model implemented in MR - MEGA. Posterior probabilities (PP) were calculated using single-SNP Bayes factors and credible sets were generated for each locus (with genomic intervals defined as described above) until the cumulative PP exceeded 99%. All SNPs in the 99% credible sets were annotated with using default criteria to select one block of annotation per variant (**Table S4**).

2.1.4.4 Assessment of allelic effect heterogeneity

Allelic effect heterogeneity between studies was assessed for all lead SNPs reaching genome-wide significance ($P < 5 \times 10^{-8}$) in the meta-regression, implemented in MR - MEGA. The meta-regression model derives axes of genetic variation from pairwise allele frequency differences between the input GWAS. Heterogeneity is then partitioned into (1) ancestry-related heterogeneity that is correlated with the axes of genetic variation and (2) residual heterogeneity that is likely due to other factors such as diagnostic accuracy, study design (e.g. covariate adjustments, phenotype

definition, imputation quality, inclusion of proxy-ADD cases) and/or geographical region. Total heterogeneity at each index SNP was quantified using the I^2 statistic in PLINK v1.9 to avoid bias due to sample size for SNPs not tested in the large European studies. The I^2 statistic describes the proportion of variation in effect estimates that is due to heterogeneity. We considered SNPs with an $I^2 > 30\%$ as having significant heterogeneity since this suggests at least moderate variation in allelic effects [76]. The percentage of this heterogeneity that is attributable to genetic ancestry was then calculated using Cochran's Q statistics for ancestral and residual heterogeneity from the meta-regression (equation 1; ANC: ancestry, RESID: residual).

$$\% \text{ Heterogeneity}_{ANC} = Q_{ANC} / (Q_{RESID} + Q_{ANC}) * 100\% \quad (1)$$

2.1.4.5 Functional inferences

To prioritize genes underlying the two novel loci, we first looked at public eQTL data to determine whether the GWAS-identified lead variants are eQTLs for nearby genes. This allowed us to cast a wide net of potential regional genes of interest. We employed Open Targets for this effort, which shares eQTL results for variants from blood, brain, and a wide array of tissues from multiple public eQTL datasets [26]. We additionally investigated a multi-ancestry brain eQTL dataset [27] which was not available on Open Targets at the time of publication. We considered the lead variants as significant eQTLs for a gene if they passed the significance threshold of $P < 1 \times 10^{-6}$, which has been shown to correspond to a genome-wide false discovery rate (FDR) of 5%, although we do acknowledge this may be overly conservative in our regional analyses [77].

Once we had nominated potential genes for which our lead variants were significant eQTLs, we used summary-based Mendelian Randomization (SMR) to make functional inferences as to whether the disease risk SNPs in these regions mediate gene expression. We integrated summary-level data from the most recent ADD GWAS [4] with data from multiple eQTL studies in different

tissues using the SMR method [78]. SMR uses summary statistics to determine if an exposure is associated with a trait through a shared causal variant. MR can be used to mimic a randomized controlled trial, as having a variant that increases or decreases expression of a gene may be comparable to life-long treatment with a drug targeting the encoded protein of that gene [79]. For example, if SNP A affects gene B expression (the exposure), and SNP A is also associated with ADD risk (the outcome), you can infer the causal effect of the expression of gene B on ADD risk. We limited our results to the genes that were prioritized by our eQTL search and considered a gene significant for expression effect on a disease if it passed an FDR-adjusted SMR significance threshold of $P < 0.05$ and a HEIDI threshold of $P > 0.01$. Filtering for a HEIDI P-value of this magnitude helps to remove associations that are likely due to polygenicity and have violated the central assumptions of SMR. Finally, we assessed the colocalization between the SMR-nominated genes in brain tissues and the multi-ancestry random effects GWAS using LocusCompare [80].

2.1.4.6 Polygenic risk scoring

PRS application cohort

Whole genomes from the Colombian population were accessed from “The Admixture and Neurodegeneration Genomic Landscape” (TANGL) study and quality controlled as previously described [68]. The TANGL cohort was further quality controlled in PLINK v1.9 to remove carriers of pathogenic variants for mendelian forms of dementia, as well as related individuals for a final cohort of 281 cases and 87 controls.

Pre-PRS variant alignment

Base summary statistics were pruned with the MungeSumStats R package [81] to remove multiallelic variants, align reference alleles to build GRCh37, and adjust weights for the

appropriate reference alleles. The target TANGL cohort was also filtered to keep only bi-allelic variants and aligned to the same reference using PLINK v2.0.

PRS method

Polygenic risk score (PRS) analyses can be used to estimate an individual's genetic liability to a phenotype by calculating the sum of risk allele effect size weights for an individual. Weights for the PRS were obtained from β estimates generated from multi-ancestry fixed and random effects meta-analyses as well as from individual ancestry summary statistics. PRS analyses were conducted using PRSice v2.3.5 including variants with minor allele frequency $> 5\%$, genotype missingness $< 10\%$, sample missingness $< 10\%$, and HWE P-value $< 1 \times 10^{-6}$. The *APOE* region (with ranges defined by FUMA as described previously) was excluded prior to variant clumping. For PRS analyses including *APOE*, the genetic polymorphisms rs7412 and rs429358 were added to the QC'd summary statistics prior to variant clumping. β estimates for the *APOE* polymorphisms were not available in the Bellenguez et al. summary statistics [4] and therefore European-ancestry estimates were taken from another recent ADD GWAS by Schwartzentruber et al. [32].

Variants were clumped in each 500 kb window with the index SNP at the center, an r^2 threshold of 0.3, and a clump P-value threshold of 1. Sex, age, and the first 5 PCs were used as covariates in the PRS analysis. PCs were generated from non-imputed genotype data using FlashPCA [82]. Variants with a MAF $< 1\%$, genotype missingness $< 10\%$, sample missingness $< 10\%$, and HWE P-values $< 5 \times 10^{-6}$ were excluded using PLINK v1.9. The remaining variants were pruned with a 1000-kb window, a 10-SNP shift per window and an r^2 threshold of 0.02 prior to PC calculation. PRS analysis was performed at select P-value thresholds to determine the best fit model ($P=5 \times 10^{-10}$, 5×10^{-9} , 5×10^{-8} , 5×10^{-7} , 5×10^{-6} , 5×10^{-5} , 5×10^{-4} , 5×10^{-3} , 5×10^{-2}). To assess the performance of each model, receiver operator characteristic curves were created using the pROC

library in R for the best fit model from each analysis as shown in **Table S6**. Since MR - MEGA does not provide standard effect estimates per SNP, an additional “composite” ROC curve was generated through a linear combination of each super population to provide a comparison to the conservative random-effects-based PRS model. Each PRS was weighted by its associated admixture population percentage, previously determined in the TANGL cohort for each individual (equation 2; AFR: African American, EUR: European (including Finnish), EAS: East Asian, NAT: Native American) [68]. Given the population history and similarities in haplotype structure between the East Asian and Native American populations, Native American admixture proportions were used to weight the East Asian PRS [83, 84].

$$PRS_{composite,i} = PRS_{AFR,i} * \%_{AFR,i} + PRS_{EUR,i} * \%_{EUR,i} + PRS_{EAS,i} * \%_{NAT,i} \quad (2)$$

2.1.5 Figures

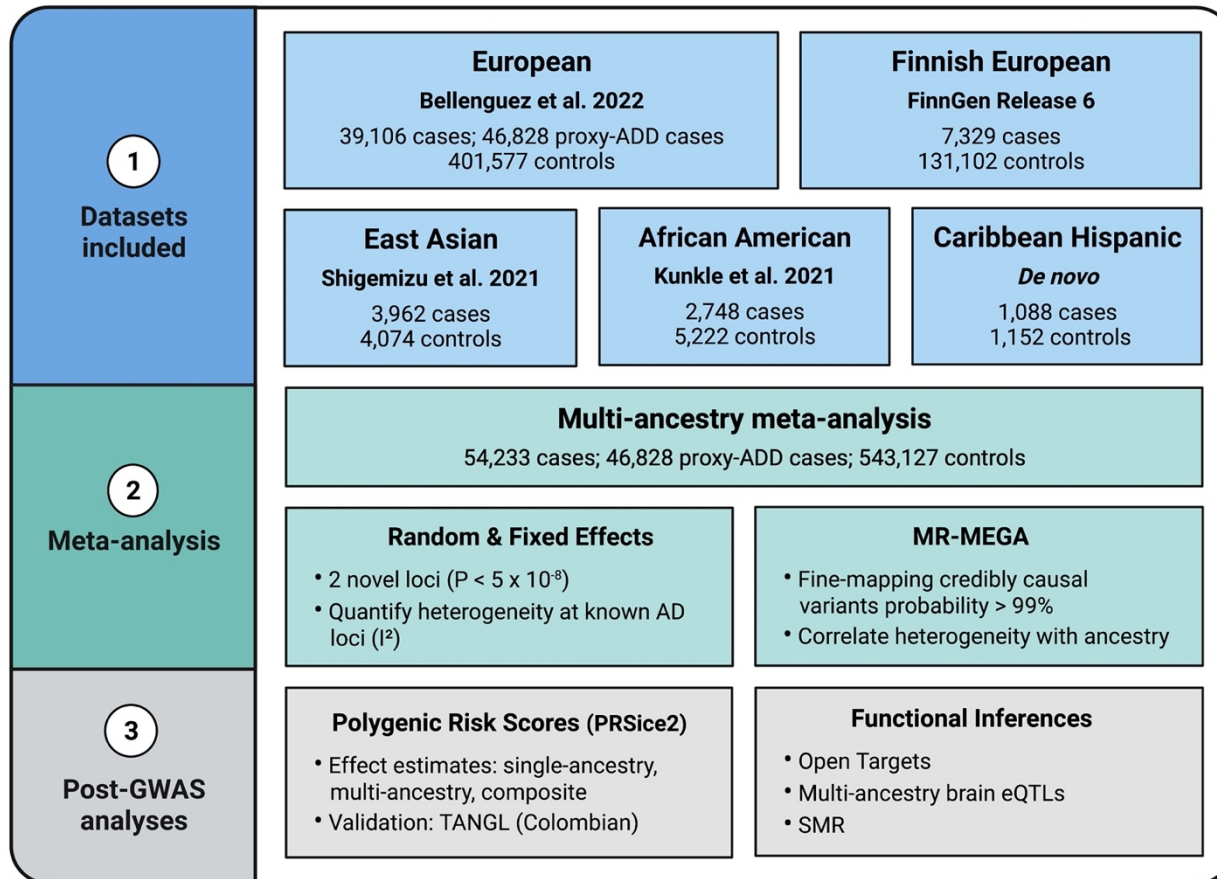


Figure 2.1.1 Outline of multi-ancestry meta-analysis.

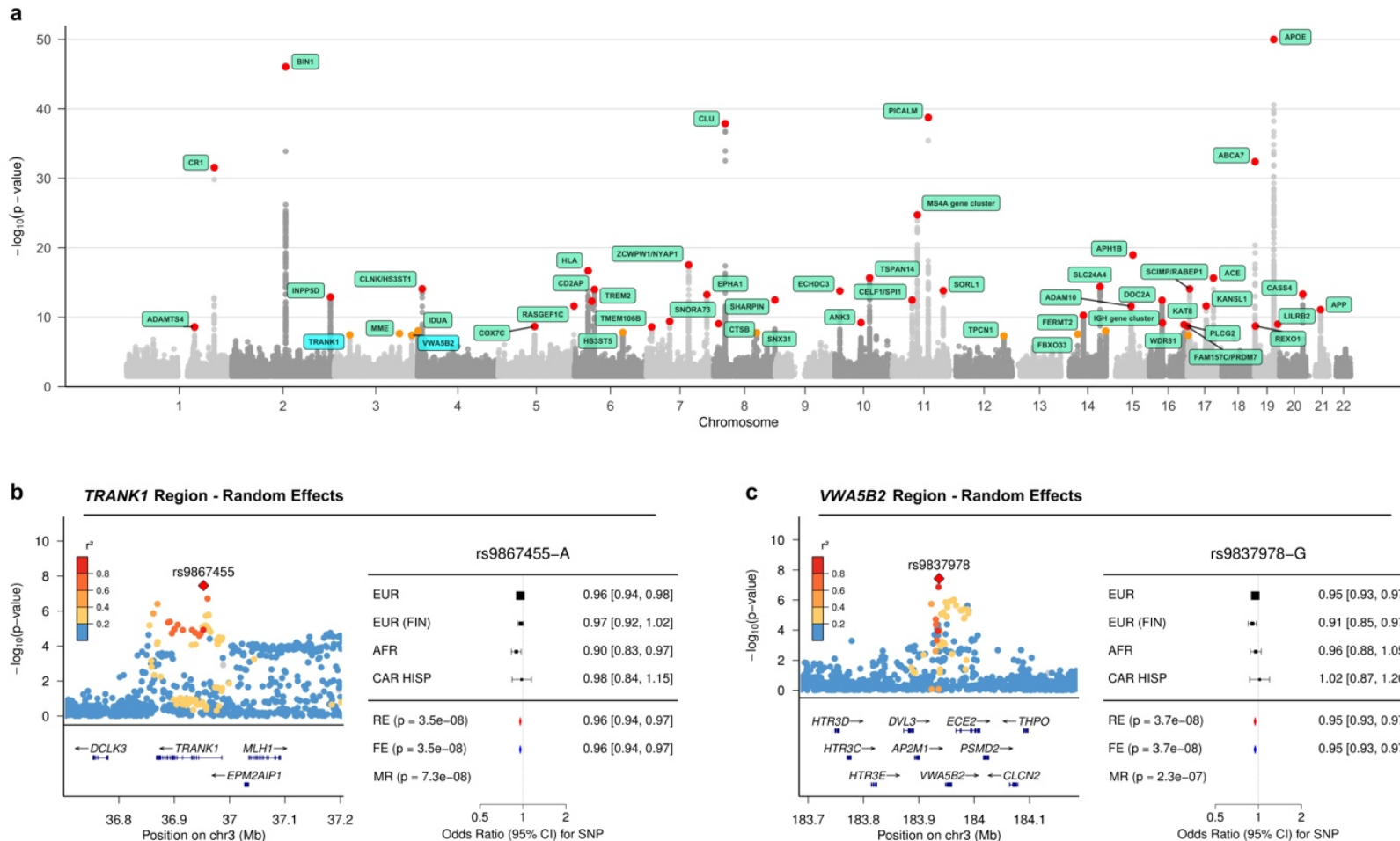


Figure 2.1.2 Summary of multi-ancestry meta-analysis

A) shows the Manhattan plot for the random effects meta-analysis P-values, truncated at $-\log_{10}(P) < 50$. An orange dot indicates that the lead SNP at a locus reached a P-value $< 5 \times 10^{-8}$, while a red dot indicates a P-value $< 5 \times 10^{-9}$. B,C) show the corresponding local association plots for the two loci of interest. C,D) show forest plots summarizing the effect estimates per ancestry group for lead SNPs at the two loci of interest. Lead SNPs from both novel loci were absent in the East Asian dataset by Shigemizu et al. used for discovery.

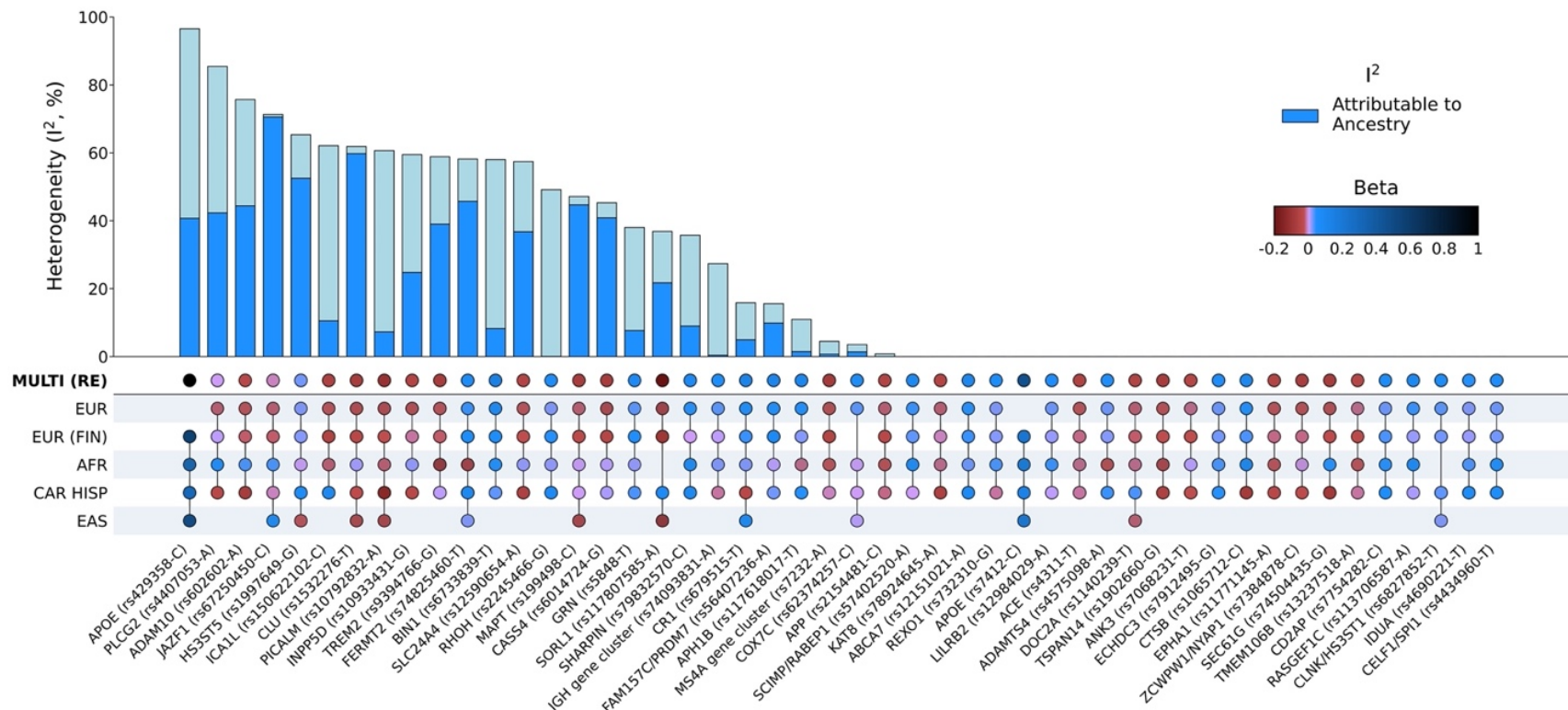


Figure 2.1.3 Graphical summary of heterogeneity at ADD genetic risk loci

Lead SNPs were derived from MR - MEGA using maximal LD blocks, apart from *APOE* rs429358 and rs7412. Both *APOE* SNPs were absent in summary statistics from the most recent European-ancestry ADD GWAS. Aggregate effects were estimated using a random effects model since MR - MEGA assumes that effects differ across populations. Allelic effect heterogeneity that is attributable to genetic ancestry was estimated using Cochran's Q statistics for ancestral and residual heterogeneity from the meta-regression (Online Methods).

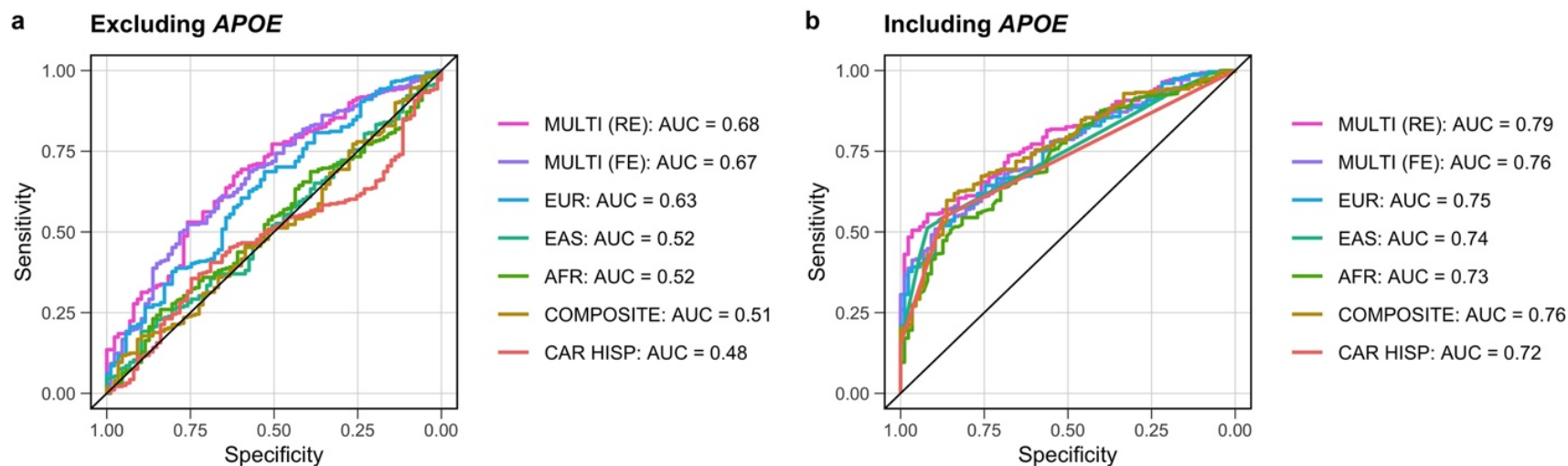


Figure 2.1.4 Graphical summary of genetic risk scores

These genetic risk scores were derived from multi-ancestry and ancestry-specific risk estimates, then applied to an admixed Colombian cohort to evaluate significance and predictive power. The European-based (EUR) PRS was derived from a fixed effect meta-analysis of the summary statistics from Bellenguez et al. and FinnGen used in the meta-GWAS. Panels A and B show the maximal AUCs for each genetic risk score with color coding to delineate the source of the risk estimates for scores excluding and then including *APOE-e4* variants. P-value thresholds that correspond to the maximal AUCs are shown in Table S6.

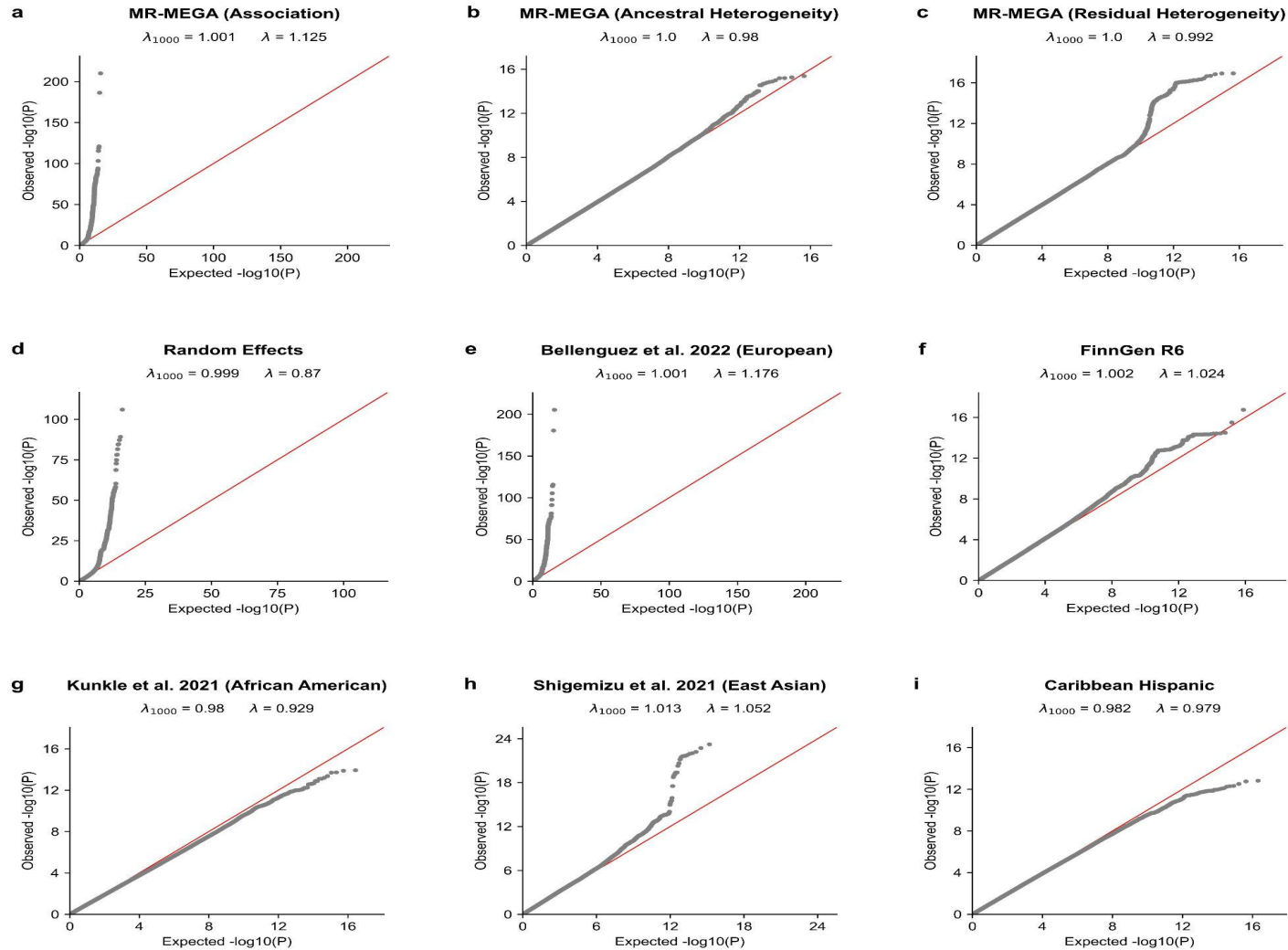


Figure 2.1.S1 Quantile-quantile plots and corresponding genomic inflation estimates
 Plots are presented with and without scaling to 1000 cases and 1000 controls. MR - MEGA P-values are shown for association, ancestral heterogeneity, and residual heterogeneity. Chromosome 19 was excluded from all datasets to avoid bias by the APOE region. All summary statistics were filtered for MAF > 1%.

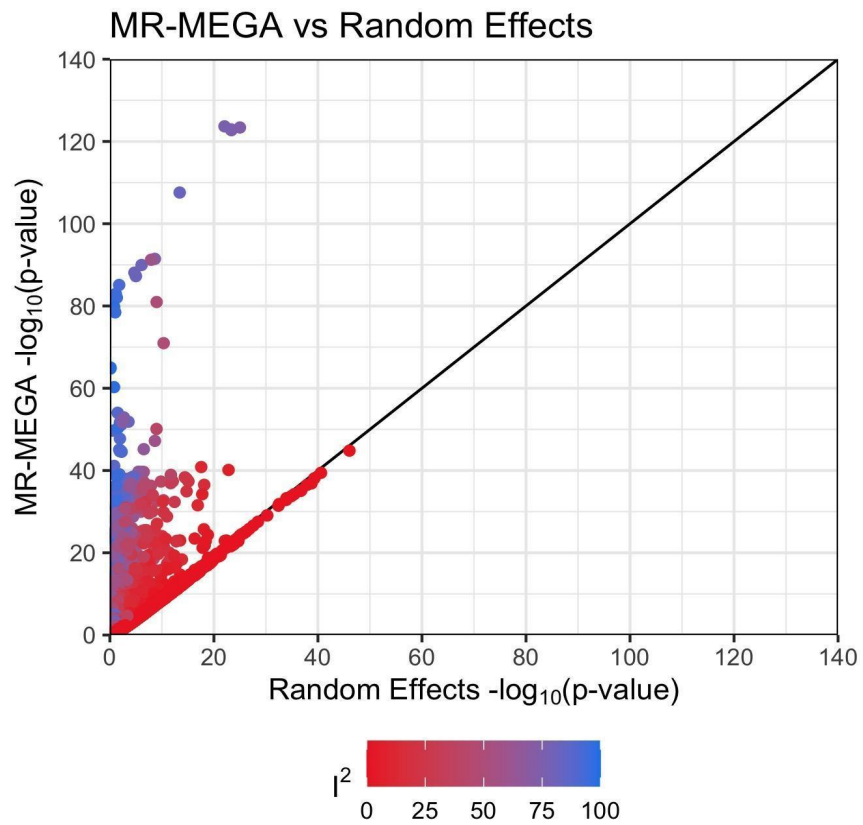


Figure 2.1.S2 MR - MEGA $-\log_{10}$ P-values plotted against random effects $-\log_{10}$ P-values SNPs are colored by I^2 value and are limited to those present in at least 4 dataset

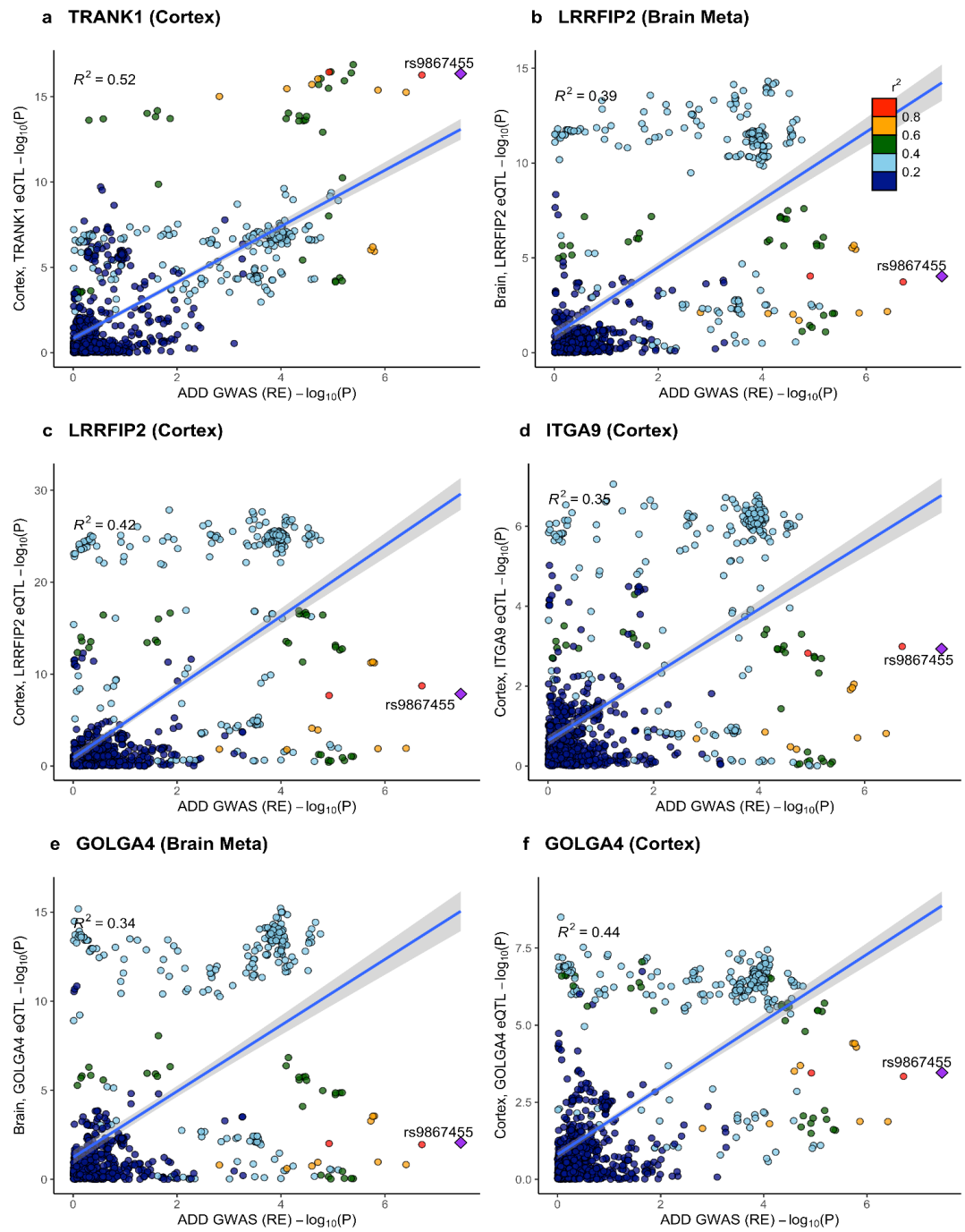


Figure 2.1.S3 LocusCompare plots showing colocalization at the *TRANK1* locus. Colocalization is presented between the random effects meta-analysis results and brain eQTLs ($P < 1 \times 10^{-6}$) in genes that were significant in SMR (FDR $P < 0.05$). Reference LD patterns are based on the European population from 1000 Genomes. Points represent SNPs plotted at their $-\log_{10} P$ -values.

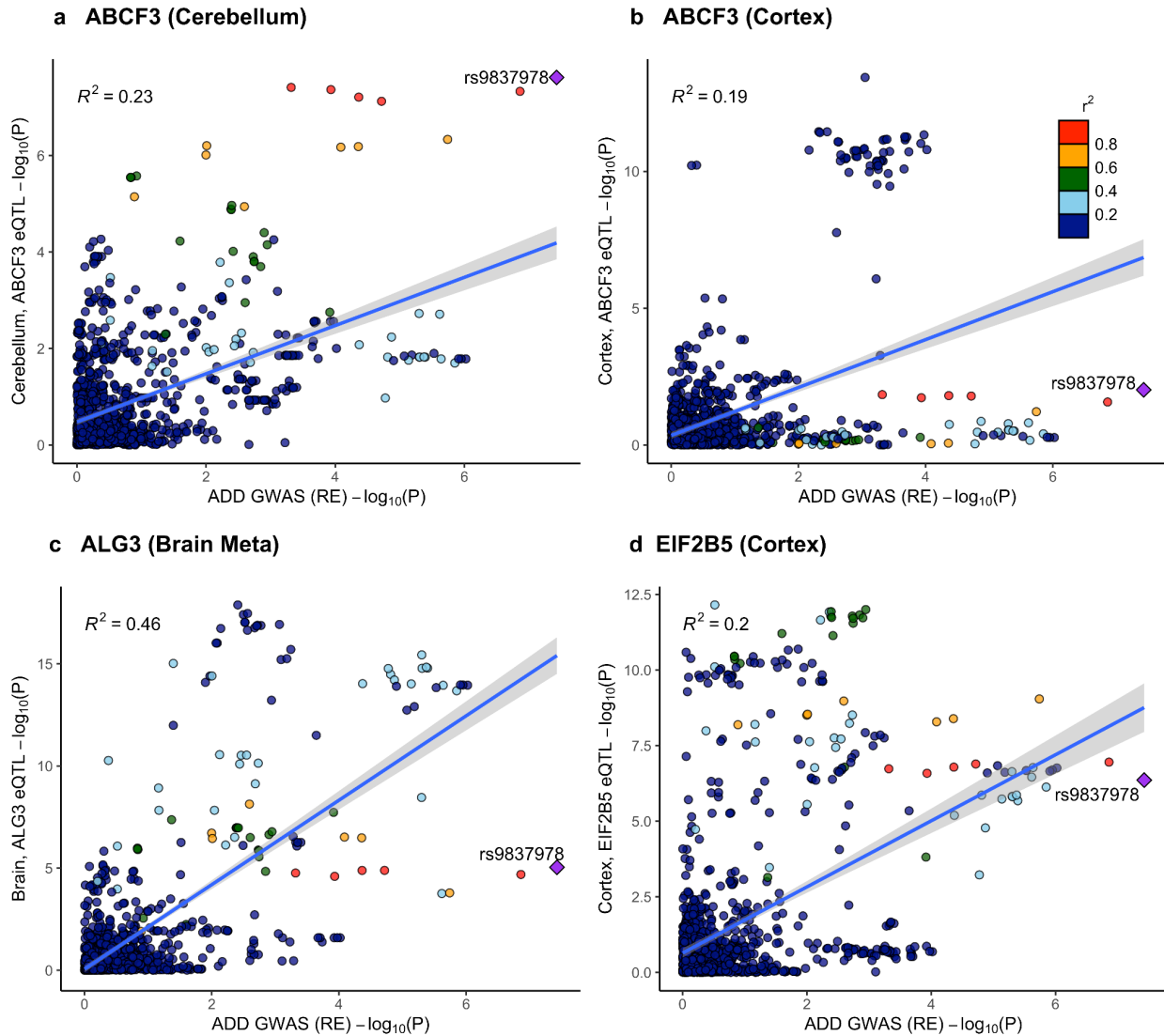


Figure 2.1.S4 LocusCompare plots showing colocalization at the *VWA5B2* locus. Colocalization is presented between the random effects meta-analysis results and brain eQTLs ($P < 1 \times 10^{-6}$) in genes that were significant in SMR ($FDR P < 0.05$). Reference LD patterns are based on the European population from 1000 Genomes. Points represent SNPs plotted at their $-\log_{10} P$ -values.

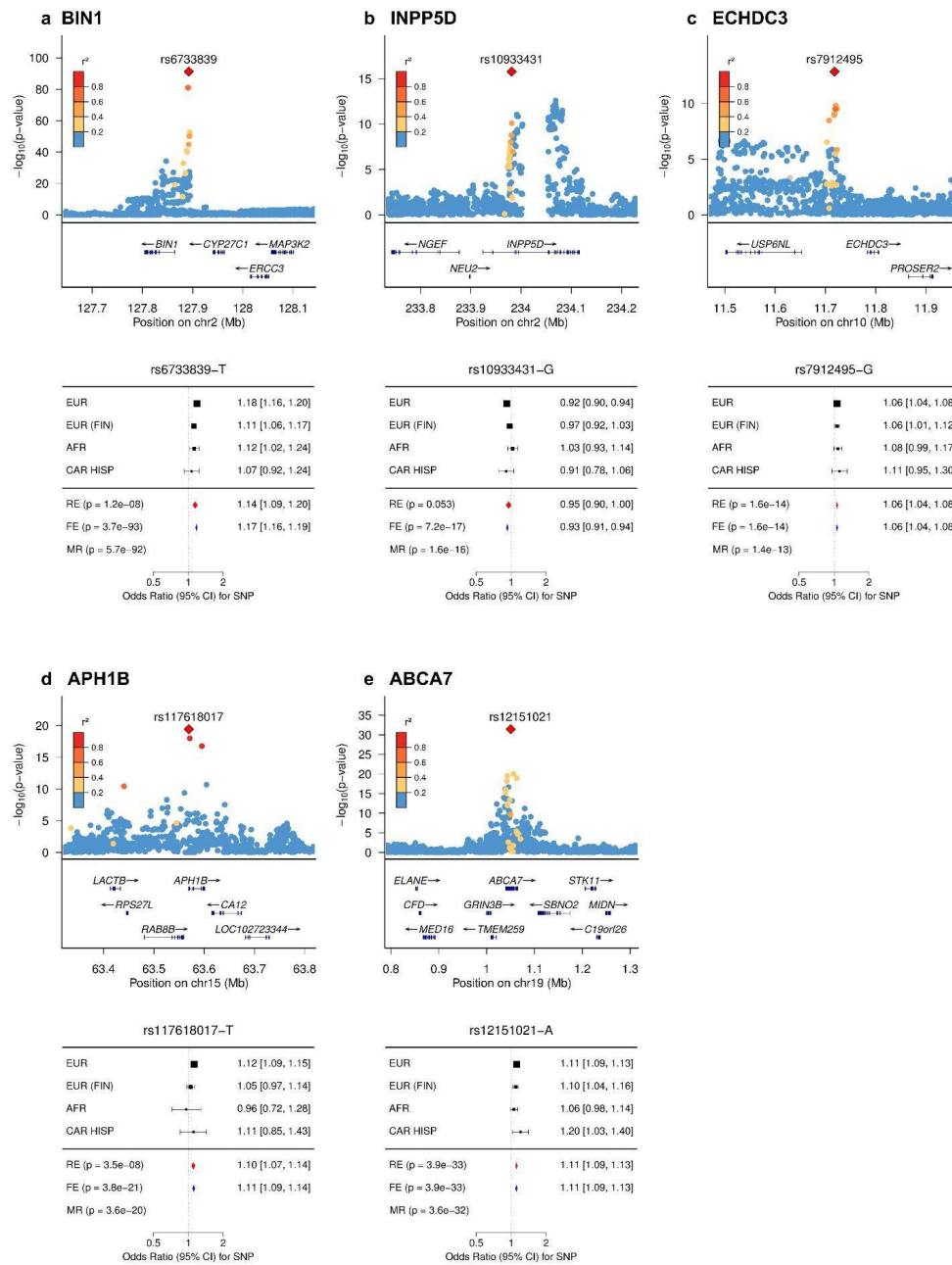


Figure 2.1.S5 LocusZoom and forest plots for fine-mapped SNPs

Fine-mapped SNPs are defined as those with posterior probability (PP) > 0.8 that have been previously fine-mapped in European studies. Reference LD patterns are based on all populations from 1000 Genomes.

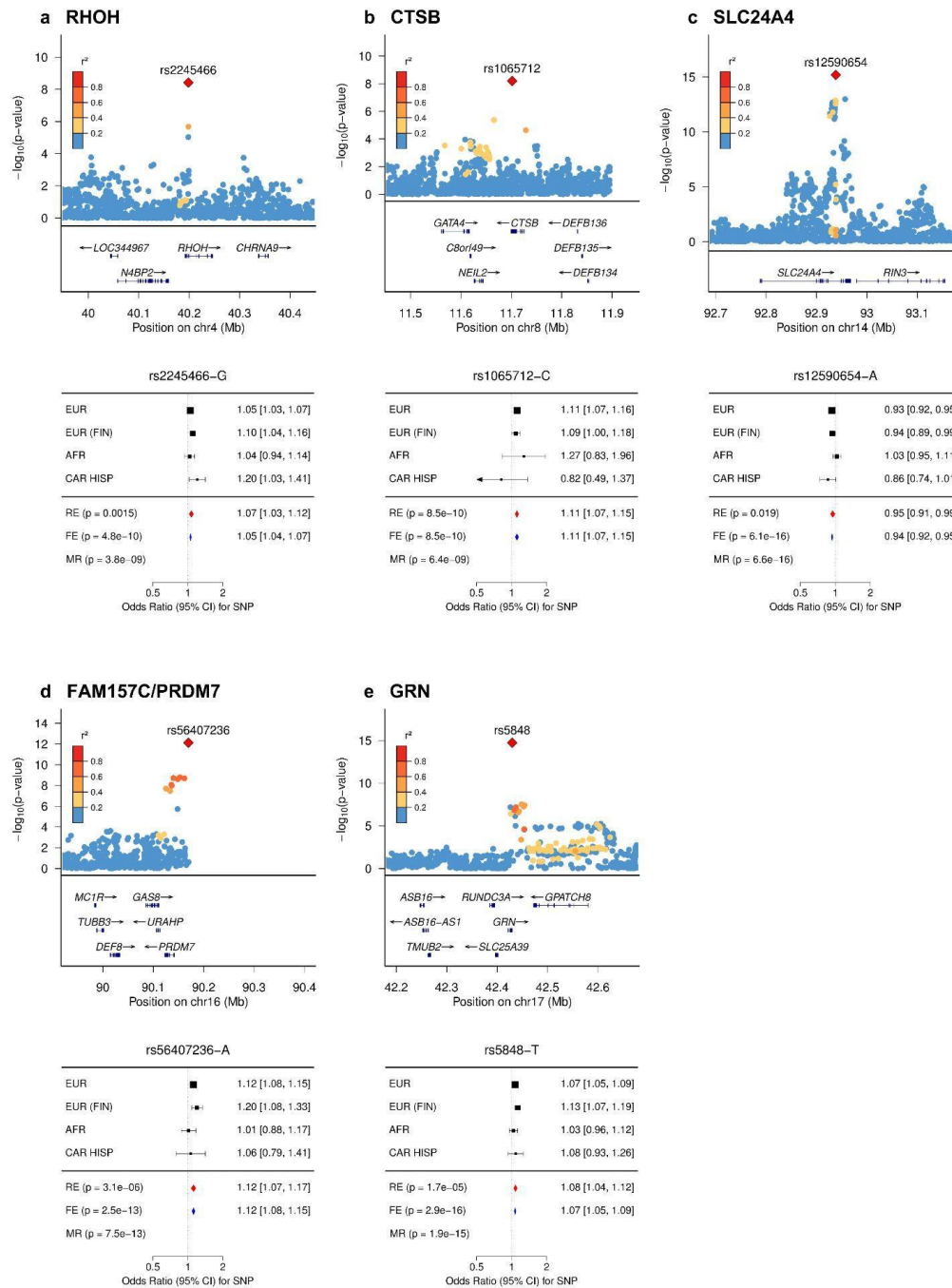


Figure 2.1.S6 LocusZoom and forest plots for fine-mapped SNPs

Fine-mapped in our study are those with posterior probability (PP) > 0.8 that have not been previously fine-mapped. Reference LD patterns are based on all populations from 1000 Genomes.

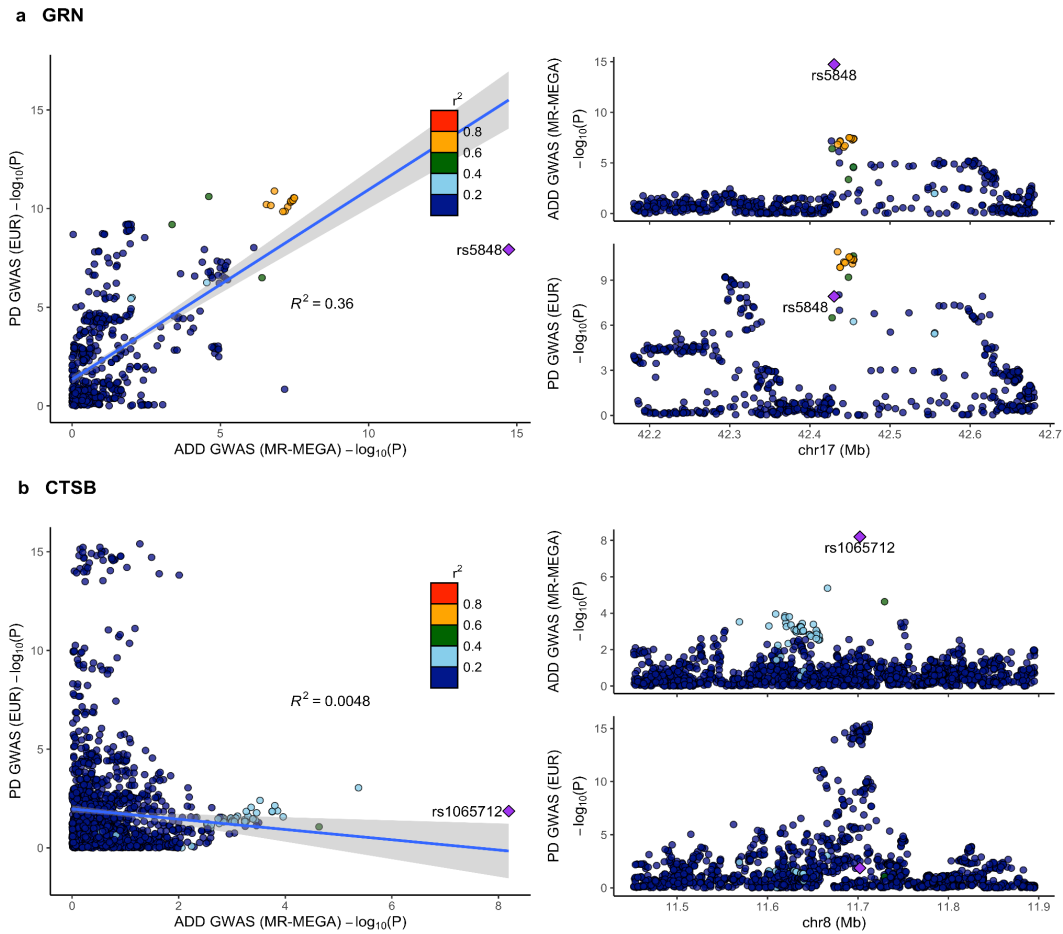


Figure 2.1.S7 LocusCompare plots for a) *GRN* and b) *CTSB* between AD and Parkinson's disease (PD)

Reference LD patterns are based on the European population from 1000 Genomes. Points represent SNPs plotted at their $-\log_{10}$ P-values.

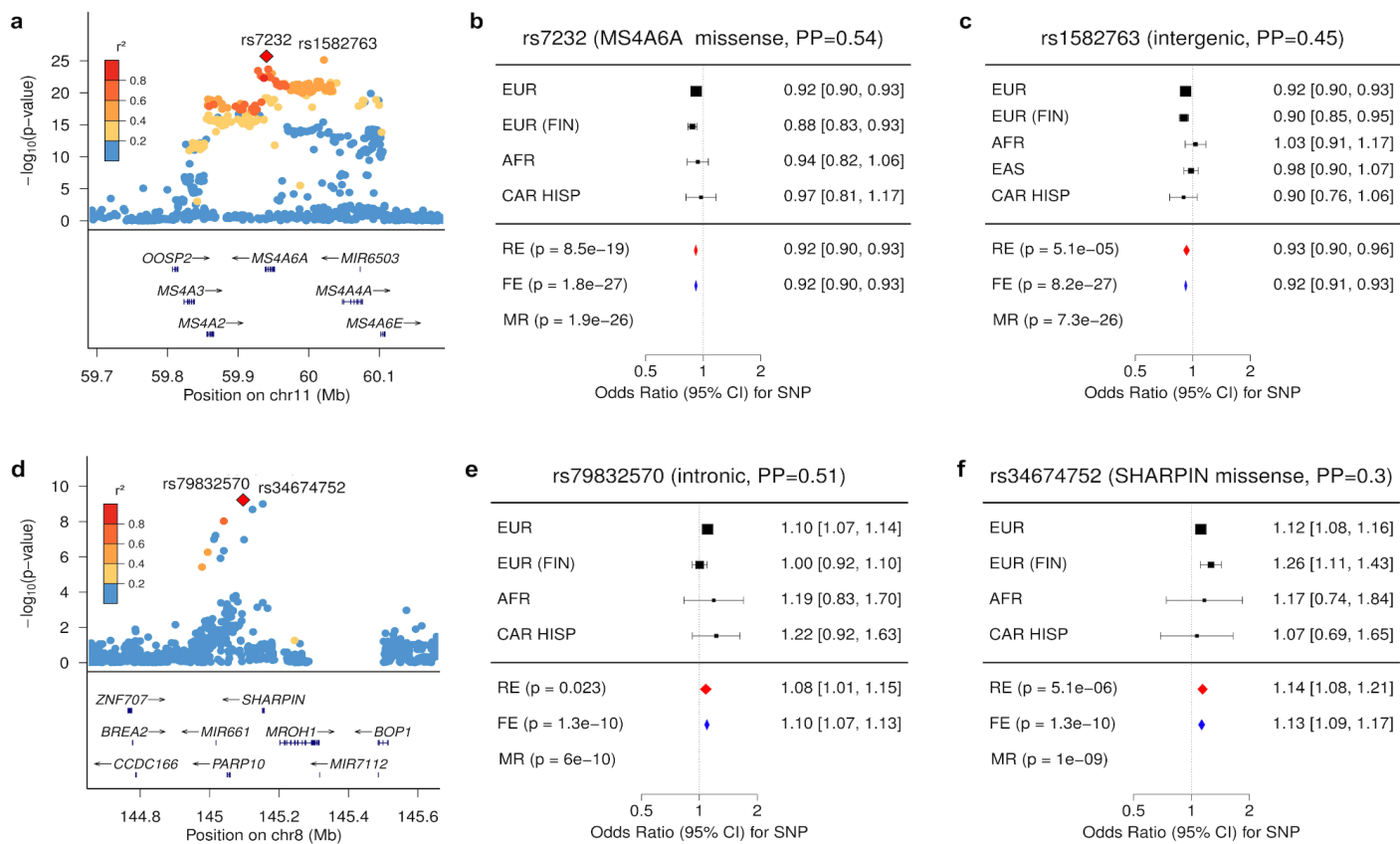


Figure 2.1.S8 LocusZoom and forest plots for loci that were fine-mapped to a missense variant with $PP \geq 0.3$

a) LocusZoom plot for MS4A6A locus highlighting the top two fine-mapped SNPs, rs7232 and rs1582763 b) forest plot for rs7232 c) forest plot for rs1582763 d) LocusZoom plot for SHARPIN locus highlighting the top two fine-mapped SNPs, rs79832570 and rs34674752 e) forest plot for rs79832570 f) forest plot for rs34674752.

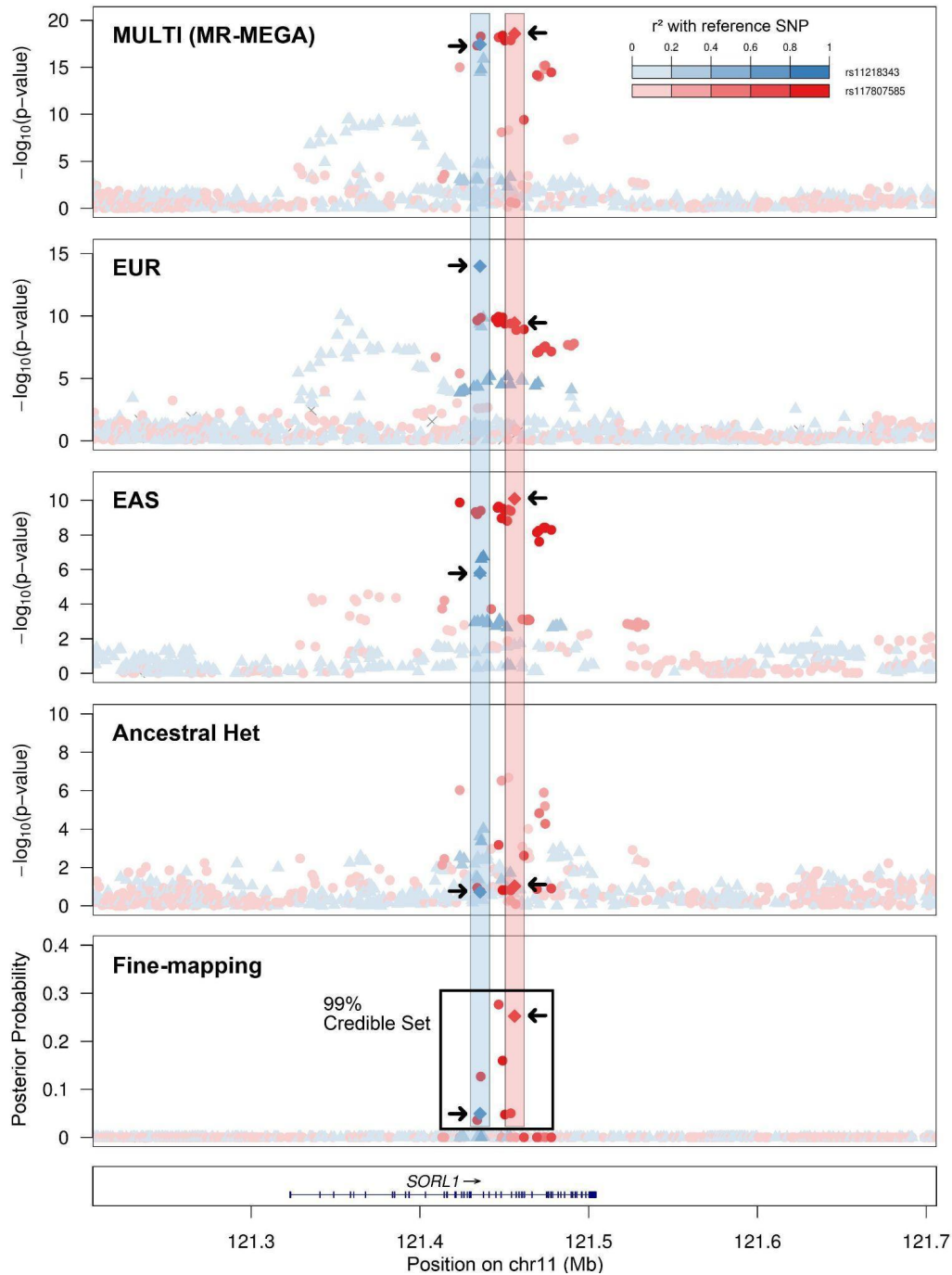


Figure 2.1.S9 LocusZoom plots showing different regional architecture at the *SORL1* locus. Plots show East Asian (Shigemizu et al.) versus European (Bellenguez et al.) populations. Diamond points represent the 2 different lead SNPs at this locus, *SORL1*-rs11218343 in Europeans (shown in blue) and *SORL1*-rs117807585 in East Asians (shown in red). Ancestral Het refers to the P-value of heterogeneity that is due to genetic ancestry as defined by MR - MEGA. SNPs are colored by LD with the respective population from 1000 Genomes, or all populations for the multi-ancestry analysis.

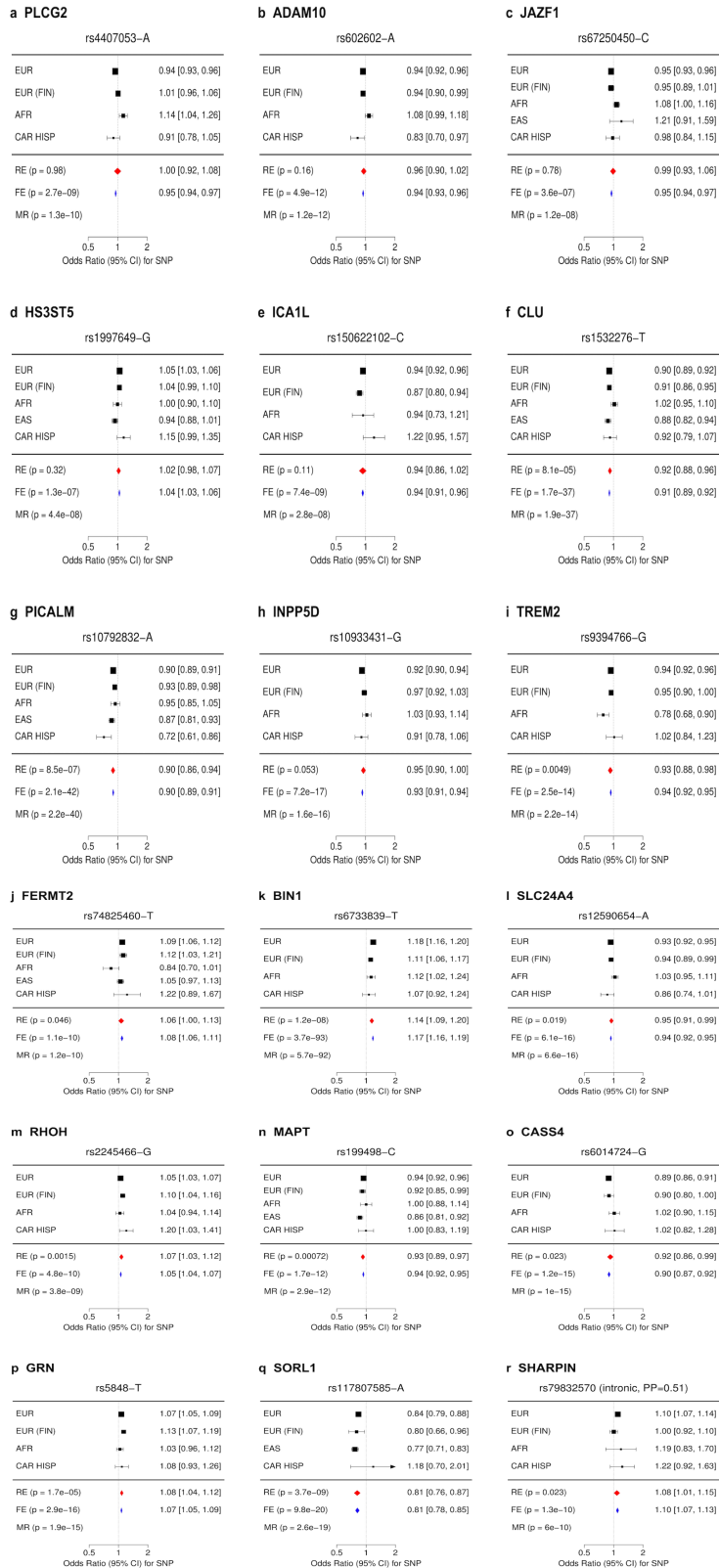


Figure 2.1.S10 Forest plots for loci with significant heterogeneity outside of the *APOE* region

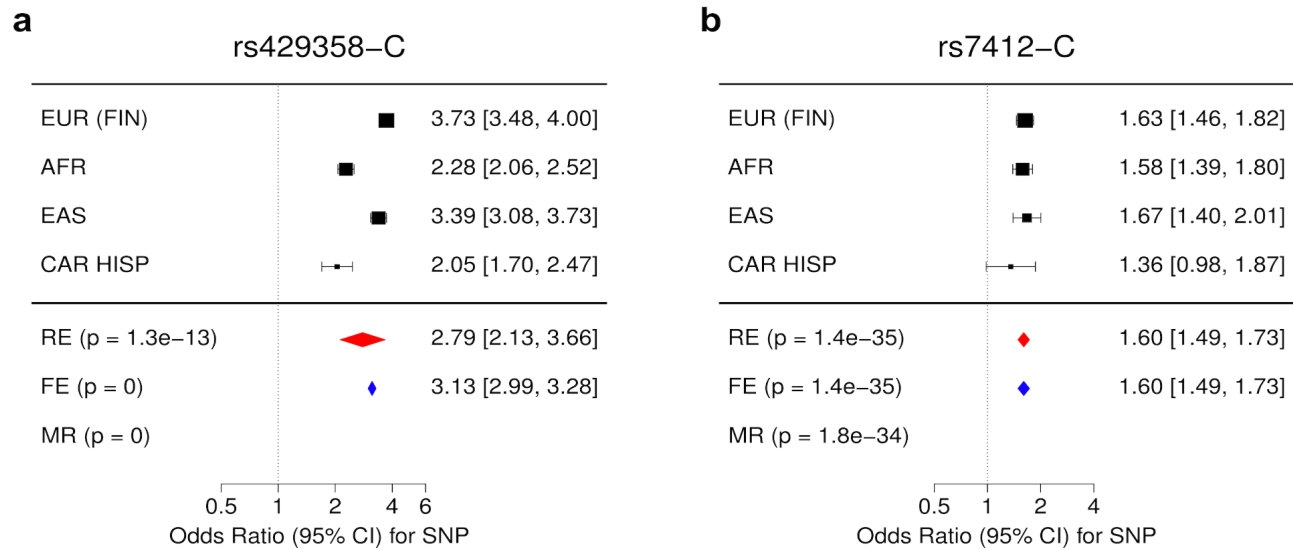


Figure 2.1.S11 Forest plots for a) *APOE*-rs429358 and b) *APOE*-rs7412

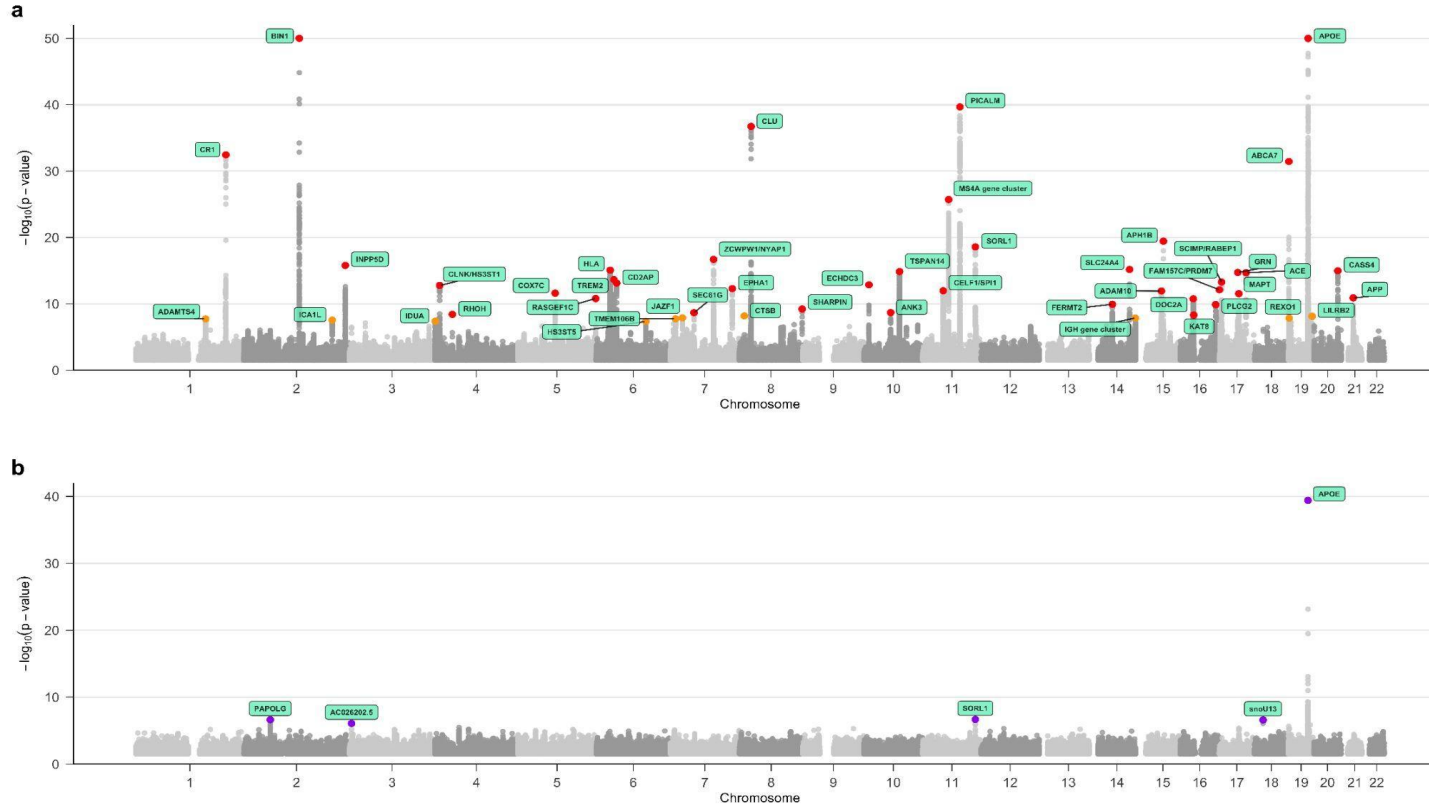


Figure 2.1.S12 a) Manhattan plot for MR - MEGA meta-analysis

The plot is truncated at $-\log_{10}(P) < 50$. Red labeling corresponds to significance at $P < 5 \times 10^{-9}$ and orange corresponds to significance at $P < 5 \times 10^{-8}$.

b) Manhattan plot for PHET from MR - MEGA truncated at $-\log_{10}(P) < 40$. Purple labeling corresponds to $P < 1 \times 10^{-1}$

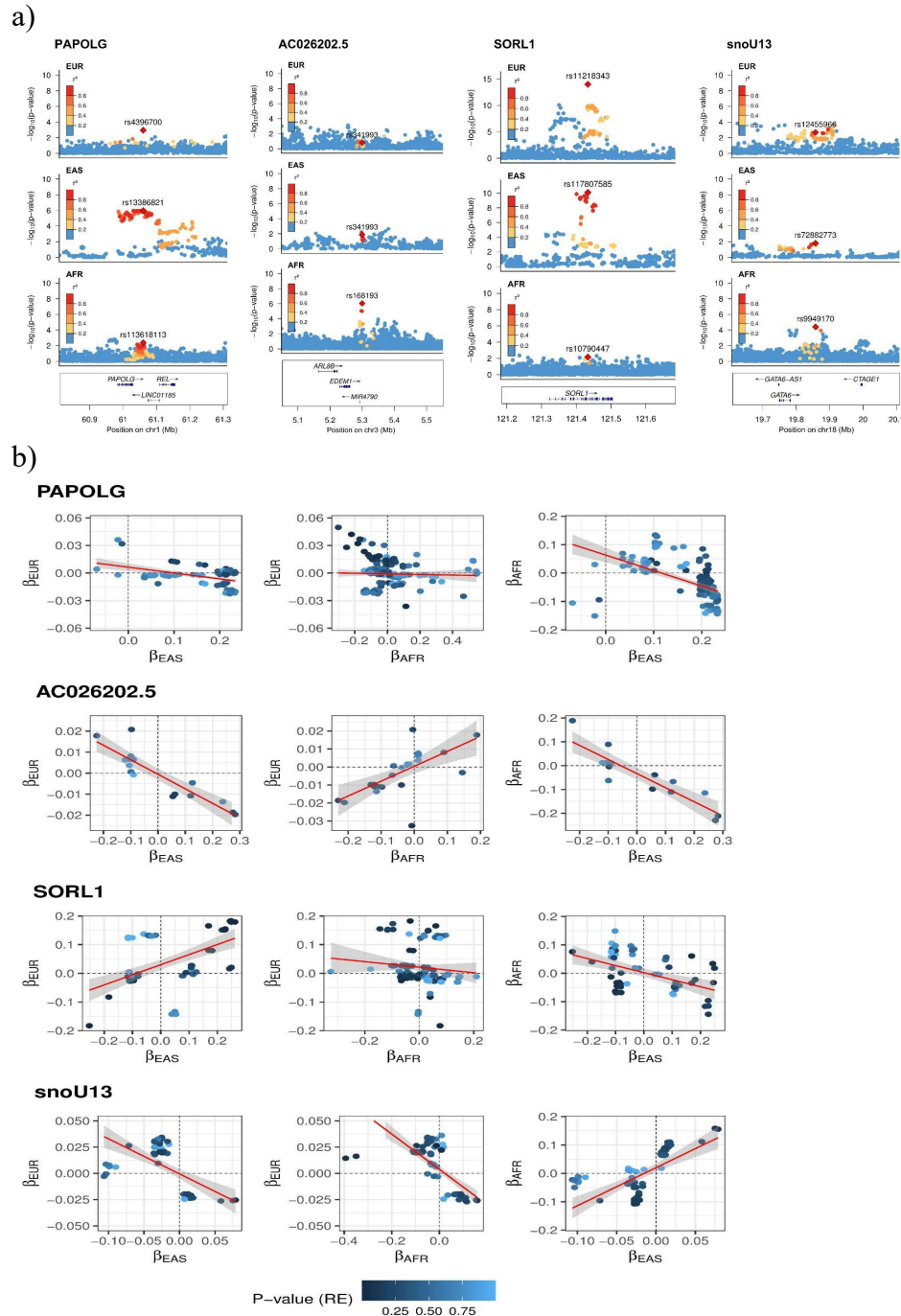


Figure 2.1.S13 LocusZoom and Beta-beta plots

a) LocusZoom plots of loci showing significant ancestry-related heterogeneity (PHET 1×10^{-6}) near SORL1, PAPOLG, AC026202.5, and snoU13. Labeled red diamonds correspond to the lead SNP in each ancestry group. b) Beta-beta plots showing effect size correlation of SORL1, PAPOLG, AC026202.5, and snoU13 across ancestry groups.

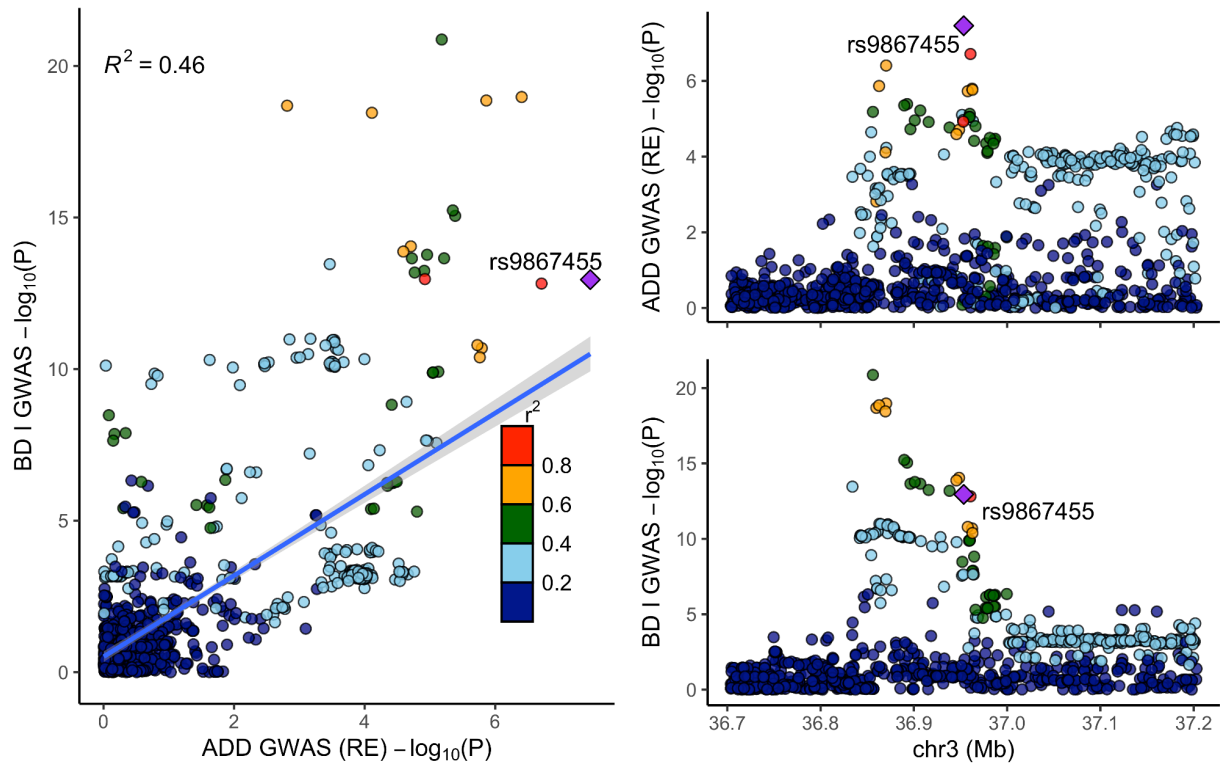


Figure 2.1.S14 LocusCompare at the TRANK 1 locus for BDI
 Plots are presented for bipolar disorder I (BD I) and the random effects ADD meta-analysis.

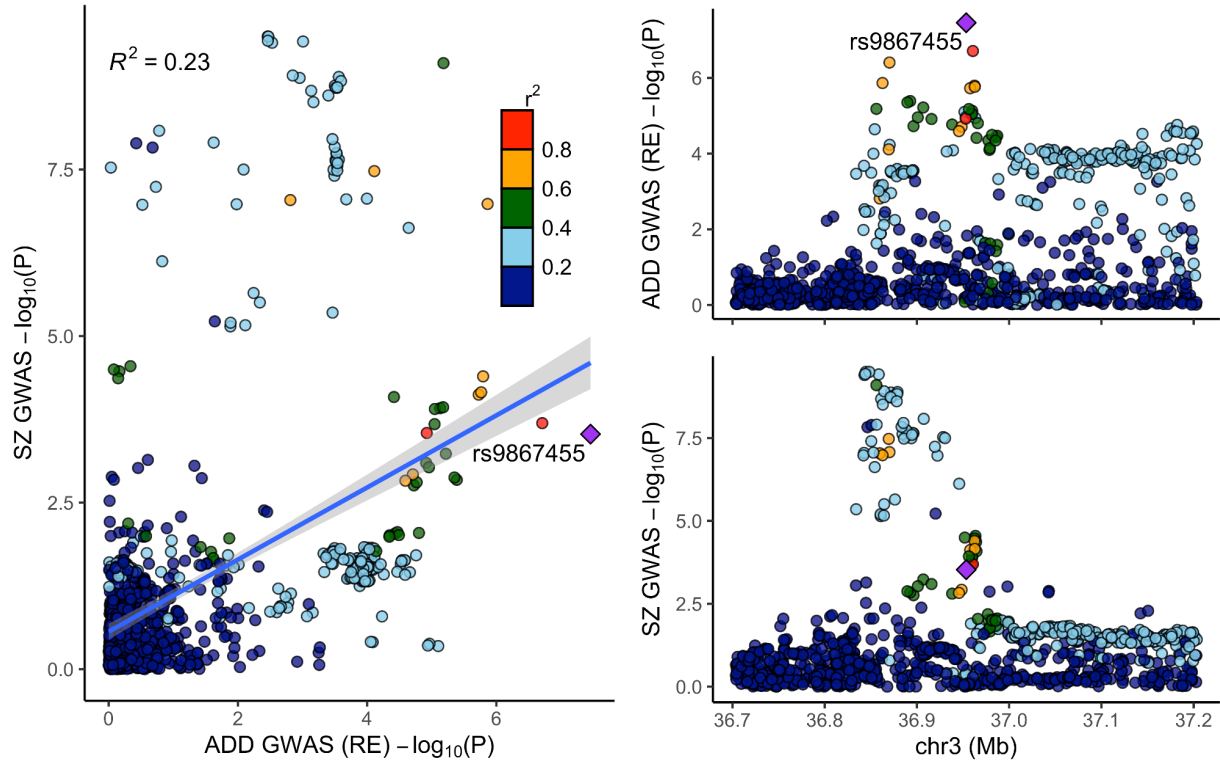


Figure 2.1.S15 LocusCompare at the TRANK 1 locus for SZ
 LocusCompare plot for schizophrenia (SZ) and the random effects ADD meta-analysis at the TRANK1 locus. Reference LD patterns are based on the European population from 1000 Genomes. Points represent SNPs plotted at their $-\log_{10}$ P-values.

About GAS Power Calculator

This Genetic Association Study (GAS) Power Calculator is a simple interface that can be used to compute statistical power for large one-stage genetic association studies. The underlying method is derived from the CaTS power calculator for two-stage association studies (2006).

Inputs

Cases/Controls = 0.955

Sample Size
Cases: 1119
Controls: 1172

Study Design

Significance Level: 0.050

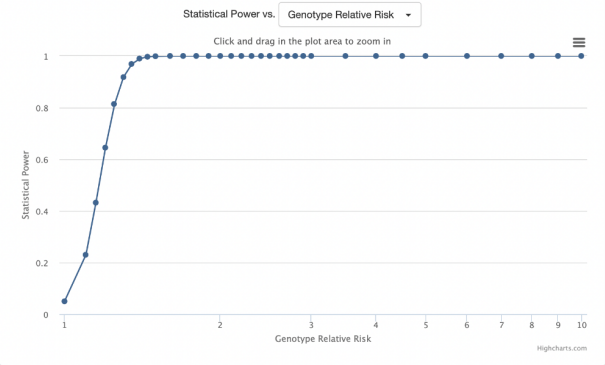
Disease Model: Additive

Prevalence: 0.0200

Disease Allele Frequency: 0.1642

Genotype Relative Risk: 1.2450

Graph



Results



Figure 2.1.S16 Power calculation for *VWA5B2*-rs9837978 in the GARD cohort using GAS Power Calculator

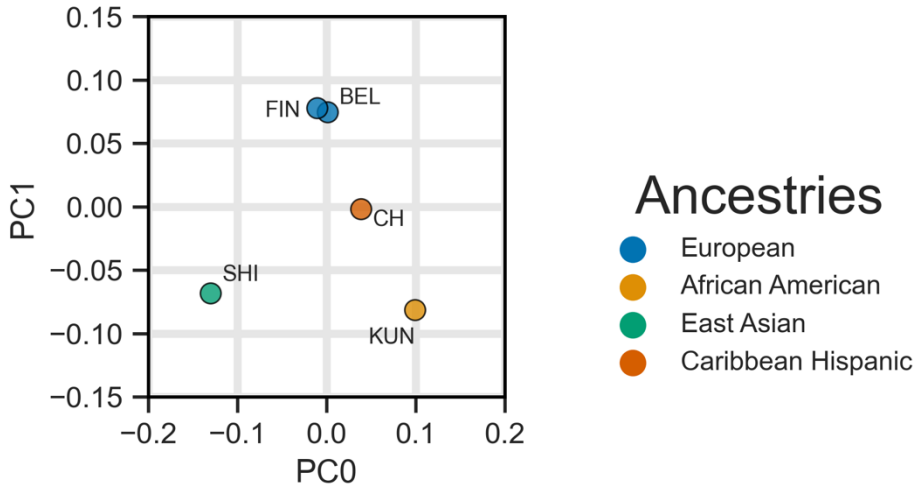


Figure 2.1.S17 The first 2 ancestral principal components (PCs) created and used by MR - MEGA. Points are labeled by dataset (FIN: FinnGen R6, BEL: Bellenguez et al., CH: Caribbean Hispanic, SHI: Shigemizu et al., KUN: Kunkle et al.) and color coded by ancestry.

2.1.6 Tables

Table 2.1.1 Summary of novel loci.

Locus	SNP	Effect allele	Other allele	P, MR - MEGA	P, ancestry heterogeneity	P, RE	Beta, RE	SE, RE	I2
VWA5B2	rs9837978	G	A	2.32E-07	5.96E-01	3.75E-08	-0.053	0.009	0.000
TRANK1	rs9867455	A	T	7.33E-08	1.18E-01	3.49E-08	-0.042	0.008	0.000

Table 2.1.2 Fine-mapping results

Results are presented for all SNPs with a posterior probability (PP) > 0.8. Fine-mapped SNPs were considered novel if they were not previously fine-mapped with PP > 0.8 in two recent European-focused studies by Wightman et al. [31, 32] and Schwartzentruber et al. [31, 32].

Locus	SNP	Chromosome	Position	Effect allele	Reference allele	Novel fine-mapping	Posterior probability	P, MR - MEGA	P, ancestry heterogeneity	chisq, ancestry (% total)	I ²
BIN1	rs6733839	2	127892810	T	C	NO	1	5.74E-92	3.13E-01	14.270	58.010
INPP5D	rs10933431	2	233981912	C	G	NO	1	1.63E-16	7.90E-02	41.675	59.470
RHOH	rs2245466	4	40198846	C	G	YES	1	3.82E-09	9.31E-01	0.128	49.160
CTSB	rs1065712	8	11702122	C	G	YES	1	6.42E-09	7.60E-01	4.551	0.000
ECHDC3	rs7912495	10	11718713	C	G	NO	1	1.35E-13	6.02E-01	48.894	0.000
SLC24A4	rs12590654	14	92938855	A	G	YES	0.910	6.56E-16	3.37E-02	63.950	57.450
APH1B	rs117618017	15	63569902	T	C	NO	0.965	3.63E-20	4.98E-01	13.608	10.950
FAM157C / PRDM7	rs56407236	16	90170095	A	G	YES	1	7.55E-13	1.33E-01	63.528	15.600
GRN	rs5848	17	42430244	T	C	YES	1	1.87E-15	3.22E-01	20.238	38.050
ABCA7	rs12151021	19	1050874	A	G	NO	1	3.62E-32	3.23E-01	34.721	0.000

2.1.7 Supplemental Online Content, Data and Code Availability

Code is available on the CARD GitHub.

Summary statistics from this study will be available to browse and download via our collaboration with the Broad's Neurodegenerative Disease Knowledge Portal.

All supplemental online content can be downloaded directly from the published material at Molecular Psychiatry.

2.1.8 Article Information

2.1.8.1 Acknowledgements

We want to acknowledge the participants and investigators of the FinnGen study. Assistance with phenotype harmonization and genotype cleaning, as well as with general study coordination, was provided by the Genetic Consortium for Late Onset Alzheimer's Disease. Additionally, we want to thank Dr. Francisco Lopera and the Neuroscience Group of Antioquia and the University of Antioquia (Colombia) for letting us use the TANGL cohort.

Funding acknowledgement

This work was supported in part by the Intramural Research Programs of the National Institute on Aging (NIA) part of the National Institutes of Health, Department of Health and Human Services; project numbers 1ZIA-NS003154, Z01-AG000949-02, Z01-AG000535 and Z01-ES101986. J.S.Y. is supported by NIH-NIA R01 AG062588, R01 AG057234, P30 AG062422; NIH-NINDS U54 NS123985; the Rainwater Charitable Foundation; the Alzheimer's Association; the Global Brain Health Institute; and the Mary Oakley Foundation. This research was supported by the Korea Brain Research Institute basic research program funded by the Ministry of Science and ICT (22- BR-03-05). Funding support for the "Genetic Consortium for Late Onset Alzheimer's Disease" was provided through the Division of Neuroscience, NIA. The Genetic Consortium for Late Onset Alzheimer's Disease includes a genome-wide association study funded as part of the Division of Neuroscience, NIA.

2.1.8.2 Author Information

These authors contributed equally: Julie Lake and Caroline Warly Solsberg as well as Jennifer Yokoyama and Hampton Leonard.

Contributions

Designed the study.

Julie Lake, Caroline Warly Solsberg, Peter Heutink, Luigi Ferrucci, Andrew B. Singleton, Mike A. Nalls, Jennifer S. Yokoyama, and Hampton L. Leonard

Acquired and processed data.

Julie Lake, Caroline Warly Solsberg, Juliana Acosta-Uribe, Mary B. Makarios, Zizheng Li, Chelsea Alvarado, Dan Vitale, Sarang Kang, Jungsoo Gim, Stefanie D. Pina-Escudero, Luigi Ferrucci, Andrew B. Singleton, Cornelis Blauwendraat, Mike A. Nalls, Jennifer S. Yokoyama, and Hampton L. Leonard

Analyzed data.

Julie Lake, Caroline Warly Solsberg, Jonggeol Jeffrey Kim, Juliana Acosta-Uribe, Mary B. Makarios, Zizheng Li, Chelsea Alvarado, Dan Vitale, Sarang Kang, Jungsoo Gim, Stefanie D. Pina-Escudero, Andrew B. Singleton, Cornelis Blauwendraat, Mike A. Nalls, Jennifer S. Yokoyama, and Hampton L. Leonard

Interpreted data, writing and revising the first draft of the manuscript.

All authors.

2.1.8.3 Correspondence

Correspondence to Hampton L. Leonard.

2.1.8.4 Ethics Declaration

Competing interests

K.L., D.V., H.L. and M.A.N.'s participation in this project was part of a competitive contract awarded to Data Tecnica International LLC by the National Institutes of Health to support open science research. M.A.N. also currently serves on the scientific advisory board for Clover Therapeutics and is an advisor to Neuron23 Inc. P.H. is also an advisor to Neuron23 Inc. J.S.Y. serves on the scientific advisory board for the Epstein Family Alzheimer's Research Collaboration.

2.1.9 References

1. Tanzi RE, Bertram L. New frontiers in Alzheimer's disease genetics. *Neuron*. 2001;32:181–184.
2. Holland D, Frei O, Desikan R, Fan C-C, Shadrin AA, Smeland OB, et al. The genetic architecture of human complex phenotypes is modulated by linkage disequilibrium and heterozygosity. *Genetics*. 2021;217.
3. Zhang Q, Sidorenko J, Couvy-Duchesne B, Marioni RE, Wright MJ, Goate AM, et al. Risk prediction of late-onset Alzheimer's disease implies an oligogenic architecture. *Nat Commun*. 2020;11:4799.
4. Bellenguez C, Küçükali F, Jansen IE, Klei L, Moreno-Grau S, Amin N, et al. New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat Genet*. 2022;54:412–436.
5. Hou K, Bhattacharya A, Mester R, Burch KS, Pasaniuc B. On powerful GWAS in admixed populations. *Nat Genet*. 2021;53:1631–1633.
6. Kunkle BW, Schmidt M, Klein H-U, Naj AC, Hamilton-Nelson KL, Larson EB, et al. Novel Alzheimer Disease Risk Loci and Pathways in African American Individuals Using the African Genome Resources Panel: A Meta-analysis. *JAMA Neurol*. 2021;78:102–113.
7. Blue EE, Arvr H, Mukherjee S, Wijsman EM, Thornton TA. Local ancestry at APOE modifies Alzheimer's disease risk in Caribbean Hispanics. *Alzheimers Dement*. 2019;15.
8. Graham SE, Clarke SL, Wu K-HH, Kanoni S, Zajac GJM, Ramdas S, et al. The power of genetic diversity in genome-wide association studies of lipids. *Nature*. 2021;600:675–679.
9. Mahajan A, Spracklen CN, Zhang W, Ng MCY, Petty LE, Kitajima H, et al. Multi-ancestry

genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. *Nat Genet.* 2022;54:560–572.

10. McCartney DL, Min JL, Richmond RC, Lu AT, Sobczyk MK, Davies G, et al. Genome-wide association studies identify 137 genetic loci for DNA methylation biomarkers of aging. *Genome Biol.* 2021;22.
11. Laisk T, Soares ALG, Ferreira T, Painter JN, Censin JC, Laber S, et al. The genetic architecture of sporadic and multiple consecutive miscarriage. *Nat Commun.* 2020;11.
12. Chen M-H, Raffield LM, Mousas A, Sakaue S, Huffman JE, Moscati A, et al. Trans-ethnic and ancestry-specific blood-cell genetics in 746,667 individuals from 5 global populations. *Cell.* 2020;182:1198.
13. Shu X, Long J, Cai Q, Kweon S-S, Choi J-Y, Kubo M, et al. Identification of novel breast cancer susceptibility loci in meta-analyses conducted among Asian and European descendants. *Nat Commun.* 2020;11.
14. Tin A, Marten J, Halperin Kuhns VL, Li Y, Wuttke M, Kirsten H, et al. Target genes, variants, tissues and transcriptional pathways influencing human serum urate levels. *Nat Genet.* 2019;51:1459.
15. Teumer A, Li Y, Ghasemi S, Prins BP, Wuttke M, Hermle T, et al. Genome-wide association meta-analyses and fine-mapping elucidate pathways influencing albuminuria. *Nat Commun.* 2019;10.
16. Wuttke M, Li Y, Li M, Sieber KB, Feitosa MF, Gorski M, et al. A catalog of genetic loci associated with kidney function from analyses of a million individuals. *Nat Genet.* 2019;51:957.
17. Morris AP, Le TH, Wu H, Akbarov A, van der Most PJ, Hemani G, et al. Trans-ethnic

- kidney function association study reveals putative causal genes and effects on kidney-specific disease aetiologies. *Nat Commun.* 2019;10.
18. Daya M, Rafaels N, Brunetti TM, Chavan S, Levin AM, Shetty A, et al. Association study in African-admixed populations across the Americas recapitulates asthma risk loci in non-African populations. *Nat Commun.* 2019;10.
 19. Shigemizu D, Mitsumori R, Akiyama S, Miyashita A, Morizono T, Higaki S, et al. Ethnic and trans-ethnic genome-wide association studies identify new loci influencing Japanese Alzheimer's disease risk. *Translational Psychiatry.* 2021;11.
 20. Kang S, Gim J, Lee J, Gunasekaran TI, Choi KY, Lee JJ, et al. Potential Novel Genes for Late-Onset Alzheimer's Disease in East-Asian Descent Identified by APOE-Stratified Genome-Wide Association Study. *J Alzheimers Dis.* 2021;82.
 21. Conti DV, Darst BF, Moss LC, Saunders EJ, Sheng X, Chou A, et al. Trans-ancestry genome-wide association meta-analysis of prostate cancer identifies new susceptibility loci and informs genetic risk prediction. *Nat Genet.* 2021;53:65–75.
 22. Ntalla I, Weng L-C, Cartwright JH, Hall AW, Sveinbjornsson G, Tucker NR, et al. Multi-ancestry GWAS of the electrocardiographic PR interval identifies 202 loci underlying cardiac conduction. *Nat Commun.* 2020;11:2542.
 23. Bentley AR, Sung YJ, Brown MR, Winkler TW, Kraja AT, Ntalla I, et al. Multi-ancestry genome-wide gene-smoking interaction study of 387,272 individuals identifies new loci associated with serum lipids. *Nat Genet.* 2019;51:636–648.
 24. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience.* 2015;4:7.
 25. Mägi R, Horikoshi M, Sofer T, Mahajan A, Kitajima H, Franceschini N, et al. Trans-

- ethnic meta-regression of genome-wide association studies accounting for ancestry increases power for discovery and improves fine-mapping resolution. *Hum Mol Genet.* 2017;26:3639–3650.
26. Ochoa D, Hercules A, Carmona M, Suveges D, Gonzalez-Uriarte A, Malangone C, et al. Open Targets Platform: supporting systematic drug-target identification and prioritisation. *Nucleic Acids Res.* 2021;49:D1302–D1310.
27. Zeng B, Bendl J, Kosoy R, Fullard JF, Hoffman GE, Roussos P. Multi-ancestry eQTL meta-analysis of human brain identifies candidate causal variants for brain-related traits. *Nat Genet.* 2022;54:161–169.
28. Steele NZR, Carr JS, Bonham LW, Geier EG, Damotte V, Miller ZA, et al. Fine-mapping of the human leukocyte antigen locus as a risk factor for Alzheimer disease: A case–control study. *PLoS Med.* 2017;14:e1002272.
29. Hirata J, Hosomichi K, Sakaue S, Kanai M, Nakaoka H, Ishigaki K, et al. Genetic and phenotypic landscape of the major histocompatibility complex region in the Japanese population. *Nat Genet.* 2019;51.
30. Sánchez-Juan P, Moreno S, de Rojas I, Hernández I, Valero S, Alegret M, et al. The MAPT H1 Haplotype Is a Risk Factor for Alzheimer’s Disease in APOE ϵ 4 Non-carriers. *Front Aging Neurosci.* 2019;0.
31. Wightman DP, Jansen IE, Savage JE, Shadrin AA, Bahrami S, Holland D, et al. A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer’s disease. *Nat Genet.* 2021;53:1276–1282.
32. Schwartzentruber J, Cooper S, Liu JZ, Barrio-Hernandez I, Bello E, Kumasaka N, et al. Genome-wide meta-analysis, fine-mapping and integrative prioritization implicate new

- Alzheimer's disease risk genes. *Nat Genet.* 2021;53:392–402.
33. Sirkis DW, Geier EG, Bonham LW, Karch CM, Yokoyama JS. Recent advances in the genetics of frontotemporal dementia. *Curr Genet Med Rep.* 2019;7:41–52.
 34. Nalls MA, Blauwendraat C, Vallerga CL, Heilbron K, Bandres-Ciga S, Chang D, et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-genome wide association study. *Lancet Neurol.* 2019;18:1091.
 35. Nalls MA, Blauwendraat C, Sargent L, Vitale D, Leonard H, Iwaki H, et al. Evidence for GRN connecting multiple neurodegenerative diseases. *Brain Commun.* 2021;3:fcab095.
 36. Miyashita A, Koike A, Jun G, Wang L-S, Takahashi S, Matsubara E, et al. SORL1 Is Genetically Associated with Late-Onset Alzheimer's Disease in Japanese, Koreans and Caucasians. *PLoS One.* 2013;8.
 37. Rajabli F, Feliciano BE, Celis K, Hamilton-Nelson KL, Whitehead PL, Adams LD, et al. Ancestral origin of ApoE ϵ 4 Alzheimer disease risk in Puerto Rican and African American populations. *PLoS Genet.* 2018;14.
 38. Tosto G, Fu H, Vardarajan BN, Lee JH, Cheng R, Reyes-Dumeyer D, et al. F-box/LRR-repeat protein 7 is genetically associated with Alzheimer's disease. *Annals of Clinical and Translational Neurology.* 2015;2:810.
 39. Chen DT, Jiang X, Akula N, Shugart YY, Wendland JR, Steele CJM, et al. Genome-wide association study meta-analysis of European and Asian-ancestry samples identifies three novel loci associated with bipolar disorder. *Mol Psychiatry.* 2011;18:195–205.
 40. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. *Genome Res.* 2002;12:996–1006.
 41. Schizophrenia Working Group of the Psychiatric Genomics Consortium, Ripke S, Neale

- BM, Corvin A, Walters JTR, Farh K-H, et al. Biological Insights From 108 Schizophrenia-Associated Genetic Loci. *Nature*. 2014;511:421.
42. Lam M, Chen C-Y, Li Z, Martin AR, Bryois J, Ma X, et al. Comparative genetic architectures of schizophrenia in East Asian and European populations. *Nat Genet*. 2019;51:1670–1678.
43. Ikeda M, Takahashi A, Kamatani Y, Momozawa Y, Saito T, Kondo K, et al. Genome-Wide Association Study Detected Novel Susceptibility Genes for Schizophrenia and Shared Trans-Populations/Diseases Genetic Effect. *Schizophr Bull*. 2019;45:824–834.
44. Goes FS, Hamshere ML, Seifuddin F, Pirooznia M, Belmonte-Mahon P, Breuer R, et al. Genome-wide association of mood-incongruent psychotic bipolar disorder. *Transl Psychiatry*. 2012;2:e180.
45. Ruderfer DM, Fanous AH, Ripke S, McQuillin A, Amdur RL, Schizophrenia Working Group of the Psychiatric Genomics Consortium, et al. Polygenic dissection of diagnosis and clinical dimensions of bipolar disorder and schizophrenia. *Mol Psychiatry*. 2014;19:1017–1024.
46. Mullins N, Forstner AJ, O’Connell KS, Coombes B, Coleman JRI, Qiao Z, et al. Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nat Genet*. 2021;53:817–829.
47. Drange OK, Smeland OB, Shadrin AA, Finseth PI, Witoelar A, Frei O, et al. Genetic Overlap Between Alzheimer’s Disease and Bipolar Disorder Implicates the MARK2 and VAC14 Genes. *Front Neurosci*. 2019;13:220.
48. Li W, Cai X, Li H-J, Song M, Zhang C-Y, Yang Y, et al. Independent replications and integrative analyses confirm TRANK1 as a susceptibility gene for bipolar disorder.

- Neuropsychopharmacology. 2021;46:1103–1112.
49. Jiang X, Detera-Wadleigh SD, Akula N, Mallon BS, Hou L, Xiao T, et al. Sodium valproate rescues expression of TRANK1 in iPSC-derived neural cells that carry a genetic variant associated with serious mental illness. *Mol Psychiatry*. 2019;24:613–624.
 50. Gambuzza ME, Sofo V, Salmeri FM, Soraci L, Marino S, Bramanti P. Toll-like receptors in Alzheimer's disease: a therapeutic perspective. *CNS Neurol Disord Drug Targets*. 2014;13:1542–1558.
 51. Wolters FJ, Boender J, de Vries PS, Sonneveld MA, Koudstaal PJ, de Maat MP, et al. Von Willebrand factor and ADAMTS13 activity in relation to risk of dementia: a population-based study. *Sci Rep*. 2018;8:5474.
 52. Li QS, De Muynck L. Differentially expressed genes in Alzheimer's disease highlighting the roles of microglia genes including OLR1 and astrocyte gene CDK2AP1. *Brain Behav Immun Health*. 2021;13:100227.
 53. Faux NG, Rembach A, Wiley J, Ellis KA, Ames D, Fowler CJ, et al. An anemia of Alzheimer's disease. *Mol Psychiatry*. 2014;19:1227–1234.
 54. Ramos-Cejudo J, Wisniewski T, Marmar C, Zetterberg H, Blennow K, de Leon MJ, et al. Traumatic Brain Injury and Alzheimer's Disease: The Cerebrovascular Link. *EBioMedicine*. 2018;28:21–30.
 55. Dean M. The Human ATP-Binding Cassette (ABC) Transporter Superfamily. National Center for Biotechnology Information (US); 2002.
 56. Cao QT, Aguiar JA, Tremblay BJ-M, Abbas N, Tiessen N, Revill S, et al. ABCF1 Regulates dsDNA-induced Immune Responses in Human Airway Epithelial Cells. *Front Cell Infect Microbiol*. 2020;10:487.

57. Momtazmanesh S, Perry G, Rezaei N. Toll-like receptors in Alzheimer's disease. *J Neuroimmunol.* 2020;348:577362.
58. Courtney SC, Di H, Stockman BM, Liu H, Scherbik SV, Brinton MA. Identification of novel host cell binding partners of Oas1b, the protein conferring resistance to flavivirus-induced disease in mice. *J Virol.* 2012;86:7953–7963.
59. Vittor AY, Long M, Chakrabarty P, Aycock L, Kollu V, DeKosky ST. West Nile Virus-Induced Neurologic Sequelae-Relationship to Neurodegenerative Cascades and Dementias. *Curr Trop Med Rep.* 2020;7:25–36.
60. de Oliveira J, Kucharska E, Garcez ML, Rodrigues MS, Quevedo J, Moreno-Gonzalez I, et al. Inflammatory Cascade in Alzheimer's Disease Pathogenesis: A Review of Experimental Findings. *Cells.* 2021;10.
61. Kang S, Gim J, Lee J, Gunasekaran TI, Choi KY, Lee JJ, et al. Potential Novel Genes for Late-Onset Alzheimer's Disease in East-Asian Descent Identified by APOE-Stratified Genome-Wide Association Study. *Journal of Alzheimer's Disease.* 2021;82:1451–1460.
62. Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet.* 2006;38:209–213.
63. Vardarajan BN, Reyes-Dumeyer D, Piriz AL, Lantigua RA, Medrano M, Rivera D, et al. Progranulin mutations in clinical and neuropathological Alzheimer's disease. *Alzheimers Dement.* 2022. 9 February 2022. <https://doi.org/10.1002/alz.12567>.
64. Tönjes A, Scholz M, Krüger J, Krause K, Schleinitz D, Kirsten H, et al. Genome-wide meta-analysis identifies novel determinants of circulating serum progranulin. *Hum Mol Genet.* 2018;27.

65. Liu C, Yu J. Genome-Wide Association Studies for Cerebrospinal Fluid Soluble TREM2 in Alzheimer's Disease. *Front Aging Neurosci.* 2019;11.
66. Hou XH, Bi YL, Tan MS, Xu W, Li JQ, Shen XN, et al. Genome-wide association study identifies Alzheimer's risk variant in MS4A6A influencing cerebrospinal fluid sTREM2 levels. *Neurobiol Aging.* 2019;84.
67. Deming Y, Filipello F, Cignarella F, Cantoni C, Hsu S, Mikesell R, et al. The MS4A gene cluster is a key modulator of soluble TREM2 and Alzheimer's disease risk. *Sci Transl Med.* 2019;11.
68. Acosta-Uribe J, Aguillón D, Cochran JN, Giraldo M, Madrigal L, Killingsworth BW, et al. A neurodegenerative disease landscape of rare mutations in Colombia due to founder effects. *Genome Med.* 2022;14:27.
69. Jiao B, Liu X, Tang B, Hou L, Zhou L, Zhang F, et al. Investigation of TREM2, PLD3, and UNC5C variants in patients with Alzheimer's disease from mainland China. *Neurobiol Aging.* 2014;35.
70. Reitz C, Jun G, Naj A, Rajbhandary R, Vardarajan BN, Wang L-S, et al. Variants in the ATP-Binding Cassette Transporter (ABCA7), Apolipoprotein E ϵ 4, and the Risk of Late-Onset Alzheimer Disease in African Americans. *JAMA.* 2013;309:1483–1492.
71. Gianattasio KZ, Prather C, Glymour MM, Ciarleglio A, Power MC. Racial disparities and temporal trends in dementia misdiagnosis risk in the United States. *Alzheimers Dement.* 2019;5:891–898.
72. Beach TG, Monsell SE, Phillips LE, Kukull W. Accuracy of the clinical diagnosis of Alzheimer disease at National Institute on Aging Alzheimer Disease Centers, 2005-2010. *J Neuropathol Exp Neurol.* 2012;71.

73. Jansen IE, Savage JE, Watanabe K, Bryois J, Williams DM, Steinberg S, et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat Genet.* 2019;51:404–413.
74. Jonsson T, Atwal JK, Steinberg S, Snaedal J, Jonsson PV, Bjornsson S, et al. A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. *Nature.* 2012;488:96–99.
75. Sirkis DW, Bonham LW, Johnson TP, La Joie R, Yokoyama JS. Dissecting the clinical heterogeneity of early-onset Alzheimer's disease. *Mol Psychiatry.* 2022;27:2674–2688.
76. Jonathan J Deeks, Julian PT Higgins, Douglas G Altman, editor. Chapter 10: Analysing data and undertaking meta-analyses. *Cochrane Handbook for Systematic Reviews of Interventions*, Cochrane; 2022.
77. Gay NR, Gloude-mans M, Antonio ML, Abell NS, Balliu B, Park Y, et al. Impact of admixture and ancestry on eQTL analysis and GWAS colocalization in GTEx. *Genome Biol.* 2020;21:233.
78. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet.* 2016;48:481–487.
79. King EA, Wade Davis J, Degner JF. Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. *PLoS Genet.* 2019;15:e1008489.
80. Liu B, Gloude-mans M, Rao AS, Ingelsson E, Montgomery SB. Abundant associations with gene expression complicate GWAS follow-up. *Nat Genet.* 2019;51:768.
81. Murphy AE, Schilder BM, Skene NG. MungeSumstats: a Bioconductor package for the

standardization and quality control of many GWAS summary statistics. *Bioinformatics*. 2021;37:4593–4596.

82. Abraham G, Inouye M. Fast principal component analysis of large-scale genome-wide data. *PLoS One*. 2014;9:e93766.
83. Montinaro F, Busby GBJ, Pascali VL, Myers S, Hellenthal G, Capelli C. Unravelling the hidden ancestry of American admixed populations. *Nat Commun*. 2015;6:1–7.
84. Genetic link between Asians and native Americans: evidence from HLA genes and haplotypes. *Hum Immunol*. 2001;62:1001–1008.

Chapter 2 Paper 2

Assessing the lack of diversity in genetics research across neurodegenerative diseases: a systematic review of the GWAS Catalog and literature

Contributing authors: Caroline Jonson, MS*, Kristin S. Levine, MS*, Julie Lake, MS*, Linnea Hertslet, BS, Lietsel Jones, MS, Dhairya Patel, BS, Jeff Kim, MS, Sara Bandres-Ciga, PharmD/PhD, Nancy Terry, MLS, Ignacio F. Mata, PhD, Cornelis Blauwendraat, PhD, Andrew B. Singleton, PhD, Mike A. Nalls, PhD, Jennifer S. Yokoyama, PhD , and Hampton L. Leonard, MS**

*Equal contribution

**Correspondence

2.2.1 Abstract

The under-representation of non-European cohorts in genome-wide association studies (GWAS) is a critical issue with significant implications, including hindering the progress of precision medicine initiatives. This issue is particularly relevant in the context of neurodegenerative diseases (NDDs), where current therapeutic approaches have shown limited success.

We conducted a systematic review of GWAS results and publications up to 2022, focusing on non-European or multi-ancestry neurodegeneration studies. Rigorous article inclusion and quality assessment methods were employed. Out of 123 NDD GWAS reviewed, 82% predominantly featured European ancestry participants. Studies focusing on non-European or multi-ancestry groups identified 52 risk loci, compared to over 90 in European-centric studies. The significant under-representation of non-European ancestries in NDD GWAS hinders comprehensive genetic understanding. It is imperative to address this disparity and foster inclusivity in genetic research to drive significant advancements in diagnosing, treating, and preventing neurodegenerative diseases among diverse global populations.

2.2.1.1 Introduction

The under-representation of non-European cohorts in genome-wide association studies (GWAS) is a critical issue with significant implications, including hindering the progress of precision medicine initiatives. This issue is particularly relevant in the context of neurodegenerative diseases (NDDs), where current therapeutic approaches have shown limited success.

2.2.1.2 Methods

We conducted a systematic review of GWAS results and publications up to 2022, focusing on non-European or multi-ancestry neurodegeneration studies. Rigorous article inclusion and quality assessment methods were employed.

2.2.1.3 Results

Out of 123 NDD GWAS reviewed, 82% predominantly featured European ancestry participants. Studies focusing on non-European or multi-ancestry groups identified 52 risk loci, compared to over 90 in European-centric studies. Encouragingly, 65% of these findings were post-2020, indicating a recent increase in diverse cohort studies.

2.2.1.4 Discussion

The significant under-representation of non-European ancestries in NDD GWAS hinders comprehensive genetic understanding. Prioritizing genomic diversity in future research is crucial for advancing NDD therapies and understanding.

2.2.2 Introduction

Genome-wide association studies (GWAS) have shown a significant bias towards individuals of European ancestry, despite comprising only 16% of the global population [1]. This underrepresentation issue is particularly salient in the realm of neurodegenerative disease (NDD) studies. For instance, while a recent Alzheimer's disease (AD) GWAS including ~800,000 individuals of European descent identified 75 disease-associated loci [2], no GWAS studies on AD currently exist for Admixed American or Native American populations. Similarly, Parkinson's disease research exhibits a glaring imbalance, with black individuals included in just ~4% of published PD studies [3].

The prevalence of NDDs varies significantly among global populations and racial/ethnic groups. This warrants a critical examination of the disparity in genetic research efforts over time. In this manuscript, we present a systematic review spanning from 2012 through 2022, focusing on neurodegenerative disease GWAS research.

Our analysis encompasses common NDDs such as Alzheimer's disease, Parkinson's disease, and amyotrophic lateral sclerosis, as well as less common atypical dementias. Our objective is to quantify the disparity in participant recruitment for genetic studies, shed light on genetic findings in underrepresented populations, and discuss ongoing initiatives aimed at addressing this pervasive issue.

2.2.3 Methods

2.2.3.1 Search Strategy

The systematic review was conducted in two phases. First, we reviewed the GWAS catalog, then, since the GWAS Catalog does not include all GWAS studies, we performed a formal literature

review in collaboration with the National Library of Medicine (NLM). The keywords used in both searches are included in Supplementary Table 1.

Results from both the GWAS Catalog and the NLM search were uploaded to Covidence [4], a web-based software platform, for further review. We removed duplicate studies and any studies published before 2012 or after 2022.

All studies were filtered to only include genome-wide associations examining neurological disease risk factors, family history of disease, disease progression, age at onset, or survival genome-wide, excluding exome wide studies and those that focused on a targeted set of SNPs or genetic loci. Studies investigating disease subtypes, biomarkers, non-English language studies, and those investigating only rare or structural variation were also excluded.

Studies were assessed for eligibility by two independent reviewers and all conflicts were resolved by a third independent reviewer. Publication date, phenotype, and cohort information were extracted from each publication. If multiple phenotypes of interest were analyzed in the same study, information was included in both phenotype categories. The number of samples per ancestry was extracted manually from each study. We looked at seven ancestry groupings: European (EUR), East Asian (EAS), Middle-Eastern (MDE), African (AFR), African American and Caribbean (AAC), Latino and indigenous Americas populations (AMR), and South Asian (SAS). A PRISMA diagram of our filtering process can be found in **Figure 2.2.1**. All 123 studies passing our filters can be found in Supplementary Table 2.

Finally, results were examined manually for all studies passing implemented filtering methods. Novel loci discovered in non-European or multi-ancestry populations, with a p-value below $5E-8$, are listed in **Table 2.2.2**.

2.2.4 Results

2.2.4.1 Search results

We identified 123 eligible GWAS studies. Unsurprisingly, we found that European populations were overrepresented in GWAS pertaining to NDDs (**Figure 2.2.2**). When non-European populations were included, the sample sizes were on average 15X smaller than the European ancestry samples included in the same disease category. The underrepresentation of non-European populations was particularly evident among the less common NDDs, including Lewy body dementia and frontotemporal dementia, where we did not identify any non-European or multi-ancestry GWAS studies using the outlined search methods. We have summarized the lack of diversity in genetic studies of NDDs in **Table 2.2.1**, and **Figure 2.2.2**.

We found 52 novel NDD loci that were identified in non-European or multi-ancestry populations (Table 2.2.2). Of these 52 loci, 28 were found in multi-ancestry studies, 21 were found in East Asian studies, and only 3 were found in other populations (AAC and AMR). No loci were discovered in AFR, MDE, or SAS ancestries. Recent studies that combined individuals of multiple ancestries by using standard random-effects and some custom meta-analytic techniques [5] have succeeded in identifying novel disease loci that reach genome-wide significance, including two novel AD loci [6] and 12 novel PD loci [7]. However, these studies leverage existing European sample sizes as a backbone for much of the statistical power needed for discovery.

In the following sections, we briefly summarize the results of our systematic review in a disease specific manner. Many of these findings have not been replicated. As datasets become larger and more inclusive, the genetic architecture of these diseases may grow and change.

2.2.4.2 Alzheimer's Disease

Largest European GWAS: Wightman 2021

Total samples: 1,126,563

Largest multi-ancestry GWAS: Lake 2022

Total samples: 644,188

Total non-European samples: 18,246

% non-European: 2.83%

Largest non-European GWAS: Sherva 2022

Ancestry: AAC

Total samples: 75,058

The largest Alzheimer's disease GWAS of European populations included ~1.1 million individuals and identified a total of 38 associated loci [8]. Another recent GWAS included ~800,000 individuals of European ancestry and identified a total of 75 loci [2]. The discrepancy between identified loci in these studies could be due to many factors, including differences in neuropathological/diagnostic criteria [9].

A 2013 GWAS conducted in African Americans replicated an association at *ABCA7* previously identified in European populations. They found that rs115550680, rare in European populations, was associated with an increased risk for AD in African Americans comparable to the highly pathogenic APOE- ϵ 4 variant observed in Europeans [10]. A 2017 GWAS in African Americans identified two novel loci at *COBL* and *SLC10A2* [11]. The most extensive African American GWAS to date, drawing from a military cohort of around 22,000 individuals and a proxy GWAS involving approximately 50,000 individuals, identified significant associations with established

AD risk genes such as *TREM2*, *CD2AP*, and *ABCA7*. Notably, distinct lead variants were observed in these loci compared to those found in European cohorts [12].

The only study conducted in Caribbean Hispanic individuals was a 2017 study with 2,451 cases and 2,063 controls. They found a novel and population specific locus near *FBXL7* [13]. The lead SNP, rs75002042, is much more common in individuals with African ancestry compared to individuals of European ancestry, with minor allele frequencies around 20% and 0.009%, respectively. This study also replicated six loci previously reported in European populations, including *FRMD4A*, *CELF1*, *FERMT2*, *SLC24A4-RIN3*, *ABCA7*, and *CD33* [13].

The largest AD study in East Asian populations was conducted in Japanese participants with 1,827 cases and 15,204 controls (discovery + replication) [14], but they did not nominate any genome-wide significant loci. More recent but smaller studies have since been conducted, including a 2021 GWAS in a Chinese cohort that reported four novel loci near *RHOBTB3/GLRX*, *CTC-278L1.1*, *CTD-2506J14.1*, and *CHODL* [15], a study in Japanese participants that nominated a locus in *FAM47E* [16], and a study including both Korean and Japanese participants that nominated two novel loci at *CACNA1A* and *LRIG* [17].

Multi-ancestry studies have nominated additional AD loci, however these studies still rely on Europeans as the majority population. *SORL1* was first identified as a risk locus for AD in a GWAS that included East Asian and European ancestry populations [18]. Other multi-ancestry GWAS identified *OR2B2* [16], *TRANK1* and *VWA5B2* [19] as novel loci for AD.

While the inclusion of diverse populations in genetic research for AD is arguably better than what is seen for some of the atypical dementias, the largest study size for a non-European population [12] was still only 7% of the total sample size for the largest European AD GWAS.

2.2.4.3 Parkinson's disease

Largest European GWAS: Nalls 2019

Total samples: 1,456,306

Largest multi-ancestry GWAS: Kim 2022

Total samples: 2,525,730

Total non-European samples: 962,735

% non-European: 38.12%

Largest non-European GWAS: Foo 2017

Ancestry: EAS

Total samples: 14,006

The largest meta-GWAS of PD risk in individuals of European ancestry found 90 significant risk signals across 78 genomic regions. The 90 nominated risk variants collectively explain roughly 16-36% of the heritable risk of non-monogenic, or complex PD [20].

The largest study in East Asian populations (with exception to a study done in Japan before our review period [21]) was conducted with Han Chinese participants, replicating loci previously identified in European populations including *SNCA*, *LRRK2* and *MCCCI* in their discovery GWAS of 14,006 participants [22]. More recent studies in Chinese populations have nominated a locus on *NDN/PWRN4* associated with age at onset and a locus on *RPL3* associated with reduced survival [23, 24].

The first and most recent PD GWAS of a South American population was conducted in 2021, replicating an association at *SNCA* with 1,497 participants [25].

Recently, more multi-ancestry studies have been conducted in PD, nominating novel loci for disease risk and age at onset including *ITGA8*, *SV2C*, and *BST1* [26,27,28]. The largest meta-

GWAS for PD, which included 4 ancestral populations, identified 12 novel loci: *MTF2*, *RP11-360P21.2*, *ADD1*, *SYBU*, *IRS2*, *USP8:RP11-562A8.5*, *PIGL*, *FASN*, *MYLK2*, *AJ006998.2*, *Y_RNA*, and *PPP6R2* [7].

The largest non-European PD GWAS was in East Asian populations, however, only a few novel loci have been nominated in that ancestry. Multi-ancestry studies have nominated more novel variants in recent studies, but much more work is needed to better understand risk for PD in non-European populations.

2.2.4.4 Amyotrophic lateral sclerosis

Largest European GWAS: van Rheenen 2016

Total samples: 41,398

Largest multi-ancestry GWAS: van Rheenen 2021

Total samples: 152,268

Total non-European samples: 18,266

% non-European: 12.00%

Largest non-European GWAS: Wei 2019

Ancestry: EAS

Total samples: 4,727

ALS GWAS in European populations have nominated a number of risk loci including *C9ORF72*, *UNC13A*, *C21orf2*, *SARM1*, *MOBP*, *SCFD1*, *TBKK1*, and *KIF5A* [29, 30].

ALS is the only disease in our review where more genome-wide significant novel loci have been identified in a non-European population than in the largest European-only study. The first GWAS of individuals with Chinese Han ancestry identified *CAMK1G* and *CABIN1/SUSD2* as susceptibility loci for ALS [31]. Later studies in the Han Chinese population nominated additional novel loci including *INPP5B*, *IQCF5/IQCF1*, *ITGA9*, *PFKP*, *MYO18B*, *ALCAM*, *OPCML*, *GPRI33*, *TYW/CRYZ* and *FGD4* [32, 33]. With a total of 12 genome-wide significant loci, East Asian ancestry GWAS for ALS have nominated the most of any single non-European population covered in our review.

Multi-ancestry GWAS for ALS, which typically consist of European and East Asian ancestry populations, have been successful at nominating additional risk loci including *GPX3-TNIP1* and *ACSL5* [30, 34, 35]. The largest ALS GWAS to date was a multi-ancestry study including over

150,000 individuals of European and East Asian ancestry. This study identified a total of 15 risk loci for ALS, replicating 8 previously-identified and nominating 7 novel loci: *SOD1*, *HLA*, *SLC9A8-SPATA2*, *ERGIC1*, *NEK1*, *COG3*, and *PTPRN2* [36].

Similar to PD, ALS GWAS including or focused on East Asian populations have made more progress than other non-European populations for these diseases. However, much more work is still needed in all populations to progress potential precision medicine initiatives for ALS.

2.2.4.5 Multiple Sclerosis

Largest European GWAS: Patsopoulos 2019

Total samples: 115,803

There are no multi-ancestry studies in MS.

Largest non-European GWAS: Isobe 2015

Ancestry: AAC

Total samples: 2,319

The largest GWAS meta-analysis for MS included 115,803 individuals of European ancestry and found 82 significant genome-wide associations with MS. This study was also the first to identify a risk locus for MS on chromosome X and the identified genetic markers accounted for nearly 50% of the hereditary risk for MS [37].

Studies in non-European populations were more limited in MS than in the previous diseases discussed. The largest GWAS in African Americans was successful at replicating 21 of the loci previously identified in European populations, but did not nominate any new risk loci at a genome-wide significant level [38]. The only other study nominated by our review process for MS in non-European populations was conducted in a Mexican population. The study found 4 significant variants, however, these variants had limited regional support and the study was severely

underpowered with only 29 cases and 132 controls [39]. Due to these limitations, we concluded that the variants identified in this study could not be classified as novel.

Sample sizes for MS GWAS are still relatively small, even for European populations. In addition, we did not find any multi-ancestry studies through our search methods, highlighting a potential opportunity for further discovery for this disease.

2.2.4.6 Frontotemporal dementia

Largest European GWAS: Ferrari 2014

Total samples: 12,928

There are no multi-ancestry or non-European studies in FTD.

Common risk loci nominated by previous European FTD studies include *C9ORF72*, *GRN*, and *MAPT* [40]. The largest FTD GWAS in our review date range included ~13,000 participants of European ancestry and nominated an additional locus in the *HLA-DRA/HLA-DRB5* region [41]. This study was conducted in 2014, and while more recent GWAS of FTD have been performed, none have surpassed the sample size from the Ferrari study, and many have focused on smaller FTD subtypes [42, 43].

No non-European or multi-ancestry GWAS were identified in our systematic review for FTD. Investigation of known genetic risk factors in non-Europeans suggest that *C9ORF72* expansions may be quite rare in Chinese populations [44] highlighting the need for further research in this area.

2.2.4.7 Myasthenia Gravis

Largest European GWAS: Sakaue 2021

Total samples: 355,142

Largest Multi-ancestry GWAS: Sakaue 2021

Total samples: 533,853

Total non-European samples: 178,711

% non-European: 33.47%

Largest non-European GWAS: Na 2014

Ancestry: EAS

Total samples: 259

Known loci for MG include *PTPN22*, *CTLA4*, *HLA-DQA1*, *ZBTB10*, and *TNFRSF11A* [45, 46], all nominated in European-based GWAS. The most recent GWAS for MG nominated an additional loci at *CHRNA1*, *SFMBT2*, and *FAM76B*, although the latter two did not replicate [47]. The largest European and multi-ancestry GWAS for MG to date were both performed in the same study, leveraging 533,853 total samples from Japanese, UK, and Finnish-based biobanks. However, with only 278 cases, the effective sample size ($4/(1/n_{\text{case}}+1/n_{\text{control}})$) for the meta-analysis was insufficiently powered and they did not nominate any new loci for MG [48].

In non-European populations, the literature review identified one Korean GWAS for MG. However, this study was small and did not identify any loci meeting genome-wide significance [49]. Other studies have found that there is earlier onset of MG in Asian populations, and higher prevalence of the ocular form in Asian children, highlighting the importance of continued discovery efforts for MG in non-European populations [50].

2.2.4.8 Lewy body dementia

Largest European GWAS: Chia 2021

Total samples: 7,372

There are no multi-ancestry or non-European studies in LBD.

Previously nominated risk loci for LBD include *GBA*, *APOE*, and *SNCA* [51, 52]. LBD can be hard to diagnose as there are a number of clinical and genetic overlaps with AD and PD, which may be one of the reasons why there is still limited genetic research for LBD in both European and non-European populations [51, 53].

We found no LBD GWAS in any single non-European ancestry populations or any multi-ancestry studies through our search methods. Concrete data on the prevalence of LBD in diverse ancestries is difficult to acquire, showing a potential opportunity for valuable future research.

2.2.4.9 Vascular dementia

Largest European GWAS: Moreno-Grau 2019

Total samples: 4,830

Largest Multi-ancestry GWAS: Fongang 2022

Total samples: 482,088

Total non-European samples: 11,590

% non-European: 2.40%

There are no non-European studies in VaD.

Despite an approximated prevalence of about 15-20% in all dementia cases [54], vascular dementia remains difficult to study because of the uncertainty of diagnosis. In fact, only two studies on vascular dementia (VaD) passed our criteria and only one of these found genome-wide significant novel loci. The first study was a European GWAS that looked at vascular, mixed, and pure AD phenotypes and nominated loci at *ANKRD31* and *NDUFA6* [55]. The second study that passed our criteria was a multi-ancestry GWAS for all-cause and vascular dementia including participants from European, African, Asian, and Hispanic ancestries, but did not find any significant novel loci [56].

VaD prevalence and risk appears to be higher in South Asian ancestries compared to European or Chinese populations [57, 58]. Additional studies have suggested that African Americans are most likely to be admitted to inpatient care with a VaD primary diagnosis [59]. Despite these findings, there are still limited genetic studies for VaD, and we found no single non-European GWAS, highlighting the need for future research.

2.2.5 Discussion

This review highlights the lack of ancestral diversity in genetic research across neurodegenerative disease GWAS over the past decade. Current research suggests that including non-European populations can improve our understanding of the genetic architecture of disease through novel ancestry-specific discoveries, increased statistical power awarded by studying diverse haplotype structures, and the identification of loci with heterogeneous effects across populations [63].

Additionally, while we looked at seven genetic ancestry groups, these “buckets” do not capture the true diversity of global populations. The African continent is known to have high genetic diversity, yet individuals of African ancestry are routinely grouped into a single category [64]. In fact, we found no studies investigating South Asian (SAS) or continental African (AFR) populations. After investigating the cohorts in our review, we noted that though there were multiple “African” labeled studies, none of them directly investigated individuals in *continental* Africa, instead looking at African-American or other African-admixed populations. It is critical to mention the reference population used to define the specific population, to prevent the misattribution of genetic features across ancestries. Grouping all participants with any African ancestry into a generalized African category obscures the significant issue of inadequate representation of continental Africans.

In addition, there has been very little research done on admixed populations, and how the combinations of different ancestries affect SNP frequencies and/or gene expression. A GWAS in a Caribbean Hispanic admixed population found that the frequency of a novel locus spanning *FBXL7* varied greatly, from 1% in those with European ancestry to 20% in African Americans [13].

Furthermore, previous research has shown that the transferability of polygenic risk scores from African Americans to various African populations is highly unreliable [65]. The substantial genetic and environmental disparities among individuals of African descent underscore the urgent need to improve diversity in genetic studies.

2.2.5.1 Diversity in SNP discovery

While 6 of the 8 NDDs we investigated had non-European representation, only PD had >1,000 cases and >30% non-European samples (Table 2.2.1). Additionally, no new significant loci have been identified in diverse population studies for MS, LBD, FTD, VaD, or MG. While the largest non-European cohort in MG included almost 180,000 samples, only 81 MG cases were included. A GWAS with less than 1,000 cases is unlikely to achieve sufficient statistical power for SNP discovery in polygenic diseases where multiple loci with small effect sizes are generally expected [60, 62]. Alzheimer's, the most well-funded of all the NDDs, has less than 3% diversity among cases in genetic studies. The incorporation of studies that lack statistical power and replicability minimizes the true imbalance between European and non-European studies, maintaining a Eurocentric bias.

In addition, many genetic association studies in East Asian populations did not meet our review criteria because they were not conducted on a genome-wide scale. Instead, these studies often investigated only one or a small group of SNPs that had been previously associated with disease

in European populations, potentially missing associations that are specific to non-European populations.

In fact, many loci identified in European populations have heterogeneous effects or ancestry-specific SNP associations. For example, while the *APOE* alleles account for around a quarter of overall heritability for AD in Europeans [8, 16], several studies suggest that the *APOE4* allele has a weaker effect in African ancestry [66, 67] and Caribbean Hispanic [68] populations. The effect has been found to be greater in Japanese populations [66, 67]. Heterogeneity of risk at *APOE4* has been quantified in a recent multi-ancestry meta-analysis, with an I² up to 85%, with 50% or more of that risk heterogeneity attributable to genetic ancestry differences [6]. We believe that examination of local ancestry at loci with such global differences may help discern whether locus-specific inheritance patterns modulate disease risk.

Similarly, *C9ORF72* is one of the most common risk factors for ALS. However, the frequency of the *C9ORF72* expansion is lower in Chinese populations (0.3%) as compared to European populations (7%) [30]. Recent research suggests that commonly used genetic tests to diagnose ALS may be less accurate in non-European ancestry patients because they are less likely to carry the *C9ORF72* structural variant [69].

Some SNPs with large effect sizes don't exist or are extremely rare in certain ancestry groups. Variants in *ABCA7*, for example, increase AD risk more in individuals of African ancestry than in those of European ancestry [70]. In fact, *ABCA7* has a comparable effect size to *APOE* in individuals of African ancestry [10]. Genetic variants in *LRRK2*, *GBA*, and *SNCA*, which have been associated with increased risk of PD in European ancestry populations, appear to have a negligible effect in individuals from India [71, 72, 73, 74]. Without studying diverse populations, researchers would miss the population-specific effects of these loci and potential therapeutic targets which modify their effects.

2.2.5.2 Looking forward

Despite the inequalities highlighted above, progress is being made. Researchers in AD are taking a strong multi-modal approach to increasing diversity. The Multi-Partner Consortium to Expand Dementia Research in Latin America (ReDLat) is leveraging “on the ground” connections with research communities in Latin America and the Caribbean to grow a diverse database of dementia resources [75,76]. The Alzheimer’s Disease NeuroImaging Study (ADNI) is growing more inclusive cell lines and generating partner data for multiple ancestrally diverse samples [77]. The NIH’s Center for Alzheimer’s and Related Dementias (CARD) is filling diversity gaps by creating training materials, generating data to complement existing efforts, and providing open science support for researchers in diverse communities.

Multiple efforts are also underway in the PD space. The Genetic Architecture of Parkinson disease in India (GAP-India) plans to develop a large clinical/genomic biobank in India [71]. The Latin American Research Consortium on the Genetics of PD (LARGE-PD) aims to address inclusivity and genomic differences within and across Latino populations. Finally, the Global Parkinson’s Genetics Program (GP2) aims to genotype >150,000 individuals from around the world. GP2-funded projects include the Black and African Americans Connections to Parkinson’s Disease Study (BLAACPD), which seeks to assess the genetic architecture of Black and African American individuals with Parkinson’s disease, as well as healthy subjects, from across the United States [78]. GP2 is motivated to increase diversity not only just among samples recruited into studies, but also in the investigators making use of the data, providing training and resources to ensure that all researchers are on an open and equal field of play [79].

A list of ongoing efforts for increasing diversity in NDD genetic research, including atypical dementias, can be found in **Supplemental Table 3**.

These efforts are paying off. More than 65% of the neurodegenerative disease-associated loci discovered in non-European or multi-ancestry populations were identified in the period between 2020 to 2022 (**Figure 2.2.3**). Over the past 10 years, there has been a steady increase in the proportion of non-European samples included in genetic studies (**Figure 2.2.2B**). With the increase in diverse samples in recent years, there has also been a growing interest in the use of multi-ancestry analyses to discover, fine-map and assess heterogeneity at disease risk loci, particularly in AD, PD, and ALS (**Figure 2.2.2A**).

In fact, we are already seeing the benefits of increased diversity on genetic discovery in NDD research. A 2023 GWAS using African and African American samples collected by GP2 and 23andMe, and co-lead by researchers in Nigeria and NIH, found a novel *GBA1* locus that is rare in other populations [82]. We anticipate that in the future, leveraging multiple ancestries will continue to improve fine-mapping resolution to prioritize causal variants [5], increase access to and reduce bias in precision medicine practices such as polygenic risk prediction [1], and drive many new discoveries in the genetics of NDDs.

2.2.6 Conclusion

Our systematic review highlights a striking disparity in the representation of diverse genetic ancestry populations in NDD research, emphasizing the urgent need for greater inclusivity to advance our understanding of these complex conditions and develop more equitable precision medicine approaches. Efforts to bridge this gap and promote diversity in genetic studies are vital for achieving meaningful progress in the diagnosis, treatment, and prevention of NDDs across global populations.

2.2.7 Figures and Tables

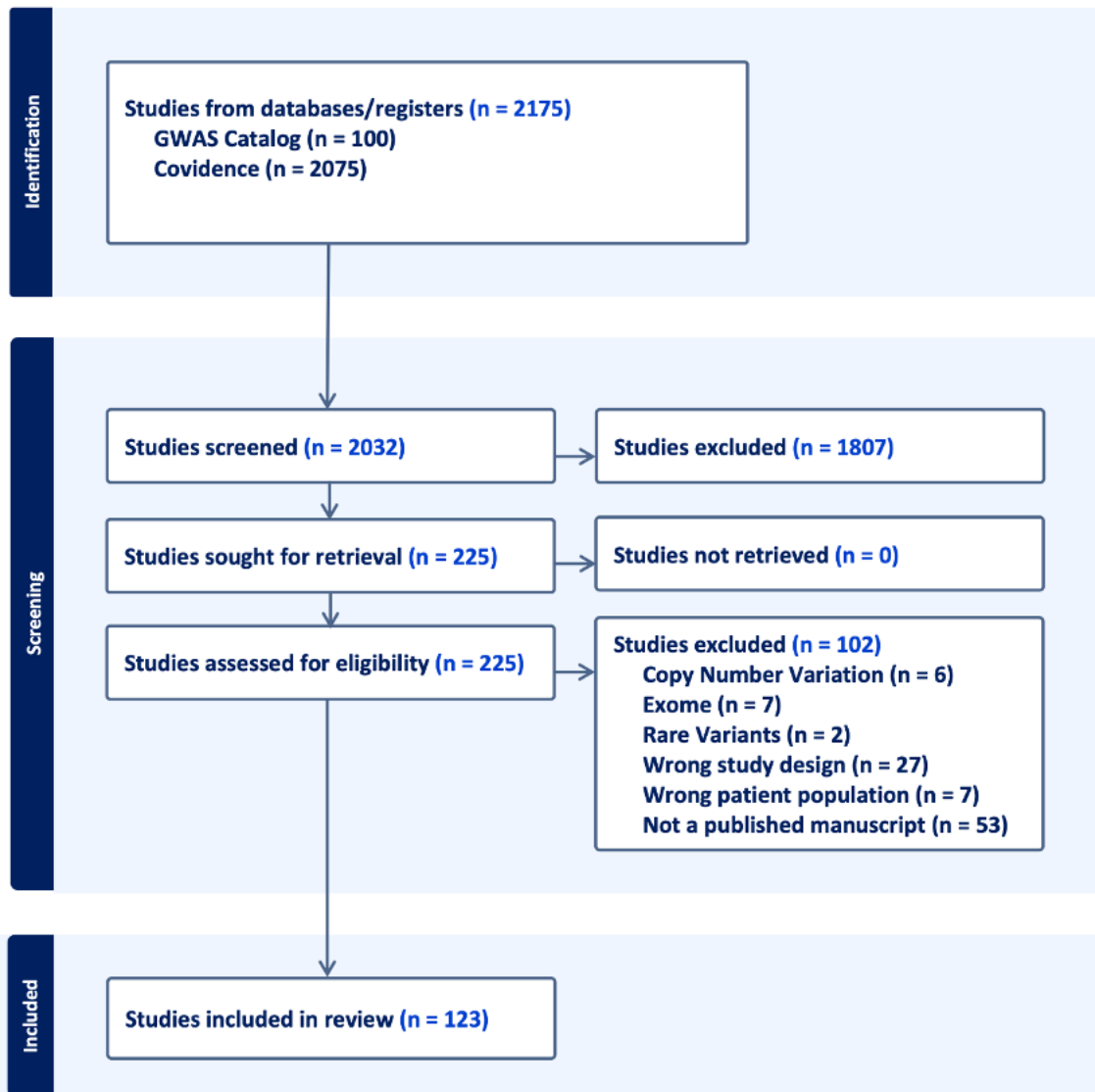


Figure 2.1.1 PRISMA flow diagram for systematic review

Diagram of published GWAS studies and pubmed literature review from NLM. N in the figure relates to the number of published studies as of April 28th, 2022.



Figure 2.2.2 Number of studies over time from 2012 to 2022
 A) Bar plot of study counts by NDD (left) with cumulative counts for each ancestry (right) . Data in this figure includes both single and multi ancestry studies. B) Time series of the annual study counts in European and Non-European populations from 2012 to 2022. The slope from a linear regression is also displayed to highlight the rate of change in the number of study counts over time.

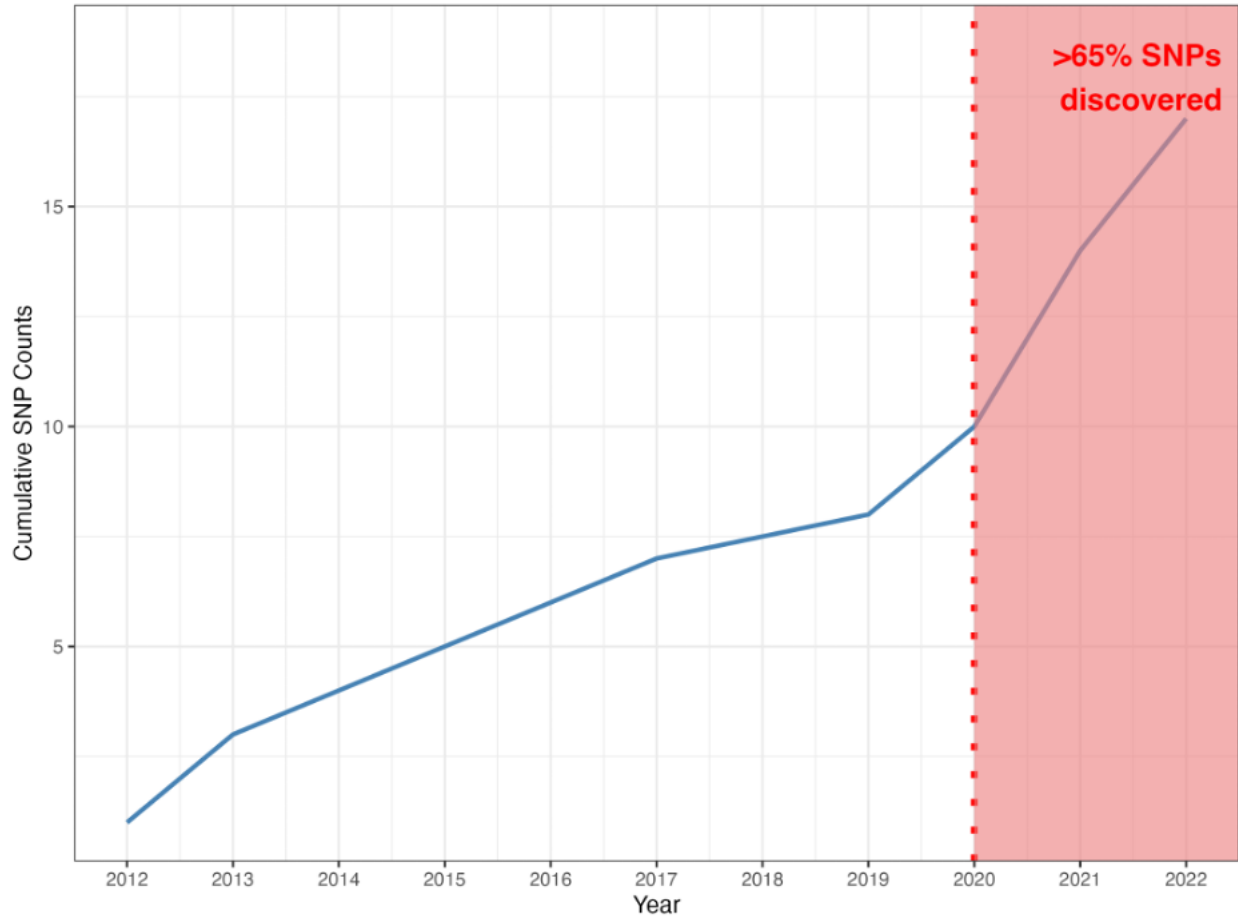


Figure 2.2.3 Cumulative count of discovered SNPs from 2012 through 2022
Notably, more than 65% of the SNPs were identified in the period between 2020 through 2022.

Table 2.2.1 Largest GWAS sample size by NDD and ancestry for single and multi ancestry studies.

NDD	Ancestry	Author	Year	Total samples (% non-European*)	Study
AD	AAC	Sherva	2022	75,058	African ancestry GWAS of dementia in a large military cohort identifies significant risk loci
	EAS	Hirano	2015	17,031	A genome-wide association study of late-onset Alzheimer's disease in a Japanese population
	EUR	Wightman	2021	1,126,563	A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease
	MULTI	Lake	2022	644,188 (2.83%)	Multi-ancestry meta-analysis and fine-mapping in Alzheimer's Disease
	AMR,MDE	NO STUDIES			
PD	AMR	Loesch	2021	1,497	Characterizing the Genetic Architecture of Parkinson's Disease in Latinos
	EAS	Foo	2017	14,006	Genome-wide association study of Parkinson's disease in East Asians
	EUR	Nalls	2019	1,474,097	Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies
	MULTI	Kim	2022	2,525,730 (38.12%)	Multi-ancestry genome-wide meta-analysis in Parkinson's disease
	AAC,MDE	NO STUDIES			
ALS	EAS	Wei	2019	4,727	Identification of TYW3/CRYZ and FGD4 as susceptibility genes for amyotrophic lateral sclerosis
	EUR	vanRheenen	2016	41,398	Genome-wide association analyses identify new risk variants and the genetic architecture of amyotrophic lateral sclerosis
	MULTI	van Rheenen	2021	152,268 (12.00%)	Common and rare variant association analyses in amyotrophic lateral sclerosis identify 15 risk loci with distinct genetic architectures and neuron-specific biology
	AAC, AMR, MDE	NO STUDIES			

NDD	Ancestry	Author	Year	Total samples (% non-European*)	Study
MS	AAC	Isobe	2015	2319	An ImmunoChip study of multiple sclerosis risk in African Americans
	AMR	Ordoñez	2015	161	Genomewide admixture study in Mexican Mestizos with multiple sclerosis
	EUR	Patsopoulos	2019	115,803	Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility
	EAS,MDE, MULTI	NO STUDIES			
FTD	EUR	Ferrari	2014	12,928	Frontotemporal dementia and its subtypes: a genome-wide association study
	AAC, AMR, EAS, MDE, MULTI	NO STUDIES			
MG	EAS	Na	2014	259	Whole-genome analysis in Korean patients with autoimmune myasthenia gravis
	EUR	Chia	2022	45,675	Identification of genetic risk loci and prioritization of genes and pathways for myasthenia gravis: a genome-wide association study
	MULTI	Sakaue	2021	533,853 (33.48 %)	A cross-population atlas of genetic associations for 220 human phenotypes
	AAC,AMR, MDE	NO STUDIES			
LBD	EUR	Chia	2021	7,372	Genome sequencing analysis identifies new loci associated with Lewy body dementia and provides insights into its genetic architecture
	AAC,AMR,E AS,MDE, MULTI	NO STUDIES			

NDD	Ancestry	Author	Year	Total samples (% non-European*)	Study
VaD	EUR	Moreno-Grau	2019	4,830	Genome-wide association analysis of dementia and its clinical endophenotypes reveal novel loci associated with Alzheimer's disease and three causality networks: The GR@ACE project
	MULTI	Fongang	2022	482,088 (2.40%)	A meta-analysis of genome-wide association studies identifies new genetic loci associated with all-cause and vascular dementia
	AAC, AMR,EAS, MDE	NO STUDIES			

*if applicable

Table 2.2.2 Genome-wide significant novel loci nominated in non-European populations
 We found no diverse or multi-ancestry loci for LBD, FTD, MG, or VaD. P-values are given in parentheses. Nominated loci were determined as the nearest gene or genomic context within 1MB of the significant SNP (P-value < 5E-8).

Disease	Ancestry	Nominated loci
AD	AAC	<i>COBL</i> (3.8E-8), <i>SLC10A2</i> (4.6E-8) ¹¹
	EAS	<i>RHOBTB3/GLRX</i> (3.07E-19), <i>CTC-278L1.1</i> (2.49E-23), <i>CTD-2506J14.1</i> (1.35E-67), <i>CHODL</i> (4.81E-9) ¹⁵ , <i>FAM47E</i> (5.34E-9) ¹⁶ , <i>CACNA1A</i> (2.49E-8), <i>LRIG1</i> (1.51E-8) ¹⁷
	AMR	<i>FBXL7</i> (6.19E-09) ¹³
	MULTI	<i>OR2B2</i> (2.14E-8) ¹⁶ , <i>SORL1</i> (1.04E-8) ¹⁸ , <i>TRANK1</i> (3.49E-8), <i>VWA5B2</i> (3.75E-8) ¹⁹
PD	EAS	<i>NDN/PWRN4</i> (3.14E-9) ²³ , <i>RPL3</i> (2.72E-8) ²⁴
	MULTI	<i>ITGA8</i> (1.3E-8) ²⁶ , <i>SV2C</i> (1.17E-10) ²⁷ , <i>MTF2</i> (1.15E-10), <i>RP11-360P21.2</i> (1.65E-10), <i>ADD1</i> (4.11E-9), <i>SYBU</i> (3.62E-9), <i>IRS2</i> (2.30E-9), <i>USP8:RP11-562A8.5</i> (6.45E-10), <i>PIGL</i> (2.93E-9), <i>FASN</i> (2.61E-9), <i>MYLK2</i> (3.86E-9), <i>AJ006998.2</i> (1.12E-9), <i>Y RNA</i> (3.81E-9), <i>PPP6R2</i> (4.09E-10) ⁷ , <i>BST1</i> (4.41E-8) ²⁸
ALS	EAS	<i>CAMKIG</i> (2.92E-8), <i>CABIN1/SUSD2</i> (2.35E-9) ³¹ , <i>INPP5B</i> (2.24E-8), <i>IQCF5/IQCF1</i> (2.06E-9), <i>ITGA9</i> (2.55E-8), <i>PFKP</i> (2.46E-9), <i>MYO18B</i> (2.28E-10), <i>ALCAM</i> (4.00E-8), <i>OPCML</i> (8.43E-9), <i>GPR133</i> (8.45E-10) ³² , <i>TYW/CRYZ</i> (2.10E-14), <i>FGD4</i> (5.19E-9) ³³
	MULTI	<i>GPX3-TNIP1</i> (1.3E-8) ³⁰ , <i>ACSL5-ZDHHC6</i> (8.3E-9) ³⁴ , <i>SOD1</i> (3.5E-18), <i>HLA</i> (3.5E-12), <i>SLC9A8-SPATA2</i> (3.2E-10), <i>ERGIC1</i> (5.6E-9), <i>NEK1</i> (6.9E-9), <i>COG3</i> (1.2E-8), and <i>PTPRN2</i> (1.8E-8) ³⁶

2.2.8 Supplemental Online Content

All supplemental online content can be downloaded directly from the preprint material at MedRxiv.

2.2.9 References

1. Martin, A. R. et al. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* 51, 584–591 (2019).
2. Bellenguez, C. et al. New insights into the genetic etiology of Alzheimer’s disease and related dementias. *Nat. Genet.* 1–25 (2022).
3. Schumacher-Schuh, A. F. et al. Diversity in Parkinson’s disease genetics research: current landscape and future directions. *bioRxiv* (2021) doi:10.1101/2021.12.07.21266995.
3. Innovation, V. H. Covidence systematic review software. Preprint at (2019).
5. Mägi, R. et al. Trans-ethnic meta-regression of genome-wide association studies accounting for ancestry increases power for discovery and improves fine-mapping resolution. *Hum. Mol. Genet.* 26, 3639–3650 (2017).
6. Lake, J. et al. Multi-ancestry meta-analysis and fine-mapping in Alzheimer’s disease. *Mol. Psychiatry* (2023) doi:10.1038/s41380-023-02089-w.
7. Kim, J. J. et al. Multi-ancestry genome-wide meta-analysis in Parkinson’s disease. *bioRxiv* (2022) doi:10.1101/2022.08.04.22278432.
8. Wightman, D. P. et al. A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer’s disease. *Nat. Genet.* 53, 1276–1282 (2021).
9. Beach, T. G., Monsell, S. E., Phillips, L. E. & Kukull, W. Accuracy of the clinical diagnosis of Alzheimer disease at National Institute on Aging Alzheimer Disease Centers, 2005–2010. *J. Neuropathol. Exp. Neurol.* 71, 266–273 (2012).
10. Reitz, C. et al. Variants in the ATP-binding cassette transporter (ABCA7), apolipoprotein E ϵ 4, and the risk of late-onset Alzheimer disease in African Americans. *JAMA* 309, 1483–1492 (2013).

11. Mez, J. et al. Two novel loci, COBL and SLC10A2, for Alzheimer's disease in African Americans. *Alzheimers. Dement.* 13, 119–129 (2017).
12. Sherva, R. et al. African ancestry GWAS of dementia in a large military cohort identifies significant risk loci. *bioRxiv* (2022) doi:10.1101/2022.05.25.22275553.
13. Tosto, G. et al. F-box/ LRR -repeat protein 7 is genetically associated with Alzheimer's disease. *Ann. Clin. Transl. Neurol.* 2, 810–820 (2015).
14. Hirano, A. et al. A genome-wide association study of late-onset Alzheimer's disease in a Japanese population. *Psychiatr. Genet.* 25, 139–146 (2015).
15. Jia, L. et al. Prediction of Alzheimer's disease using multi-variants from a Chinese genome-wide association study. *Brain* 144, 924–937 (2021).
16. Shigemizu, D. et al. Ethnic and trans-ethnic genome-wide association studies identify new loci influencing Japanese Alzheimer's disease risk. *Transl. Psychiatry* 11, 151 (2021).
17. Kang, S. et al. Potential Novel Genes for Late-Onset Alzheimer's Disease in East-Asian Descent Identified by APOE-Stratified Genome-Wide Association Study. *J. Alzheimers. Dis.* 82, 1451–1460 (2021).
18. Miyashita, A. et al. SORL1 is genetically associated with late-onset Alzheimer's disease in Japanese, Koreans and Caucasians. *PLoS One* 8, e58618 (2013).
19. Lake, J. et al. Multi-ancestry meta-analysis and fine-mapping in Alzheimer's Disease. *bioRxiv* (2022) doi:10.1101/2022.08.04.22278442.
20. Nalls, M. A. et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet Neurol.* 18, 1091–1102 (2019).
21. Satake, W. et al. Genome-wide association study identifies common variants at four loci

- as genetic risk factors for Parkinson's disease. *Nat. Genet.* 41, 1303–1307 (2009).
22. Foo, J. N. et al. Genome-wide association study of Parkinson's disease in East Asians. *Hum. Mol. Genet.* 26, 226–232 (2017).
 23. Li, C. et al. Genetic Modifiers of Age at Onset for Parkinson's Disease in Asians: A Genome-Wide Association Study. *Mov. Disord.* 36, 2077–2084 (2021).
 24. Li, C. et al. Genetic Determinants of Survival in Parkinson's Disease in the Asian Population. *Mov. Disord.* (2022) doi:10.1002/mds.29069.
 25. Loesch, D. P. et al. Characterizing the Genetic Architecture of Parkinson's Disease in Latinos. *Ann. Neurol.* 90, 353–365 (2021).
 26. Lill, C. M. et al. Comprehensive research synopsis and systematic meta-analyses in Parkinson's disease genetics: The PDGene database. *PLoS Genet.* 8, e1002548 (2012).
 27. Foo, J. N. et al. Identification of Risk Loci for Parkinson Disease in Asians and Comparison of Risk Between Asians and Europeans: A Genome-Wide Association Study. *JAMA Neurol.* 77, 746–754 (2020).
 28. Grover, S. et al. Genome-wide Association and Meta-analysis of Age at Onset in Parkinson Disease: Evidence From the COURAGE-PD Consortium. *Neurology* 99, e698–e710 (2022).
 29. Nicolas, A. et al. Genome-wide Analyses Identify KIF5A as a Novel ALS Gene. *Neuron* 97, 1268–1283.e6 (2018).
 30. Benyamin, B. et al. Cross-ethnic meta-analysis identifies association of the GPX3-TNIP1 locus with amyotrophic lateral sclerosis. *Nat. Commun.* 8, 611 (2017).
 31. Deng, M. et al. Genome-wide association analyses in Han Chinese identify two new susceptibility loci for amyotrophic lateral sclerosis. *Nat. Genet.* 45, 697–700 (2013).

32. Xie, T. et al. Genome-wide association study combining pathway analysis for typical sporadic amyotrophic lateral sclerosis in Chinese Han populations. *Neurobiol. Aging* 35, 1778.e9–1778.e23 (2014).
33. Wei, L. et al. Identification of TYW3/CRYZ and FGD4 as susceptibility genes for amyotrophic lateral sclerosis. *Neurol Genet* 5, e375 (2019).
34. Iacoangeli, A. et al. Genome-wide Meta-analysis Finds the ACSL5-ZDHHC6 Locus Is Associated with ALS and Links Weight Loss to the Disease Genetics. *Cell Rep.* 33, 108323 (2020).
35. Nakamura, R. et al. A multi-ethnic meta-analysis identifies novel genes, including ACSL5, associated with amyotrophic lateral sclerosis. *Communications Biology* 3, 1–9 (2020).
36. van Rheenen, W. et al. Common and rare variant association analyses in amyotrophic lateral sclerosis identify 15 risk loci with distinct genetic architectures and neuron-specific biology. *Nat. Genet.* 53, 1636–1648 (2021).
37. International Multiple Sclerosis Genetics Consortium. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science* 365, (2019).
38. Isobe, N. et al. An ImmunoChip study of multiple sclerosis risk in African Americans. *Brain* 138, 1518–1530 (2015).
39. Ordoñez, G. et al. Genomewide admixture study in Mexican Mestizos with multiple sclerosis. *Clin. Neurol. Neurosurg.* 130, 55–60 (2015).
40. Sirkis, D. W., Geier, E. G., Bonham, L. W., Karch, C. M. & Yokoyama, J. S. Recent advances in the genetics of frontotemporal dementia. *Curr. Genet. Med. Rep.* 7, 41–52 (2019).

41. Ferrari, R. et al. Frontotemporal dementia and its subtypes: a genome-wide association study. *Lancet Neurol.* 13, 686–699 (2014).
42. Reus, L. M. et al. Genome-wide association study of frontotemporal dementia identifies a C9ORF72 haplotype with a median of 12-G4C2 repeats that predisposes to pathological repeat expansions. *Transl. Psychiatry* 11, 451 (2021).
43. Pottier, C. et al. Potential genetic modifiers of disease risk and age at onset in patients with frontotemporal lobar degeneration and GRN mutations: a genome-wide association study. *Lancet Neurol.* 17, 548–558 (2018).
44. Che, X.-Q. et al. Genetic Features of MAPT, GRN, C9orf72 and CHCHD10 Gene Mutations in Chinese Patients with Frontotemporal Dementia. *Curr. Alzheimer Res.* 14, 1102–1108 (2017).
45. Renton, A. E. et al. A genome-wide association study of myasthenia gravis. *JAMA Neurol.* 72, 396–404 (2015).
46. Seldin, M. F. et al. Genome-Wide Association Study of Late-Onset Myasthenia Gravis: Confirmation of TNFRSF11A and Identification of ZBTB10 and Three Distinct HLA Associations. *Mol. Med.* 21, 769–781 (2016).
47. Chia, R. et al. Identification of genetic risk loci and prioritization of genes and pathways for myasthenia gravis: a genome-wide association study. *Proc. Natl. Acad. Sci. U. S. A.* 119, (2022).
48. Sakaue, S. et al. A cross-population atlas of genetic associations for 220 human phenotypes. *Nat. Genet.* 53, 1415–1424 (2021).
49. Na, S.-J. et al. Whole-genome analysis in Korean patients with autoimmune myasthenia gravis. *Yonsei Med. J.* 55, 660–668 (2014).

50. Murai, H. et al. Characteristics of myasthenia gravis according to onset-age: Japanese nationwide survey. *J. Neurol. Sci.* 305, 97–102 (2011).
51. Rongve, A. et al. GBA and APOE ϵ 4 associate with sporadic dementia with Lewy bodies in European genome wide association study. *Sci. Rep.* 9, 7013 (2019).
52. Chia, R. et al. Genome sequencing analysis identifies new loci associated with Lewy body dementia and provides insights into its genetic architecture. *Nat. Genet.* 53, 294–303 (2021).
53. Guerreiro, R. et al. Investigating the genetic architecture of dementia with Lewy bodies: a two-stage genome-wide association study. *Lancet Neurol.* 17, 64–74 (2018).
54. Schrijvers, E. M. C. et al. Genome-wide association study of vascular dementia. *Stroke* 43, 315–319 (2012).
55. Moreno-Grau, S. et al. Genome-wide association analysis of dementia and its clinical endophenotypes reveal novel loci associated with Alzheimer’s disease and three causality networks: The GR@ACE project. *Alzheimers. Dement.* 15, 1333–1347 (2019).
56. Fongang, B. et al. A meta-analysis of genome-wide association studies identifies new genetic loci associated with all-cause and vascular dementia. *Alzheimers. Dement.* 17 Suppl 3, e056081 (2021).
57. Singh, V., Dhamoon, M. S. & Alladi, S. Stroke Risk and Vascular Dementia in South Asians. *Curr. Atheroscler. Rep.* 20, 43 (2018).
58. Sahadevan, S. et al. Ethnic differences in Singapore’s dementia prevalence: the stroke, Parkinson's disease, epilepsy, and dementia in Singapore study. *J. Am. Geriatr. Soc.* 56, 2061–2068 (2008).
59. Simpkins, A. N. Impact of Race-Ethnic and Economic Disparities on Rates of Vascular

- Dementia in the National Inpatient Sample Database from 2006-2014. *J. Stroke Cerebrovasc. Dis.* 29, 104731 (2020).
60. Hindorff, L. A. et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. U. S. A.* 106, 9362–9367 (2009).
61. Wray, N. R., Goddard, M. E. & Visscher, P. M. Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Res.* 17, 1520–1528 (2007).
62. Sham, P. C. & Purcell, S. M. Statistical power and significance testing in large-scale genetic studies. *Nat. Rev. Genet.* 15, 335–346 (2014).
63. Manolio, T. A. et al. Finding the missing heritability of complex diseases. *Nature* 461, 747–753 (2009).
64. Campbell, M. C. & Tishkoff, S. A. African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. *Annu. Rev. Genomics Hum. Genet.* 9, 403–433 (2008).
65. Kamiza, A. B. et al. Transferability of genetic risk scores in African populations. *Nat. Med.* 28, 1163–1166 (2022).
66. Naslavsky, M. S. et al. Global and local ancestry modulate APOE association with Alzheimer’s neuropathology and cognitive outcomes in an admixed sample. *bioRxiv* (2022) doi:10.1101/2022.02.02.22270331.
67. Jun, G. R. et al. Transethnic genome-wide scan identifies novel Alzheimer’s disease loci. *Alzheimers. Dement.* 13, 727–738 (2017).
68. Blue, E. E., Horimoto, A. R. V. R., Mukherjee, S., Wijsman, E. M. & Thornton, T. A. Local ancestry at APOE modifies Alzheimer’s disease risk in Caribbean Hispanics.

- Alzheimers. Dement. 15, 1524–1532 (2019).
69. Mesaros, M. Investigating the genetic profile of amyotrophic lateral sclerosis in patients of diverse race, ethnicity, and ancestry (REA). (The Ohio State University, 2021).
 70. Stepler, K. E. et al. ABCA7, a Genetic Risk Factor Associated with Alzheimer’s Disease Risk in African Americans. *J. Alzheimers. Dis.* (2022) doi:10.3233/JAD-215306.
 71. Rajan, R. et al. Genetic Architecture of Parkinson’s Disease in the Indian Population: Harnessing Genetic Diversity to Address Critical Gaps in Parkinson's Disease Research. *Front. Neurol.* 11, 524 (2020).
 72. Punia, S. et al. Absence/rarity of commonly reported LRRK2 mutations in Indian Parkinson’s disease patients. *Neurosci. Lett.* 409, 83–88 (2006).
 73. Vijayan, B., Gopala, S. & Kishore, A. LRRK2 G2019S mutation does not contribute to Parkinson’s disease in South India. *Neurol. India* 59, 157–160 (2011).
 74. Vishwanathan Padmaja, M., Jayaraman, M., Srinivasan, A. V., Srikumari Srisailapathy, C. R. & Ramesh, A. The SNCA (A53T, A30P, E46K) and LRRK2 (G2019S) mutations are rare cause of Parkinson’s disease in South Indian patients. *Parkinsonism Relat. Disord.* 18, 801–802 (2012).
 75. Ibanez, A. et al. The Multi-Partner Consortium to Expand Dementia Research in Latin America (ReDLat): Driving Multicentric Research and Implementation Science. *Front. Neurol.* 12, 631722 (2021).
 76. Ibanez, A., Parra, M. A., Butler, C. & Latin America and the Caribbean Consortium on Dementia (LAC-CD). The Latin America and the Caribbean Consortium on Dementia (LAC-CD): From Networking to Research to Implementation Science. *J. Alzheimers. Dis.* 82, S379–S394 (2021).

77. Crane, P. K. et al. Cognitive assessments in ADNI: Lessons learned from the ADNI psychometrics project. *Alzheimer's & Dementia* vol. 17 Preprint at <https://doi.org/10.1002/alz.056474> (2021).
78. Bandres-Ciga, S. Black and African American Connections to Parkinson's Disease Study: Addressing Missing Diversity in Parkinson's Disease Genetics. *Mov. Disord.* 37, 1559–1561 (2022).
79. Program, T. G. P. G. & The Global Parkinson's Genetics Program. GP2 : The Global Parkinson's Genetics Program. *Movement Disorders* vol. 36 842–851 Preprint at <https://doi.org/10.1002/mds.28494> (2021).
80. Borda, M. G. et al. Colombian consortium for the study of Lewy body dementia COL-DLB. *J. Neurol. Sci.* 412, 116807 (2020).
81. Oppedal, K. et al. European DLB consortium: diagnostic and prognostic biomarkers in dementia with Lewy bodies, a multicenter international initiative. *Neurodegener. Dis. Manag.* 9, 247–250 (2019).
82. Rizig, M. et al. Identification of genetic risk loci and causal insights associated with Parkinson's disease in African and African admixed populations: a genome-wide association study. *Lancet Neurol.* 22, 1015–1025 (2023).

Gloss to Chapter 3.

When I started graduate school, I didn't know what single cell analyses were nor did I even understand that you could look at different transcriptomic profiles from different human tissues. I remember hearing presentations on single cell analyses, and I always thought it sounded so impressive. Unfortunately, it also sounded very inaccessible in presentations.

While working on chapter 3, I started collaborating more with Yokoyama lab members and I learned so much. I worked very closely with Daniel Sirkis, who taught me so much about single cell analyses and manuscript preparation. I also had the opportunity to mentor numerous rotation students while working on this chapter. I mentored almost ten individuals during my PhD including four rotation students, much of which I felt prepared for due to Dan's incredible mentorship.

As Dan provided single cell code for me to conduct my first analysis, I decided a good way to pay this forward would be to build our Yokoyama Lab Github page with this, and other resources, to help other and future lab members. This Github now has numerous repositories and job aids that have been invaluable.

Dan set an incredible example for me about what thorough research looks like. He is always sending me new papers to read and thinking about new ways to analyze data. He has taught me that when you are deciding which path to take in a research project, sometimes the best answer is "both"! Good research means looking at the questions and the data in numerous different ways.

While working through this chapter I have made some beautiful visualizations that I am very proud of and have learned how to explain the difficult concepts in single cell rna-sequencing in an accessible way. Turning difficult material into digestible content is something I am extremely passionate about and hope to continue doing in my post-graduate school career.

Since the beginning of my work in chapter 3, I have submitted two single-cell papers, and I am currently working on 3 other single cell projects. I am fascinated by the connection between genetics, brain transcriptomics, and blood transcriptomics. In the future I want to continue looking at both blood and brain transcriptomics and better understand how one may mirror the other.

Chapter 3 Paper 1

Single-cell RNA-seq reveals alterations in peripheral CX3CR1 and nonclassical monocytes in familial tauopathy

Contributing Authors: Daniel W. Sirkis*, Caroline Warly Solsberg*, Taylor P. Johnson, Luke W. Bonham, Virginia E. Sturm, Suzee E. Lee, Katherine P. Rankin, Howard J. Rosen, Adam L. Boxer, William W. Seeley, Bruce L. Miller, Ethan G. Geier, and Jennifer S. Yokoyama**

*Equal contribution

**Correspondence

3.1.1 Abstract

3.1.1.1 Background

Emerging evidence from mouse models is beginning to elucidate the brain's immune response to tau pathology, but little is known about the nature of this response in humans. In addition, it remains unclear to what extent tau pathology and the local inflammatory response within the brain influence the broader immune system.

3.1.1.2 Methods

To address these questions, we performed single-cell RNA sequencing (scRNA-seq) of peripheral blood mononuclear cells (PBMCs) from carriers of pathogenic variants in *MAPT*, the gene encoding tau (n = 8), and healthy non-carrier controls (n = 8). Primary findings from our scRNA-seq analyses were confirmed and extended via flow cytometry, droplet digital (dd)PCR, and secondary analyses of publicly available transcriptomics datasets.

3.1.1.3 Results

Analysis of ~181,000 individual PBMC transcriptomes demonstrated striking differential expression in monocytes and natural killer (NK) cells in *MAPT* pathogenic variant carriers. In particular, we observed a marked reduction in the expression of *CX3CR1* – the gene encoding the fractalkine receptor that is known to modulate tau pathology in mouse models – in monocytes and NK cells. We also observed a significant reduction in the abundance of nonclassical monocytes and dysregulated expression of nonclassical monocyte marker genes, including *FCGR3A*. Finally, we identified reductions in *TMEM176A* and *TMEM176B*, genes thought to be involved in the inflammatory response in human microglia but with unclear function in peripheral monocytes. We confirmed the reduction in nonclassical monocytes by flow cytometry and the differential expression of select biologically relevant genes dysregulated in our scRNA-seq data using ddPCR.

3.1.1.4 Conclusions

Our results suggest that human peripheral immune cell expression and abundance are modulated by tau-associated pathophysiologic changes. *CX3CR1* and nonclassical monocytes in particular will be a focus of future work exploring the role of these peripheral signals in additional tau-associated neurodegenerative diseases.

3.1.2 Introduction

Nearly 25 years after the discovery of pathogenic variants in *MAPT* (encoding the microtubule-associated protein tau) in familial frontotemporal dementia (FTD; [1, 2]; reviewed in [3]), there are still no effective therapeutics capable of halting or delaying tau-associated neurodegeneration [4, 5]. Diverse tau proteinopathies (tauopathies) also occur sporadically (i.e., in the absence of *MAPT* or other pathogenic variants) and are subdivided into primary tauopathies – which collectively fall under the umbrella term, frontotemporal lobar degeneration (FTLD)-tau – and secondary tauopathies, the most prominent example of which is Alzheimer’s disease (AD). Although much effort has gone into characterizing the natural history and longitudinal declines of *MAPT* pathogenic variant carriers (reviewed in [6]), we understand relatively little about the molecular mechanisms that impart risk for sporadic forms of tauopathy – whether primary or secondary. However, because it remains challenging to confidently diagnose sporadic tauopathy in vivo, familial tauopathy (identified via the presence of pathogenic *MAPT* variants in individuals with a family history of neurodegenerative disease) represents a powerful lens through which we can not only elucidate the pathophysiologic processes underlying hereditary tauopathy but also uncover shared processes that may contribute to sporadic tauopathy.

The last decade has witnessed a major revival in interest in immune mechanisms that may modulate risk for neurodegeneration, with a primary focus on microglia, the parenchymal macrophages of the brain (reviewed in [7]). Significantly less progress has been made in elucidating peripheral blood or cerebrospinal fluid (CSF) leukocyte perturbations in – and responses to – neurodegeneration, although this is beginning to change (reviewed in [8, 9, 10]). For example, we now know of changes to CD8⁺ T cells in AD and CD4⁺ T cells in Lewy body dementia [11, 12]. In addition, altered phospholipase C- γ 2 signaling in peripheral lymphocytes

has been reported in AD [13]. Beyond lymphocytes, changes in peripheral monocytes (particularly nonclassical [NC] monocytes) have been observed in Parkinson's disease (PD) [14] and amyotrophic lateral sclerosis (ALS; [15, 16]). In addition, patients with the hereditary white-matter disorder, adult-onset leukoencephalopathy with axonal spheroids and pigmented glia (ALSP), show striking reductions in peripheral NC monocytes [17]. Given that ALSP is thought to be driven by primary microglial defects (reviewed in [18]), the reduction in peripheral NC monocytes in this disorder suggests shared biology between these two cell types.

High-quality fluid biomarkers are being developed for AD (reviewed in [19, 20]) and FTD (reviewed in [21]), but those that can distinguish between the major neuropathologic division within FTD (i.e., FTLN-tau vs. FTLN due to TDP-43 pathology [FTLN-TDP]) have been lacking [21] until very recently [22]. We reasoned that an in-depth exploration of peripheral immune dysregulation in tauopathy may reveal novel, blood-based biomarkers associated with FTLN-tau. Therefore, in an effort to define the peripheral immune signatures of tauopathy, we carried out single-cell RNA sequencing (scRNA-seq) of peripheral blood mononuclear cells (PBMCs) in individuals with pathogenic *MAPT* variants – who have (or will develop) FTLN-tau pathology – and cognitively normal, non-carrier controls. We identified striking changes in NC monocytes – both in terms of cellular abundance and gene expression – as well as natural killer (NK) cells and other cell types. Moreover, we identified *CX3CRI* as a potentially novel peripheral marker of tauopathy, suggesting parallel changes in *CX3CRI* in microglia and peripheral leukocytes in the course of tau-mediated neurodegeneration. We also identified additional candidate genes whose expression may be altered in the periphery in tauopathy (e.g., *FCGR3A* and *TMEM176A/B*). Considering recent findings in ALS and PD, and given that NC monocytes can be detected in the human brain [23], our results add to the weight of evidence suggesting the importance of NC

monocytes across a spectrum of neurodegenerative diseases. Taken together, our results indicate that PBMCs represent an accessible and informative tissue source not only for biomarker discovery but also for elucidation of peripheral immune responses in the context of tauopathy.

3.1.3 Methods

3.1.3.1 Study Participants

All participants or their surrogates provided written informed consent prior to study participation; all aspects of the studies described here were approved by the University of California, San Francisco (UCSF) institutional review board. Sixteen individuals ($n = 8$ *MAPT* pathogenic variant carriers and $n = 8$ cognitively normal, non-carrier controls) participated in this study. Individuals were recruited as part of ongoing studies of FTD and related neurodegenerative diseases at the UCSF Memory and Aging Center (MAC). All participants underwent a multistep screening with an in-person visit at the MAC that included a neurologic exam, detailed cognitive assessment, medical history, and family history for neurodegenerative disease. Each participant's study partner was interviewed regarding functional abilities. A consensus team of clinicians then reviewed all participants. In addition, all participants were screened for pathogenic variants in established FTLD genes, including *MAPT*. *MAPT* pathogenic variant carriers had clinical syndromes of bvFTD ($n = 4$), frontal AD ($n = 1$), subjective cognitive impairment ($n = 1$), or were clinically normal and presymptomatic ($n = 2$). Pathogenic *MAPT* variants represented within this study were p.P301L ($n = 1$), p.S305I ($n = 1$), p.S305S ($n = 1$), p.R406W ($n = 3$), and IVS10+16 ($n = 2$). Pathogenic variant carriers and cognitively normal controls were sex-matched, and equal numbers of female and male participants were included in the cohort. Age was not significantly different between carrier and control groups, as assessed using an unpaired t-test. Demographic information for study participants is included in **Table 3.1.1**.

3.1.3.2 Clinical Assessment

Study participants underwent a multistep screening process prior to an in-person clinical evaluation at the UCSF MAC, which included a neurological exam, cognitive assessment, and medical history [24]. Each participant's study partner was interviewed to assess the participant's functional abilities. A multidisciplinary team composed of a behavioral neurologist, a neuropsychologist, and a registered nurse established clinical diagnoses for cases according to consensus criteria for FTD and its subtypes [25] or frontal AD [26]. Controls and presymptomatic *MAPT* carriers had a normal neurological exam and a Clinical Dementia Rating scale Sum of Boxes (CDR-SB) [27] score of 0. All controls screened negative for disease-causing pathogenic variation in established AD and FTD genes.

3.1.3.3 Cell Isolation

Human PBMCs were obtained from study participants at the UCSF MAC. Blood specimens were collected in yellow-top acid-citrate-dextrose vacutainer tubes (BD Biosciences) and processed within five hours of collection. PBMCs were isolated via Ficoll density gradient centrifugation using Lymphosep separation medium (MP Biomedicals), washed with Ca²⁺- and Mg²⁺-free PBS (ThermoFisher), and treated with one application of red blood cell lysis buffer (Biolegend). After a final wash step in PBS, PBMCs were diluted to a density of 1.5 x 10⁶ cells/ml in freezing media composed of 10% DMSO in FBS and immediately frozen at -80°C. After two weeks, samples were transferred to liquid nitrogen for long-term storage. All PBMC samples used in our primary analyses were cryopreserved. Sensitivity analysis involving freshly isolated PBMCs was performed to compare gene expression in paired fresh vs. frozen and thawed samples.

3.1.3.4 Single-Cell RNA-seq

PBMCs were thawed and prepared for scRNA-seq using the Chromium Single Cell 3' v2 kit according to the manufacturer's instructions (10x Genomics). Samples were processed in two separate batches of eight samples each, with four *MAPT* variant carriers and four controls included in each batch. To further minimize the potential for batch effects, each batch contained equal numbers of samples from female and male participants. After sample thawing, counting, and dilution, PBMCs underwent standard 10x processing, 3' gene expression library construction steps, and next-generation sequencing at the UCSF Genomics CoLab and Institute for Human Genetics (IHG).

3.1.3.5 Sequencing Data Processing

For each of the two batches, single-cell 3' libraries generated from eight samples were pooled and sequenced on one lane of a NovaSeq S4 flow cell. Raw sequencing reads were aligned to GRCh38, and feature-barcode matrices were generated using Cell Ranger version 3.0.2.

3.1.3.6 Quality Control and Clustering

We obtained a total of 7.3×10^9 reads and detected ~214,000 cells across the two independent 10x and sequencing batches, yielding a moderate sequencing depth [28,29] of ~34,000 mean reads/cell. We detected ~3,700 median UMI counts/cell and ~1,100 median genes/cell (**Table S1**). There were no significant differences in the number of cells captured per sample, the number of reads per sample, or the mean read depth per sample when comparing the *MAPT* pathogenic variant carrier group to the non-carrier control group. Subsequent quality-control (QC) and downstream analysis steps were performed using Seurat v4.1 [30, 31]. QC filtering was applied to individual-sample feature-barcode matrices and consisted of the following steps: (i) genes detected in < 10 cells were removed; (ii) cells with ≤ 500 detected genes were removed; (iii) cells with ≤ 500 counts

and those with $\geq 20,000$ counts were removed; (iv) cells with mitochondrial mapping percentages ≥ 10 were removed; (v) doublets were identified and removed using DoubletFinder v2.0.3 [32, 33] using the recommended parameter settings. After stringent QC filtering, $\sim 181,000$ cells remained for downstream analysis (**Table S2**).

After QC, we performed the following additional processing steps: (i) we applied `sctransform` [34]—a method for scRNA-seq count normalization and variance stabilization—at the individual-sample level, including mitochondrial mapping percentage as a covariate [34, 35], to minimize variability due to differences in sequencing depth between samples; (ii) the 16 individual samples were integrated with `FindIntegrationAnchors` and `IntegrateData`, specifying ‘`sctransform`’ as the normalization method and reciprocal principal component analysis (PCA) as the reduction. Subsequently, PCA was performed followed by uniform manifold approximation and projection (UMAP) reduction using the first 30 PCs; clustering was performed using a resolution parameter of 0.5. This resulted in the generation of 21 distinct clusters that were annotated on the basis of marker gene expression, identified using `FindMarkers` and literature searches.

3.1.3.7 Differential Expression Analysis

Differential expression analysis was performed using `limma` [36-39] on individual clusters and `sctransform`-normalized data, with *MAPT* pathogenic variant carrier status as the contrast and age, sex, and scRNA-seq batch as covariates. To account for multiple testing, a false discovery rate-corrected p -value ($p_{\text{FDR}} < 0.05$) was considered statistically significant. For visualization of selected differentially expressed genes (DEGs), we used violin plots displaying normalized count data generated via `Seurat`’s `NormalizeData` function.

3.1.3.8 Cluster Proportionality

Cluster proportions were determined for individual samples by dividing the number of cells in a given cluster by the total number of cells in all clusters (after QC filtering) for each individual. Differences in cluster proportionality were assessed visually for all clusters according to *MAPT* variant carrier status. Only the NC monocyte cluster (cluster 11) showed a clear difference in proportionality between carriers and controls. Significance for cluster 11 proportionality, as well as NC monocyte subcluster ratios, was determined by linear modeling, covarying for age, sex, and scRNA-seq batch.

3.1.3.9 STRING Network Analysis

DEGs with $p_{\text{FDR}} < 0.05$ and \log_2 fold-changes (LFC) > 0.1 for selected cell clusters were submitted for analysis via the STRING database (v11.5) [40] using the following parameters: the full STRING network was queried; network edge thickness indicated the confidence of the interaction; active interaction sources included databases, experiments, and text mining; a minimum interaction score of 0.4 was required; only query genes/proteins were displayed; disconnected nodes (i.e., DEGs with no interaction partners) were not displayed; gene modules were color-coded according to the results of Markov cluster algorithm (MCL) clustering [41], with modules color-coded within each panel according to their respective gene count. For clarity, within the Results section we refer to specific MCL clusters (containing DEGs) as gene/protein ‘modules,’ while reserving the term ‘cluster’ to refer to cell clusters generated via scRNA-seq analysis.

3.1.3.10 Analysis of Mouse RNA-seq Data

Publicly available mouse RNA-seq data were downloaded from GEO (accession GSE93180) and originally published in [42]. Briefly, hippocampal microglia (Cd11b⁺ myeloid cells) were sorted from 6-month-old male hMAPT-P301S transgenic mice ($n = 6$) or non-transgenic littermates

($n = 6$) using BD FACSAria sorters, then RNA was extracted. RNA-seq libraries were prepared using the Ovation RNA-Seq System V2 (NuGEN). Reads were aligned to the GRCm38 genome using GSNAP and gene counts were acquired using the HTSeqGenie Bioconductor package. Average sequencing depth was 30 million total reads with 8% of reads aligning to exons. We downloaded the available counts data, reanalyzed it using DESeq2 [43], and plotted normalized counts for mouse *Cx3cr1* using ggplot2 [44].

3.1.3.11 RNA Extraction

To minimize biological variability and facilitate orthogonal validation, for droplet digital (dd)PCR experiments, we used PBMCs isolated from 15 of the 16 participants originally selected for scRNA-seq analysis. One additional cognitively normal control sample was used for ddPCR due to unavailability of one control sample used for scRNA-seq. RNA was extracted from the PBMCs using the RNeasy Micro Kit (Qiagen) and isolated RNA was quantified and its quality was assessed using the RNA 6000 Pico Bioanalyzer kit (Agilent). PBMC RNA samples had RNA integrity number (RIN) values ranging from 9.2–9.9, indicating high-quality RNA (**Table S3**; [45]).

3.1.3.12 Droplet Digital PCR

One ng of total RNA was used for single-tube reverse transcription (RT) and ddPCR using the One-Step RT-ddPCR Advanced kit (Bio-Rad). Droplets were generated and subsequently analyzed using the QX100 system (Bio-Rad) at the UCSF Center for Advanced Technology (CAT). Reactions were prepared and run essentially according to the manufacturer's instructions. For steps in which a temperature range was specified, we used the following parameters: RT was performed at 50°C, annealing/extension occurred at 55°C, and samples were held at 12°C in the C1000 thermocycler (Bio-Rad) prior to analysis on the droplet reader. To confirm specificity, we

ran the following control reactions: wells lacking RNA but containing all other components and wells lacking reverse transcriptase but containing all other components. PrimePCR ddPCR Gene Expression primer–probe mixes coupled to FAM or HEX (Bio-Rad) were used to amplify specific genes.

3.1.3.13 Flow Cytometry

Multicolor flow cytometry was performed on thawed PBMC samples using an LSRFortessa (BD). Samples were stained using the following fluorochrome-conjugated antibodies: PE-CF594 mouse anti-human CD14 (BD), APC mouse anti-human CD16 (BD), BV 421 mouse anti-human CD56 (BioLegend), and PE rat anti-human CX3CR1 (BD). Viability was assessed using the LIVE/DEAD Aqua dye (ThermoFisher). Cells were labeled at room temperature, covered from light for 30 min in a buffer consisting of 1% fetal bovine serum in MACS buffer containing PBS and EDTA (pH 7.2; Miltenyi). After staining, samples were washed three times in the above buffer, then kept on ice until analysis on the LSRFortessa. Compensation was performed using UltraComp eBeads (ThermoFisher) for the antibodies and ArC amine-reactive compensation beads for LIVE/DEAD Aqua (ThermoFisher). Control conditions included unstained PBMCs as well as ‘fluorescence minus one’ (FMO) conditions in which a single antibody was omitted. For the primary samples to be analyzed, we acquired data on ~100,000 cells falling within the initially defined PBMC gate. All post-acquisition gating and analysis was performed in FlowJo v10 (BD). We used the following sequential gating scheme: (i) debris was excluded via an initial PBMC gate; (ii) live PBMCs were gated by low LIVE/DEAD Aqua fluorescence; (iii) singlets falling along the forward scatter (FSC)-height (-H) vs. FSC-area (-A) diagonal were gated next; (iv) finally, monocytes were gated based on their high side scatter (SSC-A) and CD14 positivity. Monocyte subtypes were gated from a starting population of all monocytes based on their characteristic CD14

vs. CD16 staining patterns. For quantifications, frequencies expressed as a percentage of all PBMCs used the number of cells in the live, singlet PBMC population as the denominator, while frequencies expressed as a percentage of all monocytes used the number of cells in the live, singlet monocyte population as the denominator.

3.1.3.14 Analysis of Publicly Available Human Brain scRNA-seq Dataset

Publicly available human brain scRNA-seq data were downloaded from GEO (accession GSE137444) and originally published in [46]. Briefly, human microglia from patients ($n = 7$) who underwent amygdalohippocampectomy were isolated from the temporal cortex, FACS-sorted, processed for scRNA-seq using the Chromium Single Cell v2 kit (10x Genomics), and sequenced on a HiSeq 4000 (Illumina). The downloaded counts data were processed and analyzed as described above for our PBMC scRNA-seq data. We used the marker genes identified in [23] to identify the cluster representing human brain CD16⁺ NC monocytes. After identification of the NC monocytes, we assessed cluster-specific expression of *CX3CR1*, *FCGR3A*, *TMEM176A/B*, and *C3AR1*.

3.1.3.15 Additional Statistical Analyses

For the analysis of ddPCR data, we performed linear modeling to assess whether *MAPT* carrier status predicted differences in gene expression while covarying for age and sex. Log₂-transformed absolute concentration data for *CX3CR1*, *FCGR3A*, *TMEM176A/B*, and *C3AR1* (or the ratios of these values with those of reference gene *EEF2*) were used for analyses assuming data normality, while non-transformed data are displayed in the plots. *EEF2* levels can be leveraged as a normalization factor to reduce technical variation across samples. In our assays, we found that the results were unaffected by normalization. To test for differences in mitochondrial genome mapping percentages and differences in the proportions of monocyte subtypes calculated from flow

cytometry data, we used Welch's t-test. Differential gene expression for the mouse bulk RNA-seq data was determined using the Wald test implemented in DESeq2. Differences were considered significant at $p \leq 0.05$ (ddPCR, mitochondrial mapping data, and flow cytometry) or $p_{\text{FDR}} < 0.05$ (bulk RNA-seq data). Analyses were performed in R and plots were generated with ggplot2.

3.1.4 Results

3.1.4.1 Reduced Nonclassical Monocytes in MAPT Pathogenic Variant Carriers

After QC filtering, clustering of ~181,000 PBMCs generated 21 primary clusters representing the major lymphoid and myeloid cell types, including CD4^+ and CD8^+ T cells, B cells, NK cells, monocytes, and dendritic cells (**Figure 3.1.1A; Figure 3.1.S1**). As expected, and in contrast to many other FTD-associated genes, *MAPT* expression was barely detectable in PBMCs (**Figure 3.1.S2**). We reasoned, therefore, that any changes detected in PBMC cell-type proportionality or gene expression in *MAPT* pathogenic variant carriers would most likely represent a response to tau neuropathology or neurodegeneration rather than cell-intrinsic changes occurring directly downstream of variant *MAPT*. Of all PBMC clusters, only one showed a clear, consistent change in abundance in *MAPT* pathogenic variant carriers relative to controls. This cluster (11), which represents *FCGR3A*⁺ (CD16^+) NC monocytes (**Figure 3.1.1A, B**), localized in UMAP space near the more abundant *CD14*⁺ classical monocyte cluster (2) and the *CLEC10A*⁺ conventional dendritic cell (cDC) cluster (14). In particular, NC monocytes, expressed as a percentage of total PBMCs for each participant, were significantly reduced in *MAPT* carriers (**Figure 3.1.1C**). To gain more fine-grained insight into the nature of this change in abundance, we subsetted and re-clustered myeloid clusters 2, 11, and 14. Assessing the myeloid subclusters (**Figure 3.1.1D, left**), we discovered that normalizing the NC monocyte subcluster to the *CLEC10A*⁺, *CD1C*⁺ cDC subcluster (corresponding to the most abundant blood cDC population, known as cDC2 [47];

Figure 3.1.1D, lower right panel) yielded better separation between *MAPT* variant carriers and non-carrier controls (**Figure 3.1.1E**). Further assessing the myeloid subclusters, we found that normalizing NC monocyte numbers to total cDCs (i.e., cDC1 + cDC2) gave a similar result (**Figure 3.1.S3A**). On the other hand, expressing NC monocyte abundance as a fraction of all monocytes or all myeloid cells did not significantly differentiate carriers from controls (**Figure 3.1.S3B, C**).

3.1.4.2 Differential Expression Analysis

We next performed differential expression analysis on each of the primary cell clusters, comparing *MAPT* variant carriers to non-carrier controls and adjusting for age, sex, and scRNA-seq batch. Focusing initially on DEGs with absolute LFC > 0.2, we determined that cDC, NC monocyte, and NK cell clusters had the highest number of DEGs (**Figure 3.1.2A, B**). We further characterized the DEGs of NC monocytes and NK cells by querying the STRING database [40] for functional and physical interactions, now using a more-permissive LFC cutoff of 0.1 to facilitate population of the gene interaction network (**Table S4** for DEG lists for all clusters).

3.1.4.3 STRING Network Analysis

MAPT carriers showed striking up-regulation of many ribosomal and mitochondrial genes that, respectively, formed large interaction modules (**Figure 3.1.3A, B**). The biological significance of these coordinately up-regulated DEGs is unclear, but tau is known to interact with and affect multiple ribosomal subunits [48–51] and mitochondrial proteins [51, 52]. Because the identified mitochondrial DEGs represent a subset of the mitochondrial genes used during QC (see Methods) to filter out putatively damaged cells [35], we considered the possibility that – despite the removal of cells with high mitochondrial mapping percentages ($\geq 10\%$) – the apparent up-regulation of mitochondrial DEGs may nevertheless be associated with higher mitochondrial mapping

percentage in *MAPT* variant carriers. Consistent with this possibility, mitochondrial mapping percentage was subtly but significantly higher in *MAPT* carriers in the NC monocyte and NK cell clusters (11 and 3, respectively; **Figure 3.1.S4A, B**). On the other hand, mitochondrial DEGs were absent from the cDC cluster (14; **Figure 3.1.S4D; Table S4**) – despite this cluster having the highest overall number of DEGs with LFC > 0.2 (**Figure 3.1.2A**) – and mitochondrial mapping percentage was not significantly increased in variant carriers in this cluster (**Figure 3.1.S4C**). This suggests (i) cell-type-specific and, presumably, biologically relevant dysregulation of mitochondrial genes in *MAPT* variant carriers, consistent with [51]; and (ii) that the up-regulation of mitochondrial DEGs in the NC monocyte and NK cell clusters is associated with increased mitochondrial mapping percentage. To test the latter possibility, we next included mitochondrial mapping percentage as an additional covariate in the differential expression analyses, and, as expected, nearly all mitochondrial DEGs that previously had estimated LFCs > 0.1 no longer passed this threshold (not shown). Thus, the presence of mitochondrial DEGs is associated with increased mitochondrial mapping percentage. Importantly, coordinated up-regulation of mitochondrial genes could lead to subtle shifts in mitochondrial mapping percentage, and therefore the causality underlying this relationship is unclear.

The other major up-regulated STRING module found in both NC monocytes and NK cells contained members of the AP-1 transcription factor family [53], including *FOS*, *JUN*, and *JUNB* (**Figure 3.1.3A, B**). Multiple members of this module (*FOS*, *DUSP1*) were previously found to be up-regulated via bulk measurements of both PBMCs and hippocampus in AD [54], suggesting consistent dysregulation of AP-1 transcription factor genes in both primary and secondary tauopathies. *NFKBIA*, encoding the NF- κ B inhibitor- α , also appears in this module in both NC

monocytes and NK cells, and this gene is highly up-regulated by treatment of microglia with tau paired-helical filaments [55] and fibrils [56] and in the brain in late-onset AD [57].

In terms of significantly down-regulated genes, both NC monocytes and NK cells contained a module populated by *FCGR3A* and *CX3CR1* (**Figure 3.1.3C, D**). Although *FCGR3A* is an established marker of NC monocytes, it is also expressed by a subset of NK cells (**Figure 3.1.1B**). *CX3CR1* is highly expressed by both NC monocytes and NK cells (**Figure 3.1.4A**) and, strikingly, is a well known modulator of tau pathophysiology in the brain [58–60]. Additional down-regulated modules in NC monocytes included those containing components of the complement pathway (*CFD*, *CFP*, *C3ARI* [61,62]), members of the S100 alarmin family (*S100A8-11* [63,64]), highly lipopolysaccharide (LPS)-responsive microglial genes (*TMEM176A*, *TMEM176B* [65]), and the lysosomal gene *PSAP*, which promotes pro-inflammatory activity in monocytes [66] (**Figure 3.1.3C**). Collectively, the down-regulation of these gene modules in *MAPT* pathogenic variant carriers suggests possible dampening of latent pro-inflammatory pathways in NC monocytes. Of note, this apparent phenotypic change also occurred within a diminished population of circulating NC monocytes.

3.1.4.4 Validation of Select Differentially Expressed Genes

We next sought to validate a handful of DEGs which showed robust changes by differential expression analysis, often in multiple cell types, focusing on those deemed most likely to be biologically relevant to tau pathophysiology. In particular, we selected the following genes for validation via an orthogonal technique, ddPCR: *CX3CR1*, *FCGR3A*, *TMEM176A*, *TMEM176B*, and *C3ARI*. As noted above, *CX3CR1* has many well-established connections to tau pathology via its role in microglia. However, to our knowledge, its role in peripheral myeloid cells has not been studied in the context of neurodegenerative disease. *FCGR3A* not only serves as a marker gene for

the significantly reduced population of NC monocytes, but was also significantly down-regulated in both NC monocytes and NK cells of *MAPT* carriers. *TMEM176A/B* are less well known but have been shown to be extremely responsive to LPS treatment in human microglia [65], are homologs of the *MS4A* gene family involved in risk for AD [67], and have also been shown to be dysregulated in PBMCs from AD patients [54]. Finally, *C3AR1* has been implicated as a key player in the spread of tau neuropathology in mouse models [68], while the complement system more generally is thought to be a key regulator of neuronal loss in primary tauopathy as well as AD [69]. We first discuss validation of the *CX3CR1* finding below.

3.1.4.5 Reduced CX3CR1 Expression in Familial Tauopathy

CX3CR1 showed robust expression in NC monocytes and NK cells (**Figure 3.1.4A**), and its expression was significantly reduced in both cell types in *MAPT* pathogenic variant carriers (**Figure 3.1.4B, C**). To determine whether down-regulation in peripheral leukocytes was mirrored by changes in brain myeloid cells in the context of tauopathy, we analyzed a publicly available brain RNA-seq dataset [42] derived from *MAPT* P301S mouse hippocampal microglia. P301S microglia also displayed down-regulation of *Cx3cr1* (**Figure 3.1.4D**), indicating consistent *CX3CR1* dysregulation between mouse microglia and human peripheral immune cells in the context of tauopathy. Importantly, the P301S mutation used in this mouse model was not represented in our cohort, suggesting that the reduced peripheral *CX3CR1* observed here may be a general feature of familial tauopathy and may therefore occur downstream of pathogenic *MAPT* variants beyond those directly assessed in this work. To validate the change in *CX3CR1* expression in peripheral immune cells, we isolated PBMC RNA from *MAPT* pathogenic variant carriers and non-carrier controls and performed ddPCR. ddPCR confirmed the reduction in *CX3CR1* (**Figure 3.1.5A**). Normalizing the *CX3CR1* signal to reference gene *EEF2* gave similar results (**Figure**

3.1.S5A, B). When *MAPT* variant carriers were stratified by variant class (splicing vs. non-splicing), both groups showed significant reductions in *CX3CR1* (**Figure 3.1.5B**), suggesting that multiple mechanistic forms of familial tauopathy converge on perturbation of *CX3CR1* expression.

3.1.4.6 Diminished FCGR3A Expression in MAPT Pathogenic Variant Carriers

We next focused on validation of *FCGR3A*. We predicted that *FCGR3A* would show a robust decrease in expression in *MAPT* variant carriers in bulk PBMC RNA given that (i) NC monocytes are reduced in abundance in carriers (**Figure 3.1.1C**), and (ii) *FCGR3A* expression was reduced in both NC monocytes and NK cells by scRNA-seq differential expression analyses (**Figure 3.1.6A, 2B**). Indeed, ddPCR analysis showed a significant reduction in *FCGR3A* expression in *MAPT* pathogenic variant carrier PBMCs (**Figure 3.1.6B**). The observed reduction likely reflects both reduced expression of *FCGR3A* and reduced levels of NC monocytes, which express *FCGR3A* at high levels. As with *CX3CR1*, normalization of the *FCGR3A* signal to reference gene *EEF2* did not affect the results (**Figure 3.1.S5C**). We also analyzed additional marker genes of the NC monocyte cluster, including *VMO1* and *IFITM3*. *VMO1*, which is specifically expressed in cluster 11 (consistent with [17]), and *IFITM3*, which is enriched in cluster 11, both showed significant differential expression in *MAPT* variant carriers (**Figure 3.1.6C, D; Table S4**), indicating changes in multiple NC monocyte marker genes. Because the ddPCR and scRNA-seq studies described here depend on the use of cryopreserved PBMCs, we tested whether cryopreservation had a major effect on the measured expression levels of *CX3CR1* and *FCGR3A*. Direct comparison by ddPCR of paired, fresh vs. frozen/thawed samples suggested that cryopreservation was associated with only a modest, ~15-20% reduction in the apparent expression of these transcripts (**Figure 3.1.S6A, B**). Cryopreservation, therefore, was not associated with a major change in the levels of *CX3CR1* and *FCGR3A*.

3.1.4.7 Analysis of TMEM176A/B in Familial Tauopathy

TMEM176A/B represent poorly characterized genes of the extended *MS4A* family [67] thought to be involved in microglial LPS response [65], antigen presentation [70], and inflammasome regulation [71]. *TMEM176A/B* were highly expressed in classical and NC monocytes (clusters 2 and 11; **Figure 3.1.7A**) and were strongly down-regulated in *MAPT* variant carrier NC monocytes (**Figure 3.1.7B**). Reduced expression of *TMEM176A* and *TMEM176B* was confirmed via ddPCR of PBMC RNA (**Figure 3.1.7C, D**). The levels of *TMEM176A* and *TMEM176B*, as detected by ddPCR, were tightly associated with one another, and a subset of *MAPT* pathogenic variant carriers (5 of 8) displayed lower expression of both genes than any non-carrier controls (**Figure 3.1.S5D**).

3.1.4.8 Analysis of C3AR1 in MAPT Pathogenic Variant Carriers

Finally, we examined the expression of *C3AR1* given its importance in models of tauopathy [68]. *C3AR1* expression was enriched in NC monocytes (**Figure 3.1.8A**) and was strikingly reduced in this cell type in *MAPT* variant carriers (**Figure 3.1.8B**). *C3AR1* expression, as measured by ddPCR of PBMC RNA, trended toward reduction but did not achieve significance (**Figure 3.1.8C**). However, given the importance of this gene and the complement pathway more generally in tauopathy and AD, analysis of peripheral *C3AR1* expression in larger, better-powered cohorts is still warranted.

3.1.4.9 Validation of the Reduction in Nonclassical Monocytes via Flow Cytometry

To determine whether a central finding of the scRNA-seq analysis could be directly validated via an independent method, we performed multicolor flow cytometry analysis on PBMCs from *MAPT* pathogenic variant carriers and non-carrier controls. PBMCs, monocytes and monocyte subtypes were gated as illustrated in **Figure 3.1.9A and B** and as described in detail in the methods section. Importantly, this analysis enabled us to confirm a significant reduction in NC monocytes

(characterized by low CD14 expression and high CD16 expression) as a percentage of all PBMCs ($p = 0.02$; **Figure 3.1.9C**, upper left). We also observed a marginally significant reduction in NC monocytes expressed as a percentage of all monocytes ($p = 0.05$; **Figure 3.1.9C**; lower left). Intermediate monocytes (characterized by high CD14 levels and moderate-to-high CD16 levels) showed a trend toward reduced abundance, relative to either all PBMCs or all monocytes, but this difference did not reach significance (**Figure 3.1.9C**, middle panels). Finally, classical monocytes (characterized by high CD14 levels and low expression of CD16) showed no change as a fraction of all PBMCs, but, as expected—given the relative reductions in NC and intermediate monocytes—showed a significant increase when expressed as a percentage of all monocytes ($p = 0.04$; **Figure 3.1.9C**, lower right). No differences in viability were detected between *MAPT* pathogenic variant carriers and non-carrier controls (**Figure 3.1.S7**). An alternative gating scheme that enabled the inclusion of additional NC monocytes with lower levels of CD14 expression resulted in very similar findings (**Figure 3.1.S8**). Taken together, our observations are consistent with a bona fide decrease in the frequency of NC monocytes in *MAPT* pathogenic variant carriers, and this alteration does not appear to be secondary to an absolute increase in classical monocyte abundance. Finally, we also asked whether alterations in cell-surface expression of CX3CR1 and CD16 could be detected via flow cytometry on NC monocytes or NK cells. To do so, we measured the median fluorescence intensity (MFI) for CX3CR1 and CD16 in each cell population, but we did not observe a difference between *MAPT* carriers and non-carrier controls (data not shown). We infer from this that the ~50% reduction in transcript abundance for *CX3CR1* and *FCGR3A* is insufficient to alter the cell-surface levels of their encoded proteins. Alternatively, there may be a more general divergence between the dysregulation of these transcripts and their encoded proteins. Future work will test these two possibilities.

3.1.4.10 Analysis of Gene Expression in Human Brain Nonclassical Monocytes

To begin to understand why peripheral NC monocytes are affected in familial tauopathy, we asked whether these cells have access to the human central nervous system (CNS). Indeed, very recent work has suggested an important role for human NC monocytes in the CSF during aging [72], and a recent analysis [23] of publicly available scRNA-seq data [46,73] has also identified these cells in the human temporal cortex. We downloaded and reanalyzed this dataset (accession GSE137444, [46]) and similarly identified a clear NC monocyte population (cluster 13 in our reanalysis; **Figure 3.1.S9A**), thus recapitulating the findings from [23]. This cluster was unique in its high expression of NC monocyte marker genes such as *LYZ*, *S100A8*, *S100A9*, and *FCNI* (marker genes from [23]; **Figure 3.1.S9B**). We next assessed expression levels of the DEGs that we focused on in this work – *CX3CR1*, *FCGR3A*, *TMEM176A/B*, and *C3AR1*. We detected robust expression of all genes except *TMEM176A* in the brain NC monocyte cluster (**Figure 3.1.S9C**). Collectively, these findings suggest that NC monocytes have ready access to the human brain and potentially explain why robust changes in NC monocytes can be observed in the context of tauopathy – that is, they may have direct access to tissue harboring tau pathology. Future functional studies will be needed to determine NC monocytes’ potential to impact the course of tauopathy. In particular, depletion of NC monocytes in mouse models of tauopathy would enable direct testing of the role these cells play in the development of tau neuropathology.

3.1.5 Discussion

This study represents an effort to discover novel, blood-based biomarkers of tauopathy, and—more broadly—to begin to understand the nature of the peripheral leukocyte response to primary tauopathy. To this end, we conducted an unbiased scRNA-seq survey of ~181,000 PBMCs in carriers of pathogenic *MAPT* variants and healthy, non-carrier controls. In doing so, we uncovered

novel, peripheral blood transcriptomic signatures of familial tauopathy at single-cell resolution. In particular, we observed a significant reduction in circulating NC monocytes and numerous DEGs that were particularly enriched in specific myeloid and NK cell clusters. Next, we validated changes in several candidate DEGs selected on the basis of plausible biological relevance. These included *CX3CR1*, *FCGR3A*, and *TMEM176A/B*. In addition, *C3AR1*, although not found to be significantly reduced in *MAPT* carriers by ddPCR, still trended toward reduced expression.

The differential expression of *CX3CR1* observed by scRNA-seq was not only replicated via ddPCR analyses but also confirmed using a publicly available mouse microglia bulk RNA-seq dataset derived from the *MAPT* P301S model. This suggests *CX3CR1* may have potential utility as a peripheral biomarker of tauopathy and warrants further study in larger cohorts. Intriguingly, mouse *Cx3cr1* has been known for over a decade to control the levels of NC monocytes [74–76]. In addition, mouse *Cx3cr1*-mediated control of NC monocyte levels has more recently been suggested to modulate the innate immune response to traumatic brain injury [77]. Collectively, these findings suggest that the reduction in *CX3CR1* expression we observe in *MAPT* pathogenic variant carrier NC monocytes (on a per-cell basis) may be directly related to the concomitant reduction in NC monocyte abundance also observed in *MAPT* carriers.

Quite aside from the role of *Cx3cr1* in controlling circulating NC monocyte levels, *Cx3cr1* exerts a well-established, microglia-mediated modulatory effect on tau neuropathology in mouse models [58–60, 78–84]. In particular, deletion of *Cx3cr1* promotes hallmark neuropathological features of tauopathy including tau hyperphosphorylation and aggregation [58–60, 84]. More-recent studies on induced pluripotent stem cell-derived microglia-like cells have confirmed the importance of *CX3CR1* in regulating human microglial homeostasis [85], consistent with *Cx3cr1*'s established role in promoting the homeostatic microglial phenotype in mice [86]. Given that microglial

homeostasis is dysregulated in tauopathy [87–90], our novel findings—coupled with the well-defined relationship between *CX3CR1* and tauopathy—provide a promising foundation for further investigation of this gene as a peripheral biomarker of tauopathy.

As noted above, we observed a significant reduction in NC monocytes in *MAPT* carriers relative to non-carrier controls. NC monocytes are recruited to sites of vascular damage, infection, or inflammation to patrol the local environment [76]. At these compromised sites, chemoattractant factors are released, and NC monocytes respond through the expression of cognate receptors, including *CX3CR1* [76]. Strikingly, NC monocytes are thought to be reduced in peripheral blood in a variety of neurodegenerative diseases, including ALS [15, 16], PD [14], and the adult-onset, hereditary leukoencephalopathy, ALSP [17]. Conversely, NC monocyte levels may be increased in the CSF in PD, suggesting a possible shift of this monocyte population from blood to CSF in the context of neurodegeneration [91]. Considering these findings, and given that: (i) ALS exists along a spectrum with FTD [92]; (ii) a portion of PD risk is mediated by variation near the *MAPT* locus [93, 94]; and (iii) ALSP can manifest clinically as FTD [18], our finding of reduced NC monocytes in familial tauopathy both strengthens and extends the purported relevance of this monocyte population in neurodegenerative disease.

We observed significantly reduced expression of the canonical NC monocyte marker gene *FCGR3A* as well as significant alterations in two additional NC monocyte marker genes (*VMO1* and *IFITM3*) in *MAPT* pathogenic variant carriers by scRNA-seq. Given that *FCGR3A* expression was reduced not only in NC monocytes but also in NK cells and considering that NC monocyte abundance was simultaneously reduced, we reasoned that bulk assessment of PBMC RNA via ddPCR would be well suited to detect reduced expression of *FCGR3A* in *MAPT* carriers, and, indeed, this is what we found. This cellular phenotype suggests several possibilities. First, the

reduced abundance of NC monocytes and diminished expression of *FCGR3A* on the remaining NC monocytes could reflect migration of mature NC monocytes (that express the highest levels of *FCGR3A*) out of the blood and into another compartment (e.g., CSF or brain). Second, our findings could reflect impaired survival of NC monocytes in tauopathy. Third, lower *FCGR3A* levels might reflect impaired differentiation of NC monocytes from classical or intermediate monocytes (the latter of which express intermediate levels of *FCGR3A*). We favor the first two possibilities, as the latter scenario would be expected to involve accumulation of other classes of monocytes, which we did not observe.

Our differential expression analyses identified cDCs (cluster 14) as the cell type with the highest number of DEGs with absolute LFC > 0.2. Although we did not focus on cDCs for our validation studies, we did find that normalizing NC monocyte abundance to cDC abundance enabled a clear separation of *MAPT* carriers from non-carrier controls. This finding begets the question: what is the biological link connecting NC monocytes to cDCs? Emerging literature has highlighted several intriguing connections between NC monocytes and particular subsets of DCs. For example, a putative DC subpopulation, identified transcriptomically and originally termed DC4 [95], is now considered to probably represent a subset of NC monocytes rather than DCs [47, 96, 97]. In addition, pathogenic variants in *STAT3* have revealed this gene's role in regulating the production of both NC monocytes and the less-abundant cDC population, cDC1 [98]. Although we could detect the rare cDC1 population in our dataset, it became apparent only upon myeloid cell re-clustering (myeloid subcluster 7, marked by *CLEC9A* expression), and these cells were too sparse to enable us to accurately gauge their abundance or use for normalization purposes. On the other hand, it remains unclear precisely how NC monocytes are biologically related to the more-abundant cDC2 population. Nevertheless, normalizing NC monocyte abundance to either cDC2 or

total cDC abundance enabled a clear separation of *MAPT* variant carriers from healthy controls. Future studies in larger cohorts will be required to determine the precise quantification metric for NC monocytes that best differentiates *MAPT* carriers from controls. It will also be important to establish whether this finding extends to sporadic forms of tauopathy; this seems likely given that similar phenomena have been reported in disparate neurodegenerative diseases.

Recently published work indicates robust transcriptional changes in NC monocytes within the CSF during healthy aging and in the context of cognitive impairment [72]. An additional recent finding [23], which we confirm here (**Figure 3.1.S9**), indicates that NC monocytes can also be found in the human temporal cortex. Given that the precise role of human NC monocytes in health and disease is still largely unknown [76], it remains unclear what a reduction in peripheral NC monocytes means vis-à-vis tauopathy. If the observed reduction in peripheral NC monocytes is accompanied by a corresponding increase in CNS NC monocytes—and if they contribute to heightened neuroinflammation—they are likely to play a net detrimental role in neurodegeneration. However, much work remains to be done to test these possibilities.

Beyond myeloid cells, our work also highlights a potentially novel role for NK cells in primary tauopathy. In particular, NK cells had a large number of DEGs with LFC > 0.2, and our findings implicating *CX3CR1* expression not only in NC monocytes but also in NK cells as a candidate peripheral biomarker of familial tauopathy is complemented by recent research suggesting an important yet previously unappreciated role for NK cells in a mouse model of AD [99]. In addition, NK cell recruitment to the CNS has been observed in ALS as well as ALS models [100]. Collectively, the available data suggest a detrimental role for NK cells in ALS and AD. More broadly, in the experimental autoimmune encephalomyelitis model of multiple sclerosis, NK cell migration into the CNS is mediated in part by CX3CR1-dependent recruitment [101, 102],

suggesting that the differential expression of *CX3CR1* in NK cells that we observed in our study could plausibly affect NK cell recruitment to the CNS in primary tauopathy.

Reliable biomarkers can improve diagnostic acumen and enable elucidation of specific forms of neuropathology underlying clinical dementia syndromes. For example, examination of brain structure and function via neuroimaging is a powerful method for the determination of neurodegenerative disease etiology. The use of positron emission tomography (PET) imaging, in particular, with radiotracers that bind to aggregated forms of tau has facilitated the in vivo detection of tau neuropathology in individuals with AD (reviewed in [19,103,104]). However, tau-PET tracers do not bind strongly to most forms of FTLD-tau pathology and may exhibit off-target binding in individuals with FTLD-TDP pathology [103, 104]. Alternatively, the use of CSF- and blood-based protein biomarkers holds great promise for AD [19, 20, 105] and FTD [21, 106], although in the case of FTD, we still cannot discriminate between underlying FTLD-tau and -TDP pathology. Important limitations apply to several of these methods. In particular, PET imaging is costly and available only at specialized medical centers, and CSF collection requires invasive lumbar puncture. In contrast to these methods, peripheral blood biomarkers are easy to collect and, when coupled with analytic techniques such as ddPCR, may eventually enable low-cost alternatives to today's better-developed biomarkers.

Despite numerous advances described above, our study has several important limitations. First, due to the significant expense of scRNA-seq and our desire to capture a relatively large number of PBMCs per individual, we necessarily used a small cohort for this study. We also opted to confirm gene expression findings via an orthogonal technique using bulk PBMC RNA measurements in essentially the same cohort that was used for the scRNA-seq analysis. While we employed this strategy to minimize the possibility that technical artifacts drove discovery of the candidate genes

we characterized, it will be important to evaluate the generalizability of our findings in larger cohorts. Larger cohorts will also enable us to determine whether any of the dysregulated genes identified here are differentially sensitive to pathogenic *MAPT* variants that affect splicing, relative to those that do not, as potentially suggested for *CX3CR1* (**Figure 3.1.5B**). In addition, future work that measures absolute numbers of NC monocytes in familial tauopathy would be a valuable addition to the field, as the scRNA-seq and flow cytometry data reported here show relative reductions in NC monocyte frequency in *MAPT* pathogenic variant carriers. Another issue to be resolved in future studies is the biological significance of the upregulation of ribosomal and mitochondrial transcripts in myeloid and other cell types observed here in familial tauopathy. For example, do these changes in gene expression in NC monocytes primarily reflect increased fragility and/or decreased survival of these cells in tauopathy, or are they reflective of more specific changes in ribosomal and mitochondrial biology?

As alluded to above, it will also be important to determine which peripheral immune changes are conserved between familial tauopathy and diverse forms of sporadic primary and secondary tauopathy. In addition, given the complex temporal trajectories of brain myeloid responses in tauopathy [107], future research on large cohorts of presymptomatic *MAPT* pathogenic variant carriers will be needed to determine which peripheral changes observed here occur prior to disease onset. Finally, it will be important to ascertain whether the peripheral leukocyte changes discovered here are reflected by parallel changes in brain myeloid cells in individuals with tauopathy.

3.1.6 Conclusions

To our knowledge, this is the first scRNA-seq study of peripheral blood cells in primary tauopathy. Beyond our initial discoveries, we validated a handful of DEGs via an orthogonal technique, ddPCR. In particular, we have connected longstanding observations from mouse models regarding microglial *Cx3cr1* and tau neuropathology to reduced *CX3CR1* in peripheral leukocytes in individuals with familial tauopathy. Moreover, we discovered a significant reduction in the abundance of circulating NC monocytes, a cell type that is similarly reduced in several additional neurodegenerative diseases. We also discovered large numbers of DEGs in NK cells, including *CX3CR1*, which is thought to be involved in recruitment of NK cells to the CNS. Further studies are now required to investigate the generalizability of our findings through replication in larger cohorts and extension to other tauopathies and related neurodegenerative diseases. Analogous studies of PBMCs in *GRN* pathogenic variant carriers and *C9orf72* hexanucleotide repeat expansion carriers should enable the discovery of peripheral biomarkers of FTLD-TDP. Ultimately, comparative studies should clarify the role of peripheral immune responses in distinct proteinopathies and enable discovery of novel peripheral biomarkers that can successfully discriminate between tau and TDP-43 neuropathology, providing critical new tools for diagnostics and clinical trials.

3.1.7 Article Information

3.1.7.1 Declarations

Ethics approval and consent to participate. All participants or their surrogates provided written informed consent prior to study participation; all aspects of the studies described here were approved by the UCSF institutional review board (committee number 10-03946). The research described herein conformed to the principles of the Helsinki Declaration.

Availability of data and materials. The scRNA-seq dataset described here has been uploaded to the FAIR Data Sharing Portal within the Alzheimer's Disease Workbench, which is supported by the Alzheimer's Disease Data Initiative.

Competing interests. JSY serves on the scientific advisory board for the Epstein Family Alzheimer's Research Collaboration. ALB has served as a consultant for Aeovian, AGTC, Alector, Arkuda, Arvinas, AviadoBio, Boehringer Ingelheim, Denali, GSK, Life Edit, Humana, Oligomerix, Oscotec, Roche, Transposon, TrueBinding and Wave. The remaining authors declare that they have no competing interests.

Funding. Funding to JSY and for this project comes from NIH-NIA K01AG049152, R01AG062588, R01AG057234, P30AG062422, P01AG019724; NIH-NINDS U54NS123985; the Rainwater Charitable Foundation; the Larry L. Hillblom Foundation; the Bluefield Project to Cure Frontotemporal Dementia; the Alzheimer's Association; the Global Brain Health Institute; the French Foundation; and the Mary Oakley Foundation. VES is supported by R01AG052496. SEL is supported by R01AG058233. ALB receives research support from NIH (U19AG063911, R01AG038791, R01AG073482), the Tau Research Consortium, the Association for Frontotemporal Degeneration, Bluefield Project to Cure Frontotemporal Dementia, Corticobasal

Degeneration Solutions, the Alzheimer's Drug Discovery Foundation, and the Alzheimer's Association, and has received research support from Biogen, Eisai and Regeneron. The content of this publication is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Authors' contributions. DWS initiated the study; analyzed the scRNA-seq, ddPCR, and flow cytometry data; and wrote the manuscript. CWS analyzed the scRNA-seq data and contributed to writing the manuscript. TPJ isolated PBMCs and performed ddPCR and flow cytometry experiments with DWS. LWB analyzed scRNA-seq data and edited the manuscript. VES, SEL, KPR, HJR, ALB, WWS, and BLM were involved in research participant recruitment and clinical characterization. EGG was involved in early stages of the scRNA-seq analysis and provided guidance on study design. JSY initiated the study, oversaw the analysis, and edited the manuscript. All authors read and approved the final manuscript.

Acknowledgements. We thank the UCSF Genomics CoLab and IHG staff, including Catherine Chu, for expert technical assistance. In addition, we thank UCSF CAT staff, including Tyler Miyasaki, for expert advice on the use of ddPCR equipment. Finally, we are grateful to Collin Spencer in the UCSF Department of Neurology for his help and expert advice related to flow cytometry. The cartoon in **Figure 3.1.4D** was created with BioRender.com.

3.1.8 Figures and Tables

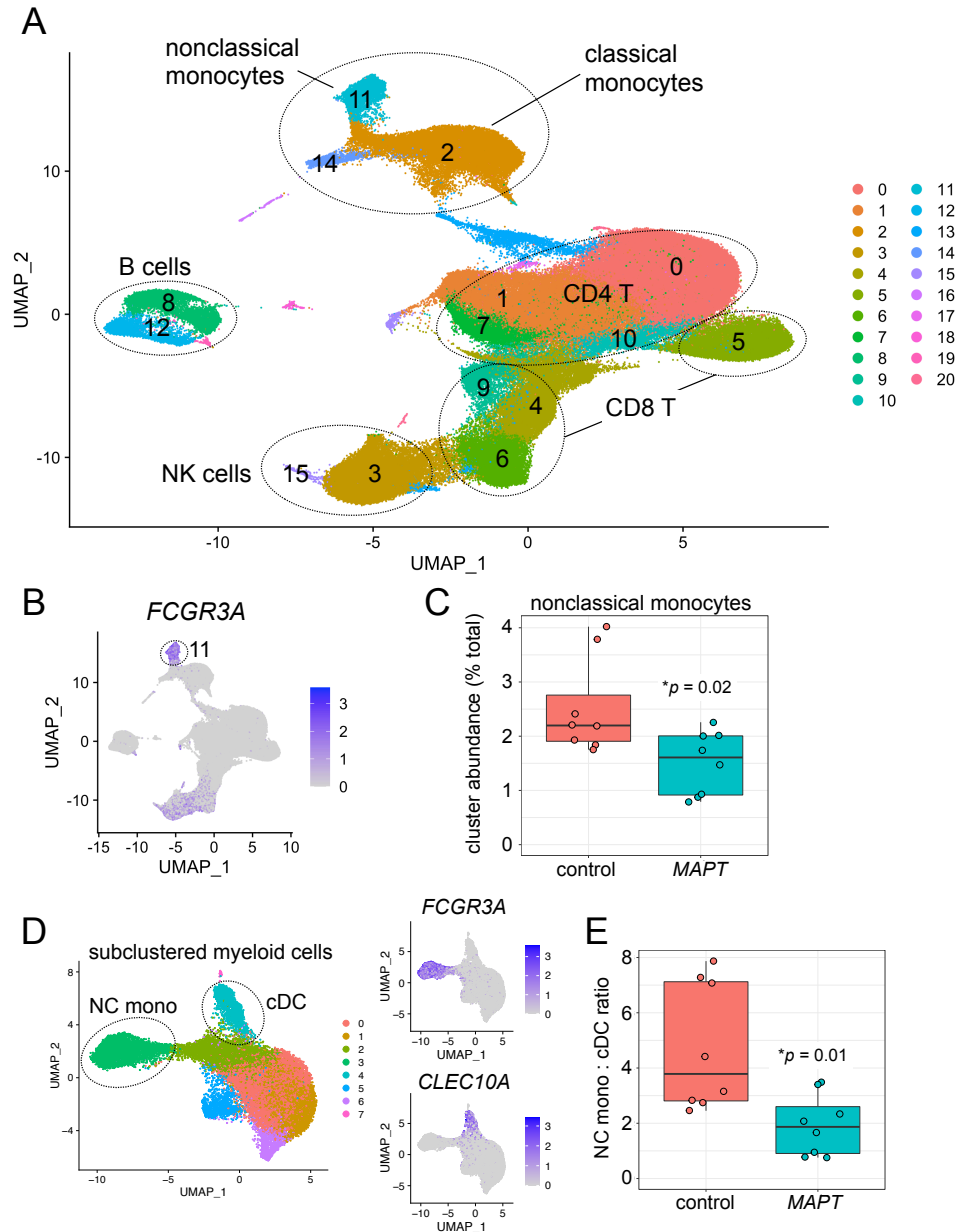


Figure 3.1.1 Single-cell RNA-seq reveals reductions in NCM in *MAPT* pathogenic variant carriers

A) Two-dimensional UMAP plot of ~181,000 PBMCs from *MAPT* variant carriers and non-carrier controls, colored by cluster identity. Major cell types are labeled within the plot. B, C) Cluster 11, marked by high *FCGR3A* expression and identified as NC monocytes, was significantly reduced in *MAPT* carriers ($p = 0.02$; data are expressed as percentage of total PBMCs for each sample). D) Myeloid cells (clusters 2, 11, 14) were subset and re-clustered. NC monocyte and cDC2 subclusters were identified by *FCGR3A* and *CLEC10A* expression, respectively (D, right). E) The ratio of NC monocytes to cDC2 was significantly reduced in *MAPT* variant carriers ($p = 0.01$)

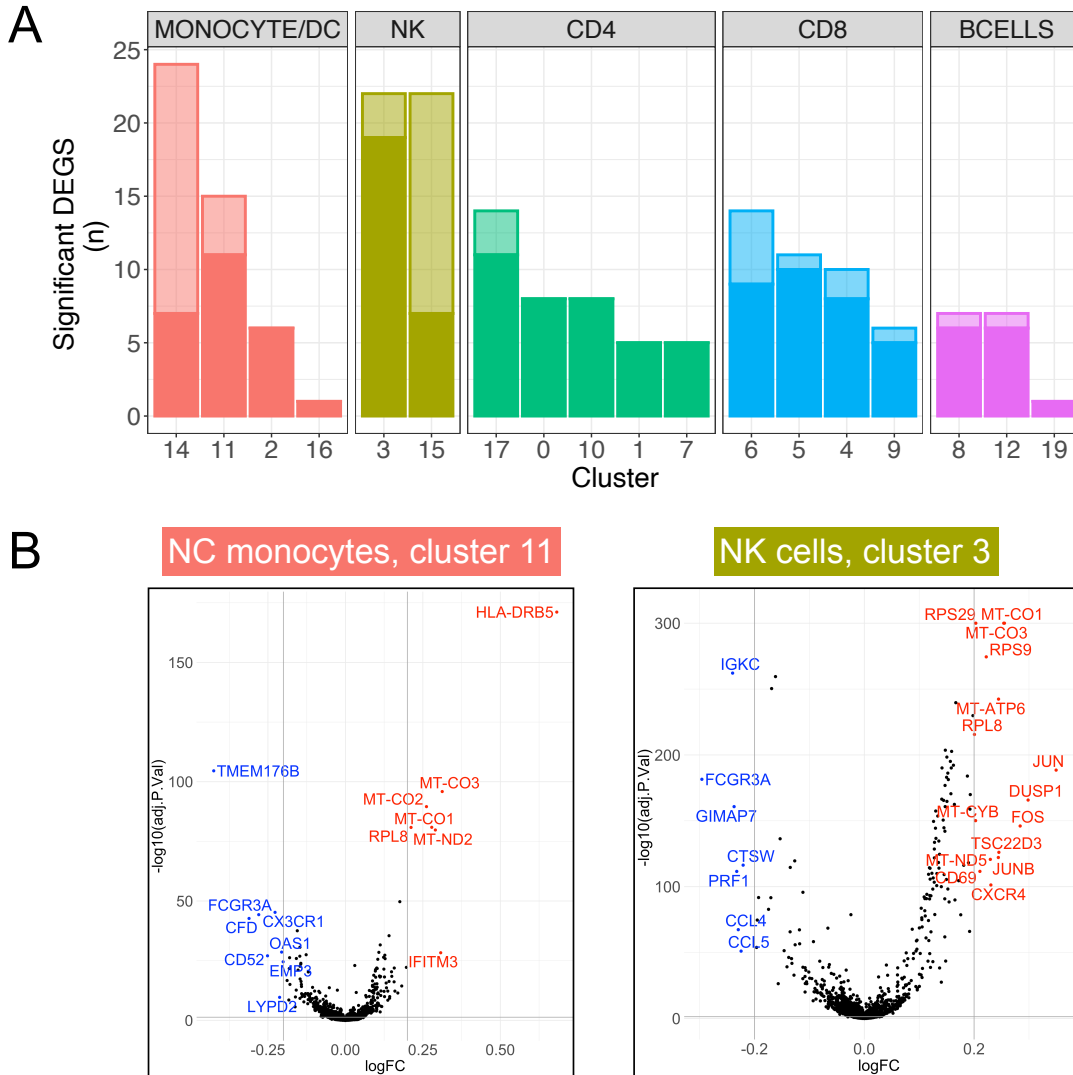


Figure 3.1.2 Differential expression in *MAPT* variant carriers by cell cluster

A) Clusters are grouped by cell type and ranked by the number of DEGs with $p_{FDR} < 0.05$ and absolute LFC > 0.2 . Differential expression was determined in *MAPT* variant carriers relative to non-carrier controls while covarying for age, sex, and scRNA-seq batch. Solid-colored portions of the bars indicate DEGs shared by at least one other cluster, while the translucent portions indicate DEGs unique to a given cluster. cDCs (cluster 14), NK cells (clusters 3 and 15), and NC monocytes (cluster 11) had the highest numbers of DEGs with absolute LFC > 0.2 . B) Volcano plots of the NC monocyte cluster and major NK cell cluster; DEGs with absolute LFC > 0.2 are labeled in blue (down-regulated) or red (up-regulated). Several NK cell DEGs (right) with $-\log_{10}(p_{FDR})$ values > 300 were set to 300 for visualization purposes.

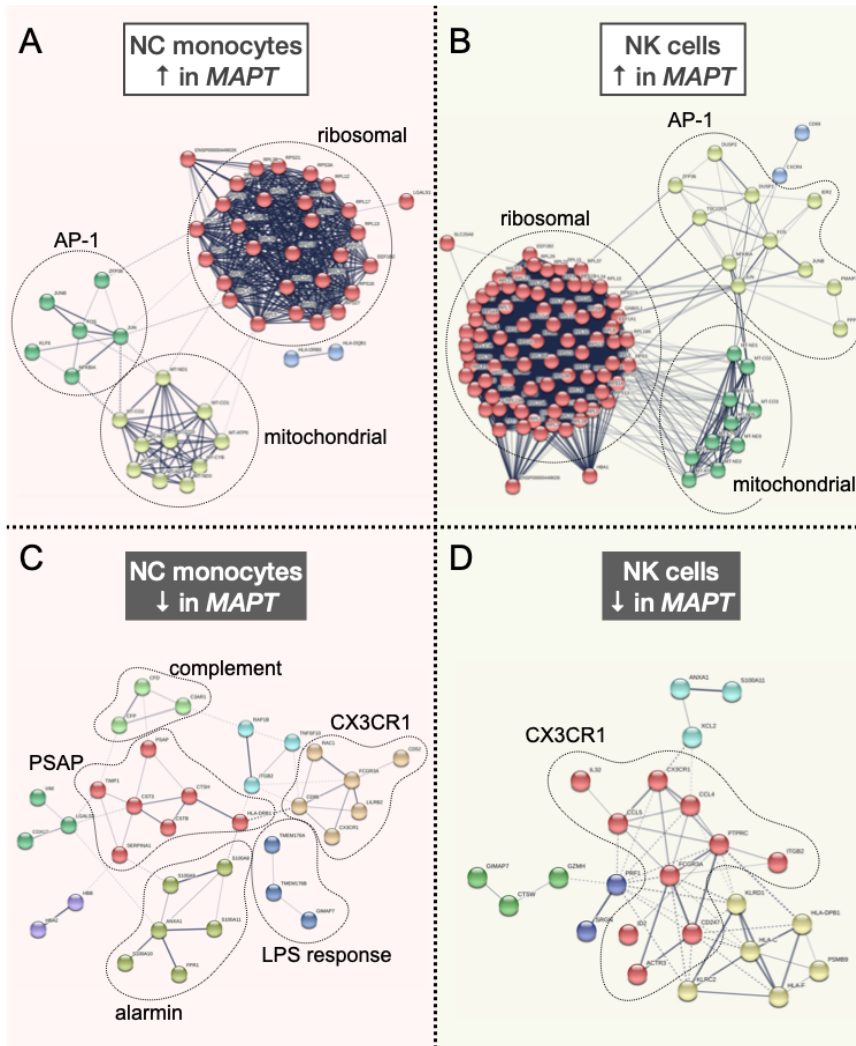


Figure 3.1.3 STRING interactions in NCM and NK cell differentially expressed genes. A, B) Up-regulated DEGs in NC monocytes and NK cells had similar overall network architecture, with large ribosomal and mitochondrial modules, and a third module containing members of the AP-1 transcription factor, among other genes. C) The down-regulated DEGs in NC monocytes contained a module featuring *CX3CR1* and *FCGR3A* as members, in addition to modules harboring genes involved in LPS response, the alternative and classical complement cascades, the *SI100* alarmin molecules, as well as *PSAP*. D) Down-regulated DEGs in NK cells also featured a large module featuring both *CX3CR1* and *FCGR3A*. All DEGs with $p_{FDR} < 0.05$ and absolute LFC > 0.1 from clusters 3 and 11 were input into the STRING database as described in the Methods section. Modules are colored according to the results of MCL clustering.

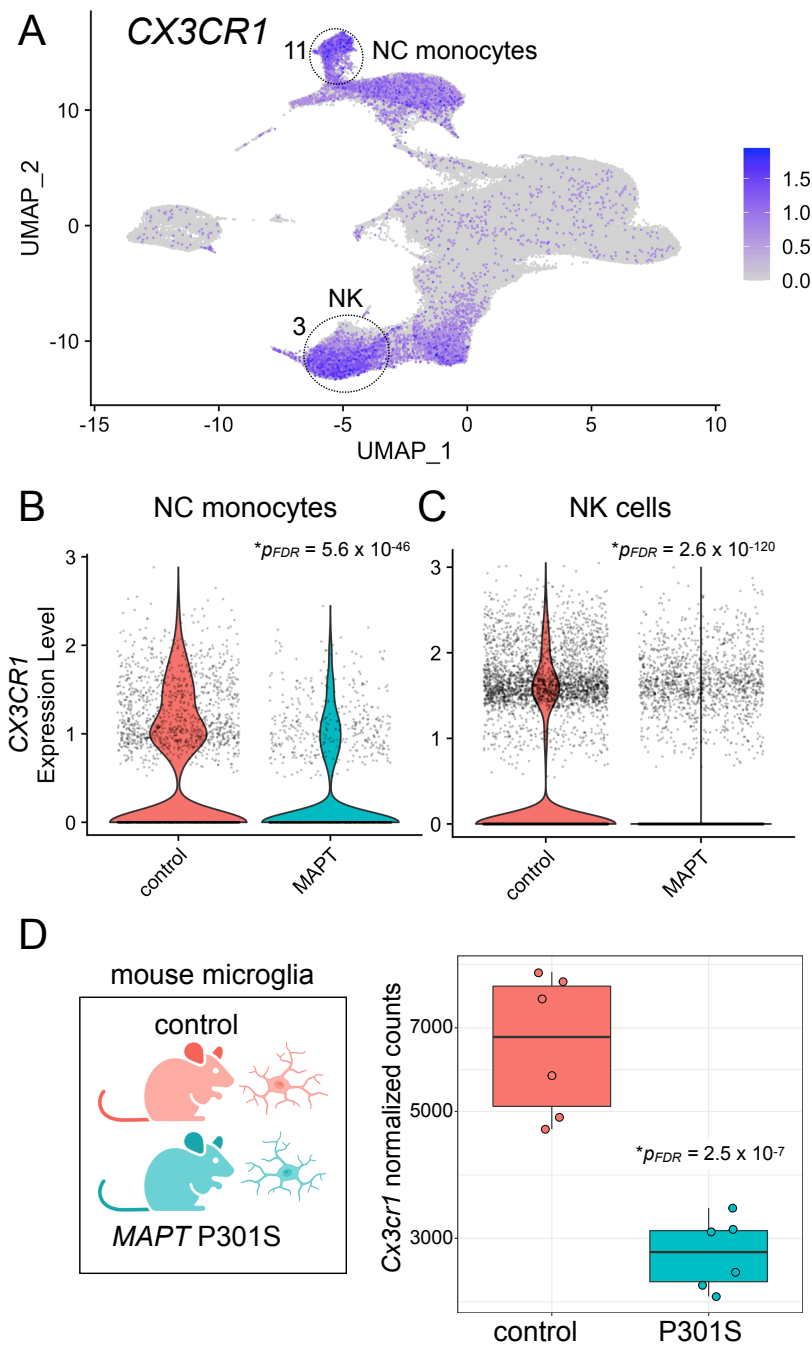


Figure 3.1.4 *CX3CR1* expression is reduced in peripheral myeloid and lymphoid cells in familial tauopathy

A) *CX3CR1* is robustly expressed in both NC monocytes (cluster 11) and NK cells (cluster 3). NC monocytes (B; $p_{FDR} = 5.6 \times 10^{-46}$) and NK cells (C; $p_{FDR} = 2.6 \times 10^{-120}$) both show significantly reduced expression of *CX3CR1* in *MAPT* pathogenic variant carriers. D) Reanalysis of publicly available bulk RNA-seq data from mouse hippocampal CD11b⁺ microglia demonstrated a significant reduction ($p_{FDR} = 2.5 \times 10^{-7}$) of *Cx3cr1* in the *MAPT* P301S model.

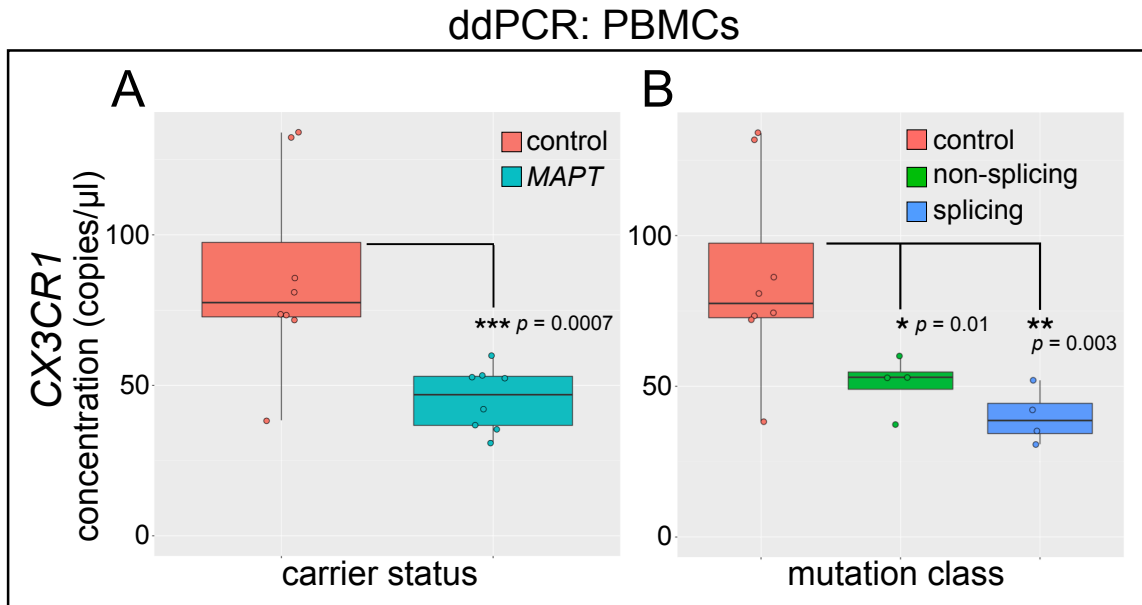


Figure 3.1.5 Confirmation of reduced *CX3CR1* expression in *MAPT* variant carrier PBMCs via ddPCR

RNA was isolated from PBMCs from *MAPT* variant carriers and healthy, non-carrier controls; gene expression was determined by RT-ddPCR. A) *CX3CR1* was significantly reduced in *MAPT* carrier PBMCs ($p = 0.0007$) relative to controls. B) Separation of samples according to *MAPT* variant class (non-splicing and splicing) reveals that *CX3CR1* was significantly reduced in both groups, relative to controls (non-splicing, $p = 0.01$; splicing, $p = 0.003$).

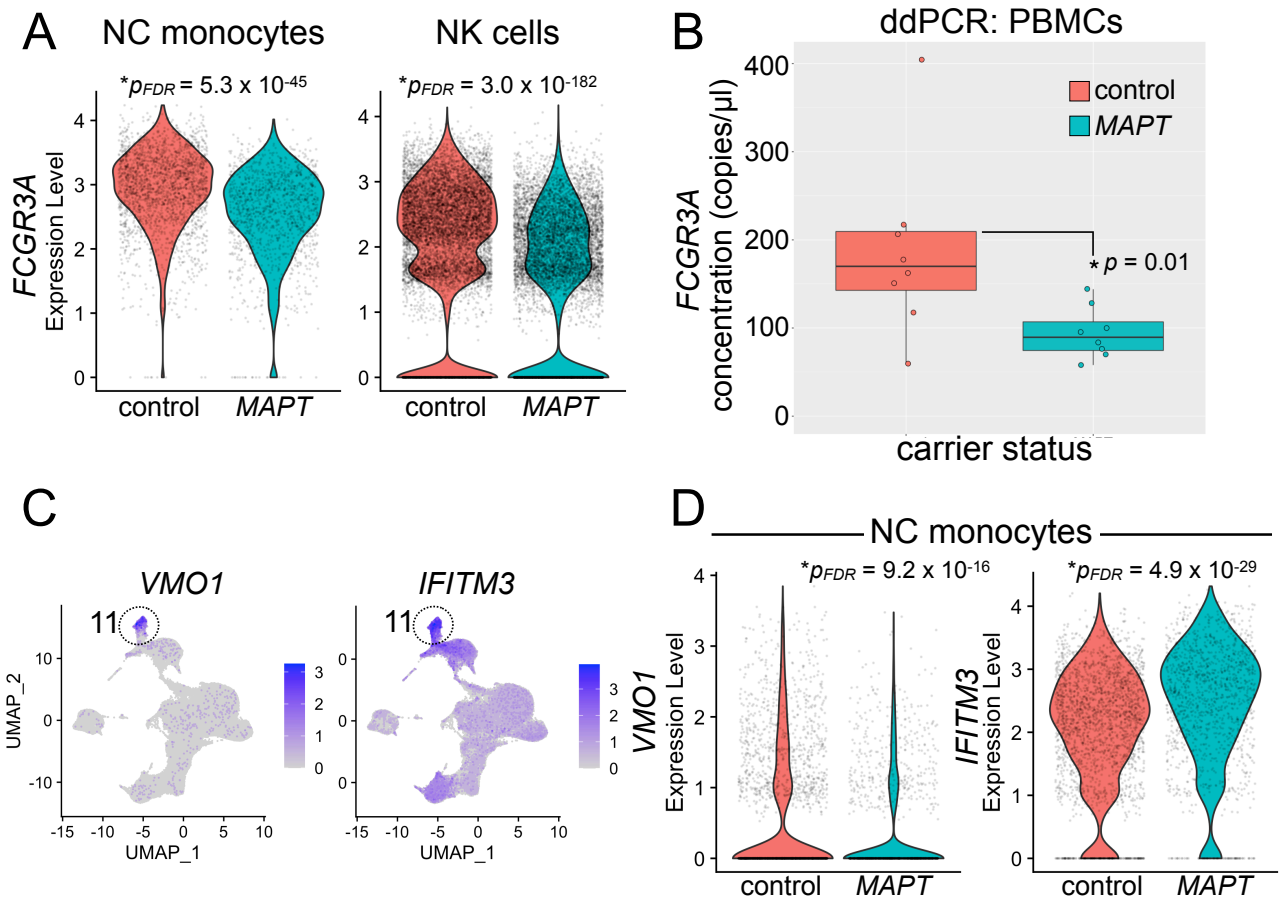


Figure 3.1.6 Analysis of nonclassical monocyte marker genes in *MAPT* variant carriers

A) *FCGR3A*, the NC monocyte marker gene encoding CD16, is robustly expressed not only in NC monocytes but also in NK cells. *FCGR3A* is significantly reduced in both NC monocytes (left; $p_{FDR} = 5.3 \times 10^{-45}$) and NK cells (right; $p_{FDR} = 3.0 \times 10^{-182}$) in *MAPT* pathogenic variant carriers. B) ddPCR confirmed a reduction in *FCGR3A* reduction in *MAPT* variant carrier PBMCs ($p = 0.01$). C) Additional genes expressed specifically (*VMO1*, left) or enriched in (*IFITM3*, right) NC monocytes showed significant alterations D) in *MAPT* variant carrier NC monocytes. D) *VMO1* (left) was significantly reduced ($p_{FDR} = 9.2 \times 10^{-16}$), while *IFITM3* (right) was significantly increased ($p_{FDR} = 4.9 \times 10^{-29}$) in *MAPT* carriers.

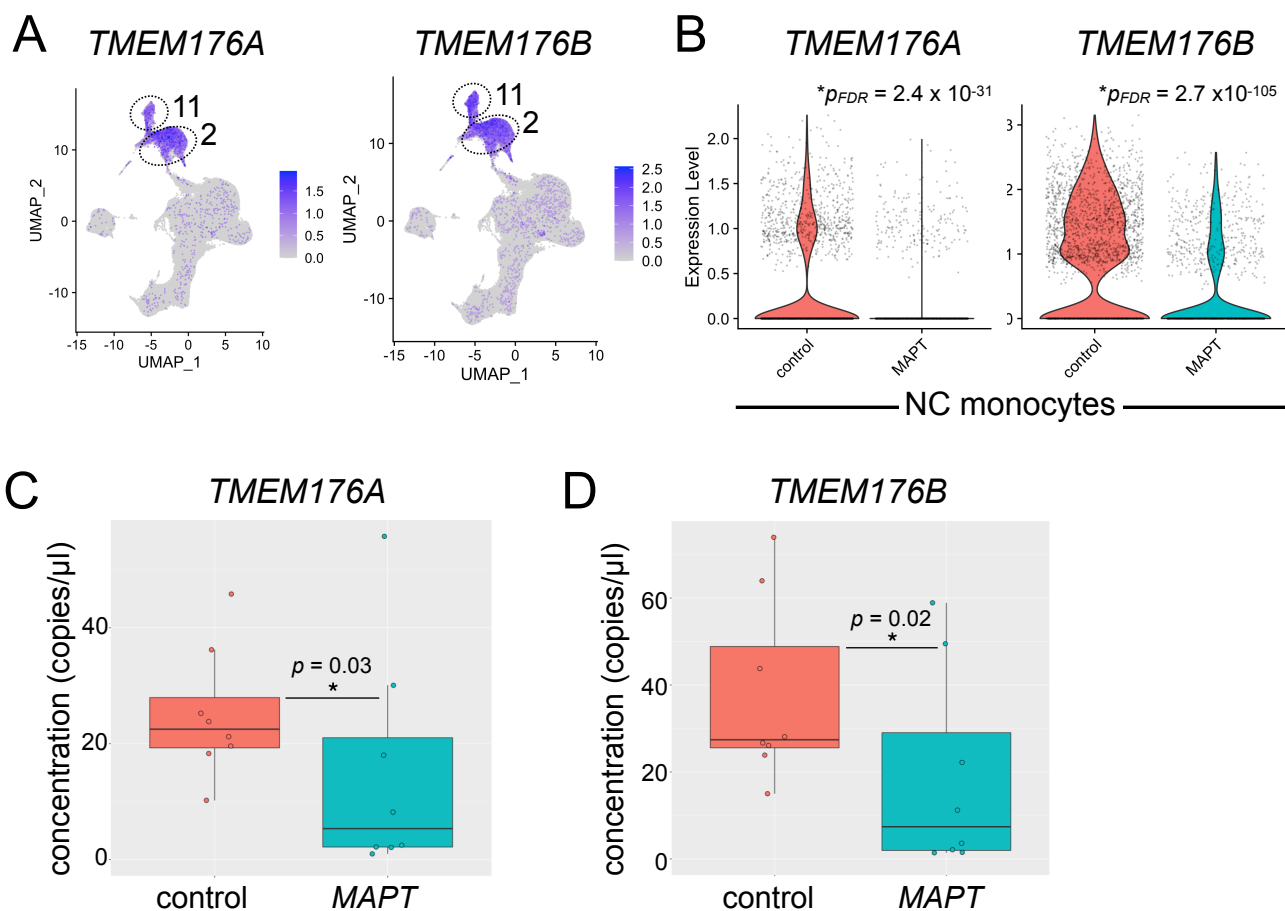


Figure 3.1.7 Analysis of *TMEM176A/B* in *MAPT* pathogenic variant carriers
 A) *TMEM176A/B* are highly expressed in both classical (cluster 2) and NC (cluster 11) monocytes. B) *TMEM176A/B* are significantly reduced in NC monocytes (*TMEM176A*, $p_{FDR} = 2.4 \times 10^{-31}$; *TMEM176B*, $p_{FDR} = 2.7 \times 10^{-105}$) from *MAPT* carriers. These genes were similarly reduced in *MAPT* carrier classical monocytes (cluster 2; Additional file 5: Table S4). The reduction in *TMEM176A* C) and *TMEM176B* D) in *MAPT* variant carriers was confirmed using bulk PBMC RNA and ddPCR (*TMEM176A*, $p = 0.03$; *TMEM176B*, $p = 0.02$).

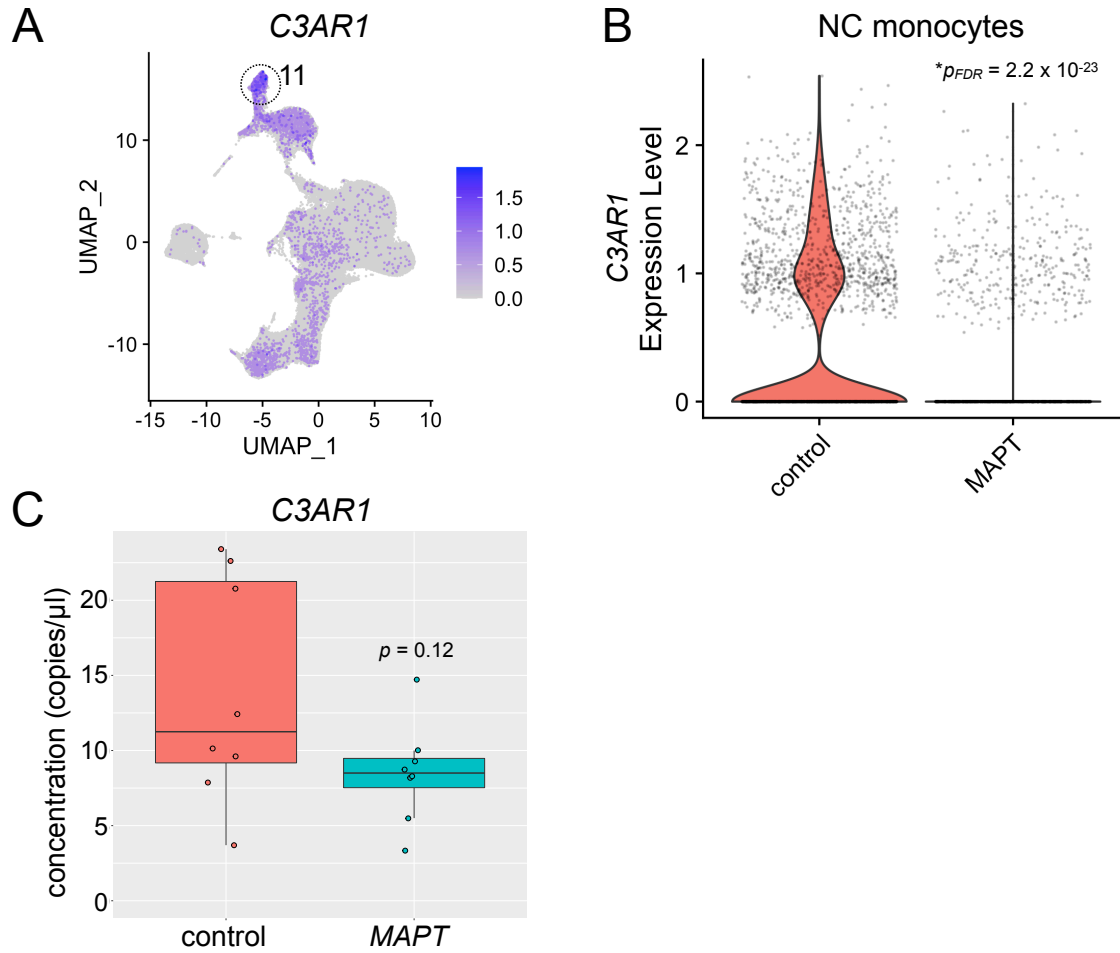


Figure 3.1.8 Potential dysregulation of *C3AR1* in *MAPT* pathogenic variant carriers
 A) *C3AR1* expression was enriched in the NC monocyte cluster (11). B) *C3AR1* expression in NC monocytes was significantly reduced in *MAPT* variant carriers ($*p_{FDR} = 2.2 \times 10^{-23}$). C) ddPCR analysis of PBMC RNA revealed a trend toward reduced expression of *C3AR1* in *MAPT* variant carriers which did not reach significance ($p = 0.12$).

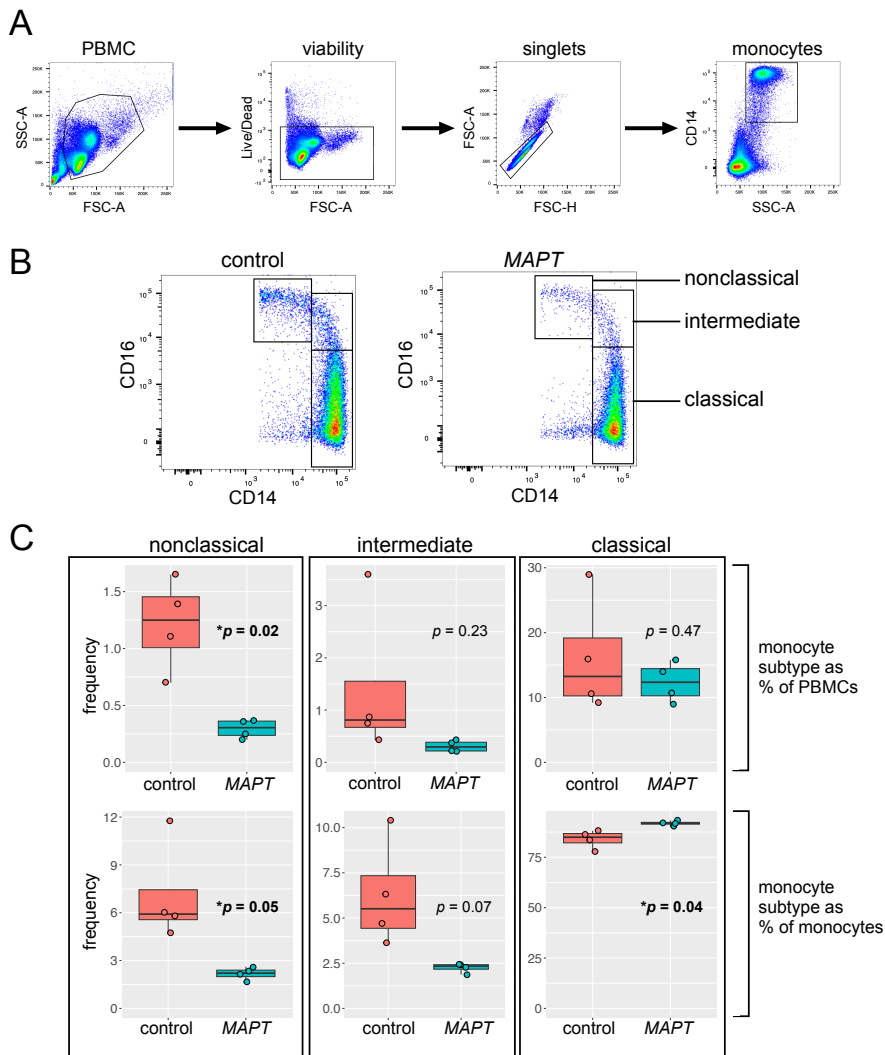


Figure 3.1.9 Validation of the Reduction in Nonclassical Monocytes via Flow Cytometry
 A) Gating scheme for identification and analysis of monocyte subtypes. PBMCs were gated as follows: debris was excluded, non-viable cells were excluded, then doublets were excluded. Next, monocytes were gated based on their high side scatter and CD14 expression. B) Monocyte subtypes were gated based on their characteristic CD14 and CD16 expression, with classical monocytes having high CD14 expression and low CD16 expression, intermediate monocytes having high CD14 expression and moderate-to-high CD16 expression, and NC monocytes having low CD14 expression and high CD16 expression. C) Quantification of the frequency of NC (left), intermediate (center), and classical monocytes (right), either as a percentage of PBMCs (top row) or all monocytes (bottom row). NC monocytes were reduced in *MAPT* pathogenic variant carriers as a fraction of PBMCs (upper left, $p = 0.02$) and as a fraction of monocytes (lower left, $p = 0.05$). Intermediate monocytes (center) showed a trend toward reduction relative to both PBMCs and monocytes. Classical monocytes (right) showed no change as a fraction of PBMCs but were significantly increased in *MAPT* pathogenic variant carriers as a fraction of all monocytes.

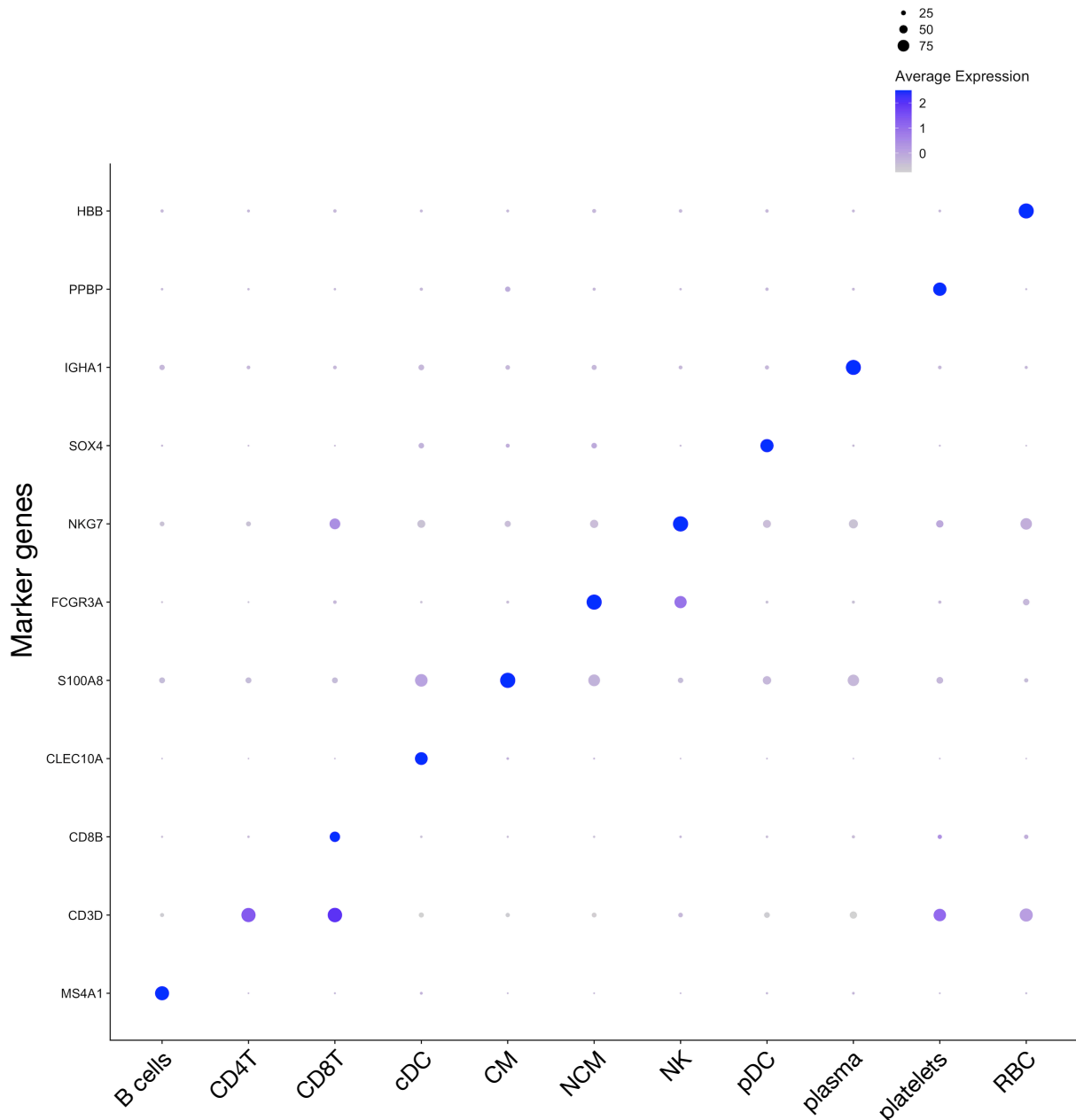


Figure 3.1.S1 Annotation of cell types from cluster markers

Cell types were annotated using a combination of canonical marker genes and additional markers identified via Seurat's FindMarkers function. Clusters were assigned to the following cell-type categories: B cells (clusters 8, 12, 19), CD4+ T cells (CD4T; clusters 0, 1, 7, 10, 17), CD8+ T cells (CD8T; clusters 4, 5, 6, 9), cDC (cluster 14), classical monocytes (CM; cluster 2), NC monocytes (NCM; cluster 11), NK cells (clusters 3 and 15), plasmacytoid dendritic cells (pDC; cluster 16), plasma cells (cluster 18), and contaminating platelets (cluster 13) and red blood cells (RBC; cluster 20). Expression of marker genes is shown using Seurat's DotPlot visualization, with dot size indicating the percentage of cells in a given category having detectable expression of a given marker gene and color indicating the average expression level.

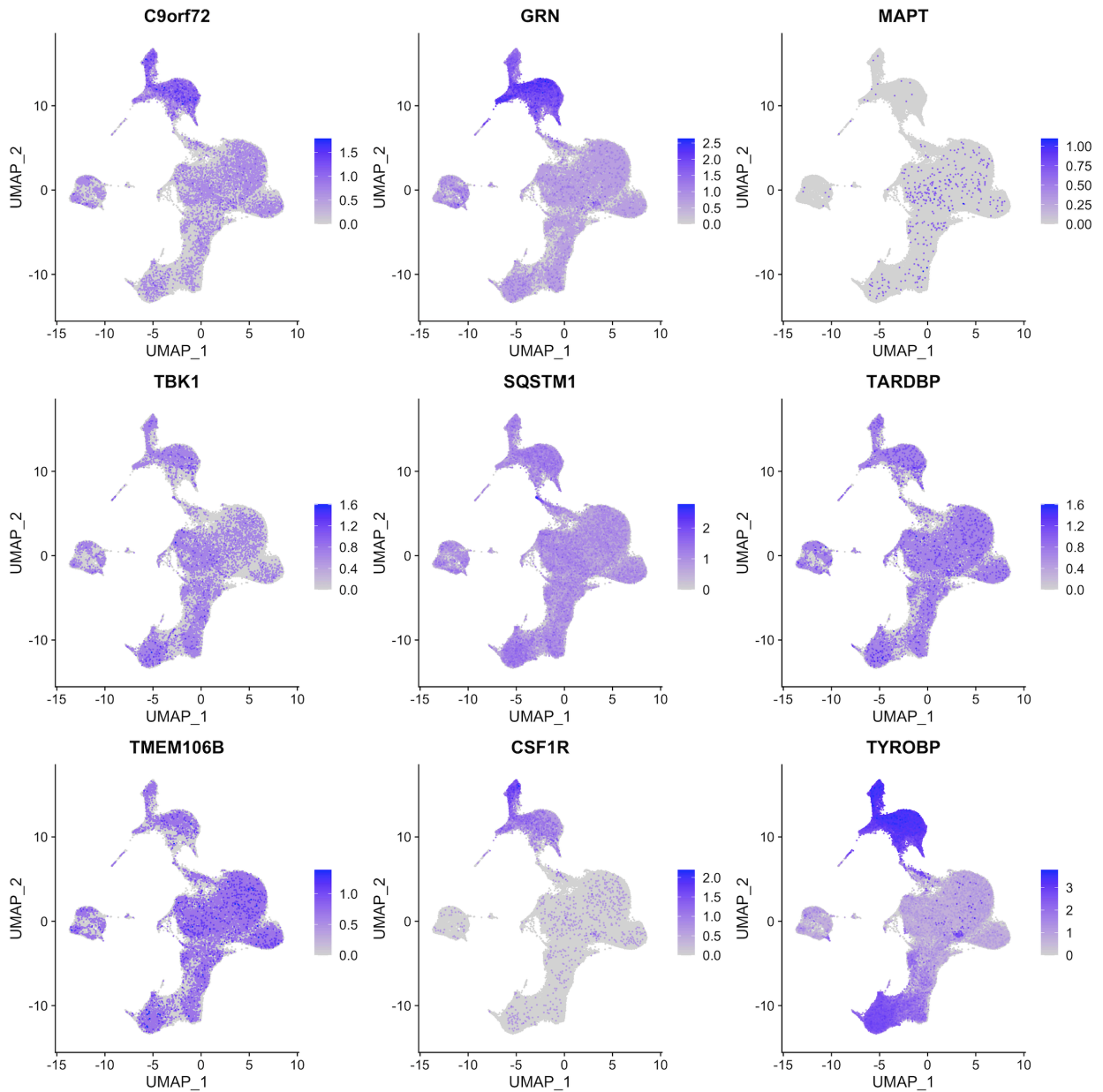


Figure 3.1.S2 Expression of FTD-associated genes in PBMCs

The expression level of nine FTD-associated genes (C9orf72, GRN, MAPT, TBK1, SQSTM1, TARDBP, TMEM106B, CSF1R, TYROBP) across all PBMCs in the dataset is depicted. Some genes displayed widespread expression across many PBMC cell types (TBK1, SQSTM1, TARDBP, TMEM106B), while others showed enriched expression in myeloid clusters (C9orf72, GRN, CSF1R, TYROBP). In contrast to other FTD-associated genes, MAPT expression was very rarely detected in PBMCs

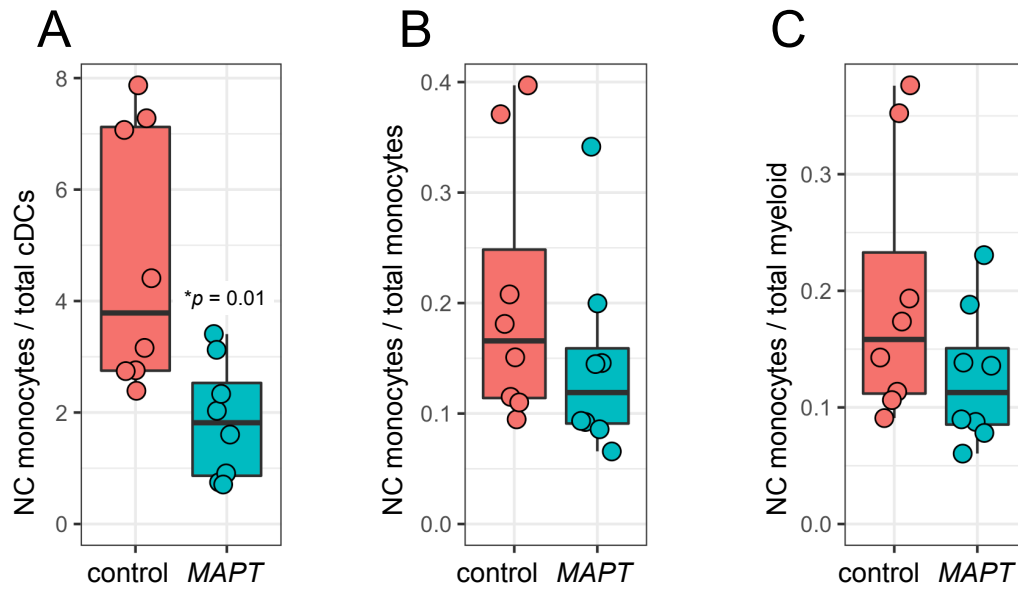


Figure 3.1.S3 Nonclassical monocyte abundance relative to other myeloid populations
 A) MAPT pathogenic variant carriers showed a significant reduction ($p = 0.01$) in the ratio of NC monocytes to total cDCs (cDC1 + cDC2). B, C) In contrast, the ratio of NC monocytes to total monocytes B) or total myeloid cells C) was not significantly reduced in MAPT variant carriers compared to non-carrier controls.

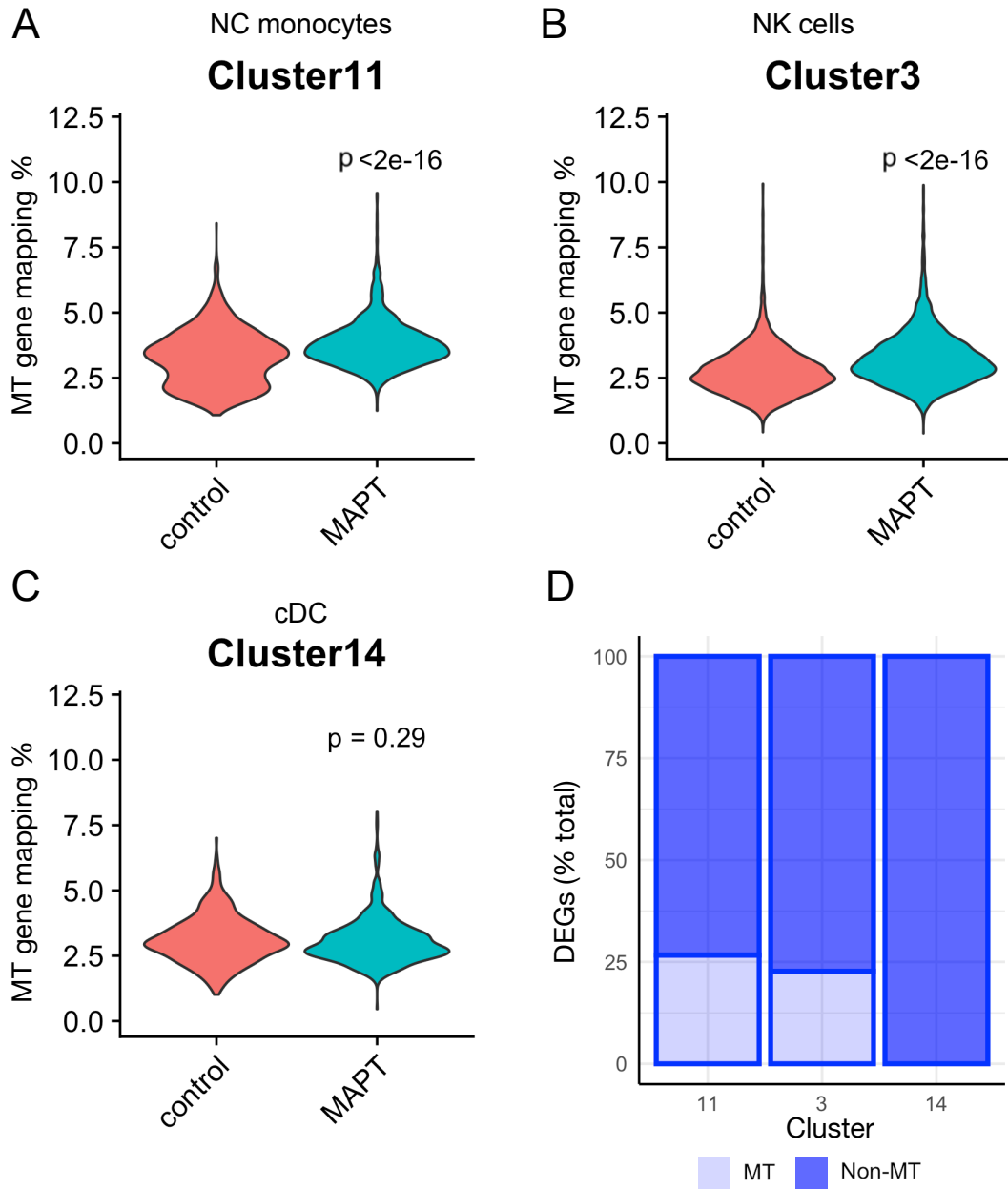


Figure 3.1.S4 Mitochondrial genome mapping percentages and mitochondrial DEGs in selected clusters

A, B) Mitochondrial mapping percentage was subtly but significantly increased ($p < 2 \times 10^{-16}$) in MAPT pathogenic variant carriers in clusters 11 (NC monocytes; A) and 3 (NK cells; B). C) On the other hand, mitochondrial mapping percentage was not significantly increased in cluster 14 (cDC). D) Clusters 11 and 3 both harbored appreciable fractions of mitochondrial DEGs relative to total DEGs, while cluster 14 did not contain mitochondrial DEGs

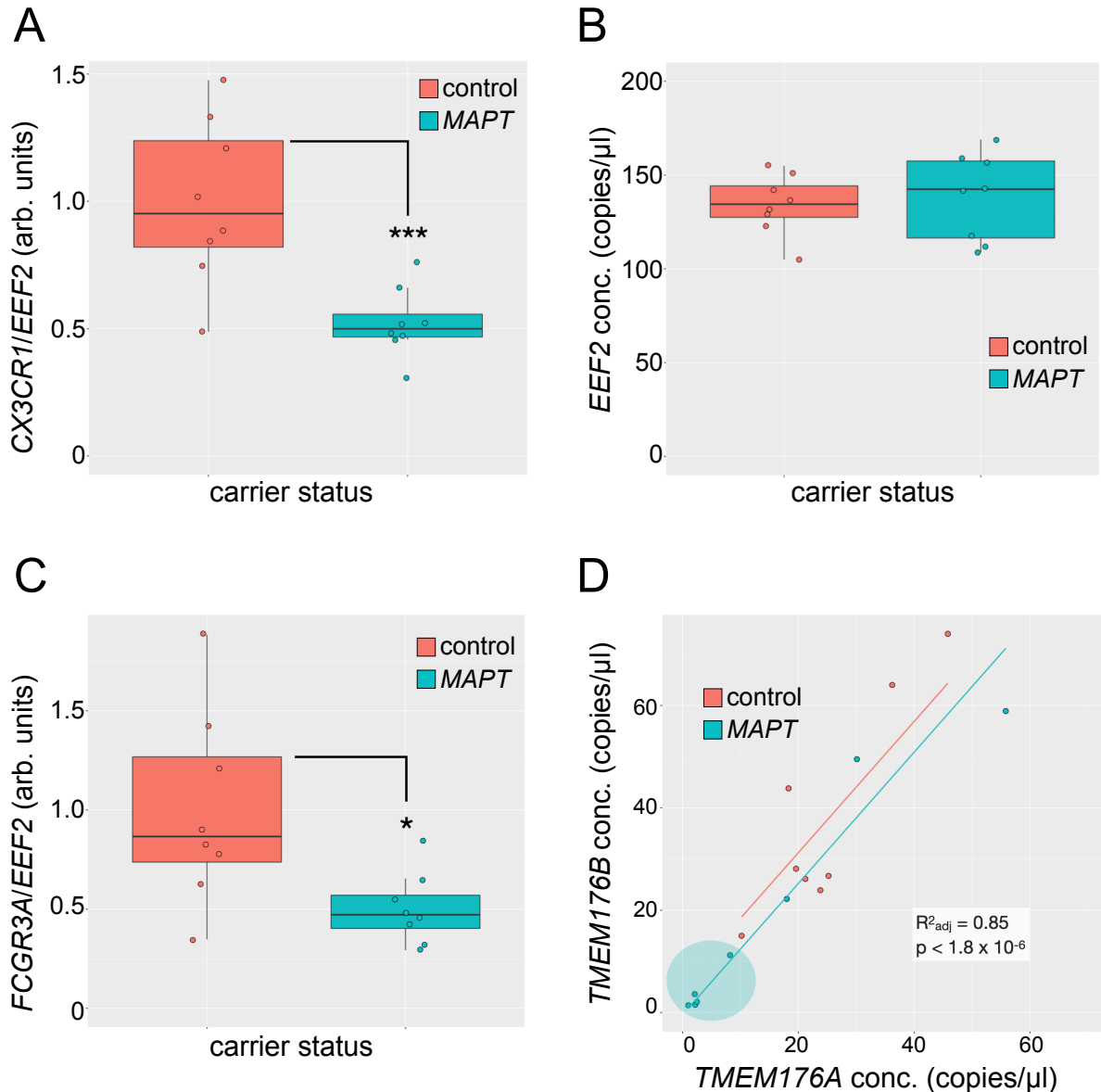


Figure 3.1.S5 Additional droplet digital PCR analysis

A) Normalization of the CX3CR1 concentration to that of reference gene EEF2 produced results similar (***, $p = 0.0006$) to what is depicted in Fig. 5A. B) EEF2 absolute concentration values were similar between MAPT pathogenic variant carriers and non-carrier controls. C) Normalization of the FCGR3A concentration to that of EEF2 did not alter the results depicted in Fig. 6B (*, $p = 0.01$). D) TMEM176A/B levels were closely associated with one another ($R^2_{adj} = 0.85$, $p < 1.8 \times 10^{-6}$). A subset of MAPT variant carriers (5 of 8) showed levels of TMEM176A/B that were lower than what was observed for any non-carrier control.

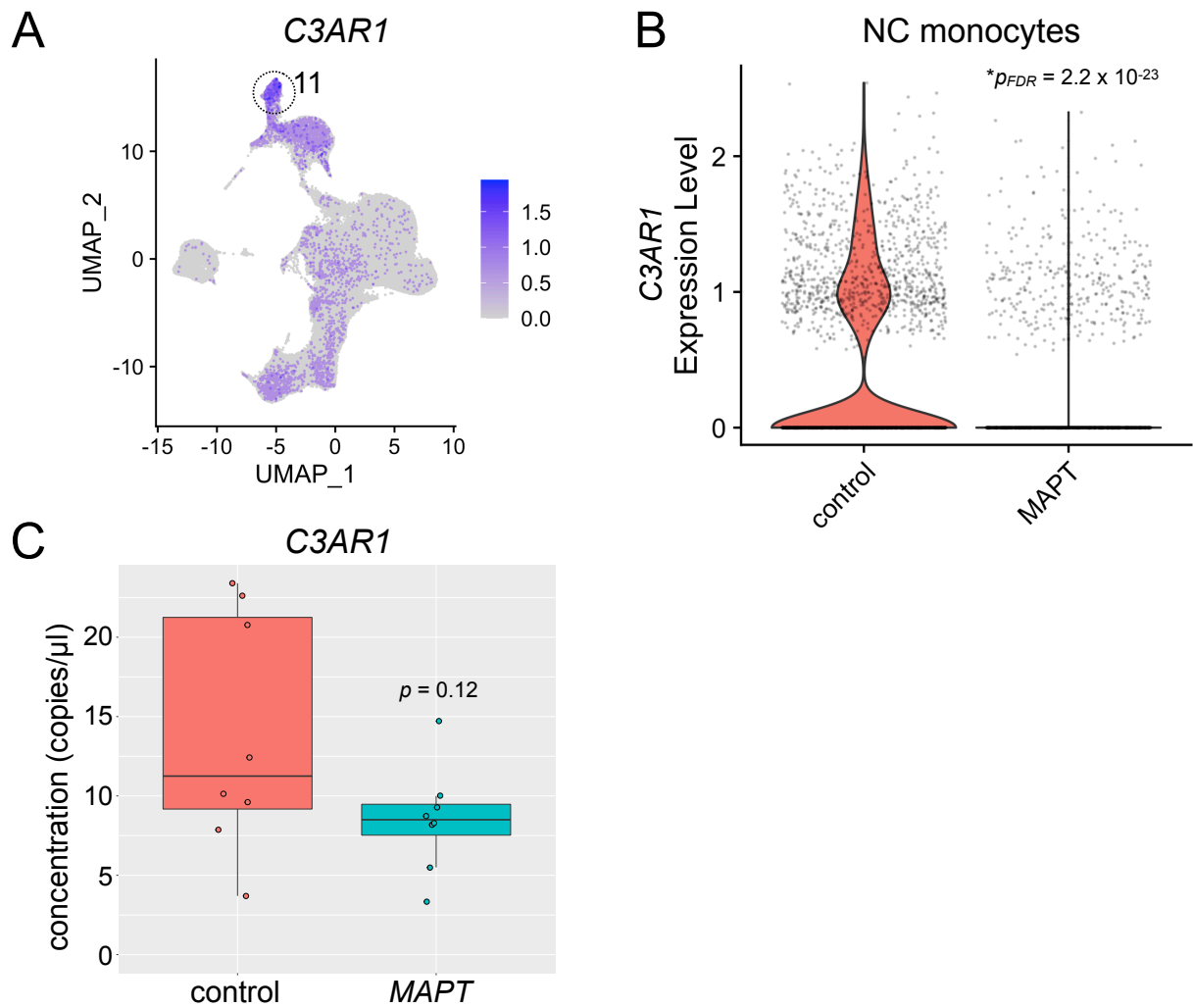


Figure 3.1.S6 Potential dysregulation of *C3AR1* in *MAPT* pathogenic variant carriers
 A) *C3AR1* expression was enriched in the NC monocyte cluster (11). B) *C3AR1* expression in NC monocytes was significantly reduced in *MAPT* variant carriers ($*p_{FDR} = 2.2 \times 10^{-23}$). C) ddPCR analysis of PBMC RNA revealed a trend toward reduced expression of *C3AR1* in *MAPT* variant carriers which did not reach significance ($p = 0.12$).

Table 3.1.3 Demographic characteristics of cohort.

	Control	<i>MAPT</i> variant carrier
<i>n</i>	8	8
Age, mean (SD)	52.8 (10.0)	54.4 (11.4)
Sex, <i>n</i> female	4	4
Clinical syndrome (<i>n</i>)	clinically normal (8)	bvFTD (4), frontal AD (1), subjective cognitive impairment (1), presymptomatic (2)
<i>MAPT</i> variants (<i>n</i>)	N/A	p.P301L (1), p.S305I (1), p.S305S (1), p.R406W (3), IVS10+16 (2)
Splicing variants, <i>n</i>	N/A	4
CDR-SB, mean (SD)	0.0 (0.0)	5.8 (5.7) 7.8 (5.3) [#] [#] symptomatic carriers only

3.1.8 Supplemental Online Content, Data and Code Availability

All supplemental online content can be downloaded directly from the published material at Genome Medicine.

3.1.9 References

1. Hutton M, Lendon CL, Rizzu P, Baker M, Froelich S, Houlden H, et al. Association of missense and 5'-splice-site mutations in tau with the inherited dementia FTDP-17. *Nature*. 1998;393:702–5.
2. Spillantini MG, Murrell JR, Goedert M, Farlow MR, Klug A, Ghetti B. Mutation in the tau gene in familial multiple system tauopathy with presenile dementia. *Proc Natl Acad Sci U S A*. 1998;95:7737–41.
3. Sirkis DW, Geier EG, Bonham LW, Karch CM, Yokoyama JS. Recent advances in the genetics of frontotemporal dementia. *Curr Genet Med Rep*. 2019;7:41–52.
4. VandeVrede L, Boxer AL, Polydoro M. Targeting tau: Clinical trials and novel therapeutic approaches. *Neurosci Lett*. 2020;731:134919.
5. Chang C-W, Shao E, Mucke L. Tau: Enabler of diverse brain disorders and target of rapidly evolving therapeutic strategies. *Science* [Internet]. 2021;371. Available from: <http://dx.doi.org/10.1126/science.abb8255>
6. Rosen HJ, Boeve BF, Boxer AL. Tracking disease progression in familial and sporadic frontotemporal lobar degeneration: Recent findings from ARTFL and LEFFTDS. *Alzheimers Dement*. 2020;16:71–8.
7. Song WM, Colonna M. The identity and function of microglia in neurodegeneration. *Nat Immunol*. 2018;19:1048–58.
8. Haage V, De Jager PL. Neuroimmune contributions to Alzheimer's disease: a focus on human data. *Mol Psychiatry* [Internet]. 2022; Available from: <http://dx.doi.org/10.1038/s41380-022-01637-0>

9. Chen Y, Colonna M. Spontaneous and induced adaptive immune responses in Alzheimer's disease: new insights into old observations. *Curr Opin Immunol.* 2022;77:102233.
10. Sirkis DW, Bonham LW, Karch CM, Yokoyama JS. Immunological signatures in frontotemporal lobar degeneration. *Curr Opin Neurol.* 2019;32:272–8.
11. Gate D, Saligrama N, Leventhal O, Yang AC, Unger MS, Middeldorp J, et al. Clonally expanded CD8 T cells patrol the cerebrospinal fluid in Alzheimer's disease. *Nature.* 2020;577:399–404.
12. Gate D, Tapp E, Leventhal O, Shahid M, Nonninger TJ, Yang AC, et al. CD4 T cells contribute to neurodegeneration in Lewy body dementia. *Science.* 2021;374:868–74.
13. Phongpreecha T, Fernandez R, Mrdjen D, Culos A, Gajera CR, Wawro AM, et al. Single-cell peripheral immunoprofiling of Alzheimer's and Parkinson's diseases. *Sci Adv* [Internet]. 2020;6. Available from: <http://dx.doi.org/10.1126/sciadv.abd5575>
14. Grozdanov V, Bliederaeuser C, Ruf WP, Roth V, Fundel-Clemens K, Zondler L, et al. Inflammatory dysregulation of blood monocytes in Parkinson's disease patients. *Acta Neuropathol.* 2014;128:651–63.
15. Zondler L, Müller K, Khalaji S, Bliedehäuser C, Ruf WP, Grozdanov V, et al. Peripheral monocytes are functionally altered and invade the CNS in ALS patients. *Acta Neuropathol.* 2016;132:391–411.
16. McGill RB, Steyn FJ, Ngo ST, Thorpe KA, Heggie S, Ruitenber MJ, et al. Monocytes and neutrophils are associated with clinical features in amyotrophic lateral sclerosis. *Brain Commun.* 2020;2:fcaa013.

17. Hofer TP, Zawada AM, Frankenberger M, Skokann K, Satzler AA, Gesierich W, et al. Slan- defined subsets of CD16-positive monocytes: impact of granulomatous inflammation and M-CSF receptor mutation. *Blood*. 2015;126:2601–10.
18. Sirkis DW, Bonham LW, Yokoyama JS. The Role of Microglia in Inherited White-Matter Disorders and Connections to Frontotemporal Dementia. *Appl Clin Genet*. 2021;14:195–207.
19. Sirkis DW, Bonham LW, Johnson TP, La Joie R, Yokoyama JS. Dissecting the clinical heterogeneity of early-onset Alzheimer’s disease. *Mol Psychiatry*. 2022;27:2674–88.
20. Karikari TK, Ashton NJ, Brinkmalm G, Brum WS, Benedet AL, Montoliu-Gaya L, et al. Blood phospho-tau in Alzheimer disease: analysis, interpretation, and clinical utility. *Nat Rev Neurol*. 2022;18:400–18.
21. Swift IJ, Sogorb-Esteve A, Heller C, Synofzik M, Otto M, Graff C, et al. Fluid biomarkers in frontotemporal dementia: past, present and future. *J Neurol Neurosurg Psychiatry*. 2021;92:204–15.
22. Cousins KAQ, Shaw LM, Chen-Plotkin A, Wolk DA, Van Deerlin VM, Lee EB, et al. Distinguishing Frontotemporal Lobar Degeneration Tau From TDP-43 Using Plasma Biomarkers. *JAMA Neurol* [Internet]. 2022; Available from: <http://dx.doi.org/10.1001/jamaneurol.2022.3265>
23. Silvin A, Uderhardt S, Piot C, Da Mesquita S, Yang K, Geirsdottir L, et al. Dual ontogeny of disease-associated microglia and disease inflammatory macrophages in aging and neurodegeneration. *Immunity*. 2022;55:1448–65.e6.

24. Miller ZA, Mandelli ML, Rankin KP, Henry ML, Babiak MC, Frazier DT, et al. Handedness and language learning disability differentially distribute in progressive aphasia variants. *Brain*. 2013;136:3461–73.
25. Rascovsky K, Hodges JR, Knopman D, Mendez MF, Kramer JH, Neuhaus J, et al. Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. *Brain*. 2011;134:2456–77.
26. Ossenkoppele R, Pijnenburg YAL, Perry DC, Cohn-Sheehy BI, Scheltens NME, Vogel JW, et al. The behavioural/dysexecutive variant of Alzheimer’s disease: clinical, neuroimaging and pathological features. *Brain*. 2015;138:2732–49.
27. Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology*. 1993;43:2412–4.
28. Haque A, Engel J, Teichmann SA, Lönnberg T. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Med*. 2017;9:75.
29. Nguyen QH, Pervolarakis N, Nee K, Kessenbrock K. Experimental Considerations for Single-Cell RNA Sequencing Approaches. *Front Cell Dev Biol*. 2018;6:108.
30. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM 3rd, et al. Comprehensive Integration of Single-Cell Data. *Cell*. 2019;177:1888–902.e21.
31. Hao Y, Hao S, Andersen-Nissen E, Mauck WM 3rd, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. *Cell*. 2021;184:3573–87.e29.
32. McGinnis CS, Murrow LM, Gartner ZJ. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst*. 2019;8:329–37.e4.
33. Xi NM, Li JJ. Benchmarking Computational Doublet-Detection Methods for Single-Cell RNA Sequencing Data. *Cell Syst*. 2021;12:176–94.e6.

34. Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* 2019;20:296.
35. Ilicic T, Kim JK, Kolodziejczyk AA, Bagger FO, McCarthy DJ, Marioni JC, et al. Classification of low quality cells from single-cell RNA-seq data. *Genome Biol.* 2016;17:29.
36. Law CW, Chen Y, Shi W, Smyth GK. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* 2014;15:R29.
37. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43:e47.
38. Sonesson C, Robinson MD. Bias, robustness and scalability in single-cell differential expression analysis. *Nat Methods.* 2018;15:255–61.
39. Mou T, Deng W, Gu F, Pawitan Y, Vu TN. Reproducibility of Methods to Detect Differentially Expressed Genes from Single-Cell RNA Sequencing. *Front Genet.* 2019;10:1331.
40. Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, et al. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* 2021;49:D605–12.
41. van Dongen S, Abreu-Goodger C. Using MCL to extract clusters from networks. *Methods Mol Biol.* 2012;804:281–95.
42. Friedman BA, Srinivasan K, Ayalon G, Meilandt WJ, Lin H, Huntley MA, et al. Diverse Brain Myeloid Expression Profiles Reveal Distinct Microglial Activation States and

- Aspects of Alzheimer's Disease Not Evident in Mouse Models. *Cell Rep.* 2018;22:832–47.
43. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15:550.
44. Wickham H. *ggplot2: Elegant Graphics for Data Analysis.* Springer Science & Business Media; 2009.
45. Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, Gassmann M, et al. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol Biol. BioMed Central;* 2006;7:1–14.
46. Mancuso R, Van Den Daele J, Fattorelli N, Wolfs L, Balusu S, Burton O, et al. Stem-cell-derived human microglia transplanted in mouse brain to study human disease. *Nat Neurosci.* 2019;22:2111–6.
47. Cabeza-Cabrerizo M, Cardoso A, Minutti CM, Pereira da Costa M, Reis e Sousa C. Dendritic Cells Revisited. *Annu Rev Immunol.* 2021;39:131–66.
48. Evans HT, Benetatos J, van Rooijen M, Bodea L-G, Götz J. Decreased synthesis of ribosomal proteins in tauopathy revealed by non-canonical amino acid labelling. *EMBO J.* 2019;38:e101174.
49. Koren SA, Hamm MJ, Meier SE, Weiss BE, Nation GK, Chishti EA, et al. Tau drives translational selectivity by interacting with ribosomal proteins. *Acta Neuropathol.* 2019;137:571–83.
50. Evans HT, Taylor D, Kneynsberg A, Bodea L-G, Götz J. Altered ribosomal function and protein synthesis caused by tau. *Acta Neuropathol Commun.* 2021;9:110.

51. Tracy TE, Madero-Pérez J, Swaney DL, Chang TS, Moritz M, Konrad C, et al. Tau interactome maps synaptic and mitochondrial processes associated with neurodegeneration. *Cell*. 2022;185:712–28.e14.
52. Liu C, Song X, Nisbet R, Götz J. Co-immunoprecipitation with Tau Isoform-specific Antibodies Reveals Distinct Protein Interactions and Highlights a Putative Role for 2N Tau in Disease. *J Biol Chem*. 2016;291:8173–88.
53. Karin M, Liu Z g., Zandi E. AP-1 function and regulation. *Curr Opin Cell Biol*. Elsevier BV; 1997;9:240–6.
54. Leandro GS, Evangelista AF, Lobo RR, Xavier DJ, Moriguti JC, Sakamoto-Hojo ET. Changes in Expression Profiles Revealed by Transcriptomic Analysis in Peripheral Blood Mononuclear Cells of Alzheimer’s Disease Patients. *J Alzheimers Dis*. 2018;66:1483–95.
55. Jiang S, Maphis NM, Binder J, Chisholm D, Weston L, Duran W, et al. Proteopathic tau primes and activates interleukin-1 β via myeloid-cell-specific MyD88- and NLRP3-ASC-inflammasome pathway. *Cell Rep*. 2021;36:109720.
56. Wang C, Fan L, Khawaja RR, Liu B, Zhan L, Kodama L, et al. Microglial NF- κ B drives tau spreading and toxicity in a mouse model of tauopathy. *Nat Commun*. 2022;13:1969.
57. Li X, Long J, He T, Belshaw R, Scott J. Integrated genomic approaches identify major pathways and upstream regulators in late onset Alzheimer’s disease. *Sci Rep*. 2015;5:12393.
58. Bhaskar K, Konerth M, Kokiko-Cochran ON, Cardona A, Ransohoff RM, Lamb BT. Regulation of tau pathology by the microglial fractalkine receptor. *Neuron*. 2010;68:19–31.

59. Cho S-H, Sun B, Zhou Y, Kauppinen TM, Halabisky B, Wes P, et al. CX3CR1 protein signaling modulates microglial activation and protects against plaque-independent cognitive deficits in a mouse model of Alzheimer disease. *J Biol Chem*. 2011;286:32713–22.
60. Maphis N, Xu G, Kokiko-Cochran ON, Jiang S, Cardona A, Ransohoff RM, et al. Reactive microglia drive tau pathology and contribute to the spreading of pathological tau in the brain. *Brain*. 2015;138:1738–55.
61. Whaley K. Biosynthesis of the complement components and the regulatory proteins of the alternative complement pathway by human peripheral blood monocytes. *J Exp Med*. 1980;151:501–16.
62. Martin U, Bock D, Arseniev L, Tornetta MA, Ames RS, Bautsch W, et al. The human C3a receptor is expressed on neutrophils and monocytes, but not on B or T lymphocytes. *J Exp Med*. 1997;186:199–207.
63. Wang S, Song R, Wang Z, Jing Z, Wang S, Ma J. S100A8/A9 in Inflammation. *Front Immunol*. 2018;9:1298.
64. Austermann J, Spiekermann C, Roth J. S100 proteins in rheumatic diseases. *Nat Rev Rheumatol*. 2018;14:528–41.
65. Hasselmann J, Coburn MA, England W, Figueroa Velez DX, Kiani Shabestari S, Tu CH, et al. Development of a Chimeric Model to Study and Manipulate Human Microglia In Vivo. *Neuron*. 2019;103:1016–33.e10.
66. van Leent MMT, Beldman TJ, Toner YC, Lameijer MA, Rother N, Bekkering S, et al. Prosaposin mediates inflammation in atherosclerosis. *Sci Transl Med [Internet]*. 2021;13. Available from: <http://dx.doi.org/10.1126/scitranslmed.abe1433>

67. Naj AC, Jun G, Beecham GW, Wang L-S, Vardarajan BN, Buross J, et al. Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat Genet.* 2011;43:436–41.
68. Litvinchuk A, Wan Y-W, Swartzlander DB, Chen F, Cole A, Propson NE, et al. Complement C3aR Inactivation Attenuates Tau Pathology and Reverses an Immune Network Deregulated in Tauopathy Models and Alzheimer's Disease. *Neuron.* 2018;100:1337–53.e5.
69. Wu T, Dejanovic B, Gandham VD, Gogineni A, Edmonds R, Schauer S, et al. Complement C3 Is Activated in Human AD Brain and Is Required for Neurodegeneration in Mouse Models of Amyloidosis and Tauopathy. *Cell Rep.* 2019;28:2111–23.e6.
70. Segovia M, Louvet C, Charnet P, Savina A, Tilly G, Gautreau L, et al. Autologous dendritic cells prolong allograft survival through Tmem176b-dependent antigen cross-presentation. *Am J Transplant.* 2014;14:1021–31.
71. Segovia M, Russo S, Jeldres M, Mahmoud YD, Perez V, Duhalde M, et al. Targeting TMEM176B Enhances Antitumor Immunity and Augments the Efficacy of Immune Checkpoint Blockers by Unleashing Inflammasome Activation. *Cancer Cell.* 2019;35:767–81.e6.
72. Piehl N, van Olst L, Ramakrishnan A, Teregulova V, Simonton B, Zhang Z, et al. Cerebrospinal fluid immune dysregulation during healthy brain aging and cognitive impairment. *Cell.* 2022;185:5028–39.e13.
73. Thrupp N, Sala Frigerio C, Wolfs L, Skene NG, Fattorelli N, Poovathingal S, et al. Single-Nucleus RNA-Seq Is Not Suitable for Detection of Microglial Activation Genes in Humans. *Cell Rep.* 2020;32:108189.

74. Landsman L, Bar-On L, Zerneck A, Kim K-W, Krauthgamer R, Shagdarsuren E, et al. CX3CR1 is required for monocyte homeostasis and atherogenesis by promoting cell survival. *Blood*. 2009;113:963–72.
75. Auffray C, Fogg DK, Narni-Mancinelli E, Senechal B, Trouillet C, Saederup N, et al. CX3CR1⁺ CD115⁺ CD135⁺ common macrophage/DC precursors and the role of CX3CR1 in their response to inflammation. *J Exp Med*. 2009;206:595–606.
76. Narasimhan PB, Marcovecchio P, Hamers AAJ, Hedrick CC. Nonclassical Monocytes in Health and Disease. *Annu Rev Immunol*. 2019;37:439–56.
77. Makinde HM, Cuda CM, Just TB, Perlman HR, Schwulst SJ. Non-classical monocytes mediate secondary injury, neurocognitive outcome, and neutrophil infiltration after Traumatic Brain Injury. *J Immunol*. NIH Public Access; 2017;199:3583.
78. Lee S, Xu G, Jay TR, Bhatta S, Kim K-W, Jung S, et al. Opposing effects of membrane-anchored CX3CL1 on amyloid and tau pathologies via the p38 MAPK pathway. *J Neurosci*. 2014;34:12538–46.
79. Maphis N, Xu G, Kokiko-Cochran ON, Cardona AE, Ransohoff RM, Lamb BT, et al. Loss of tau rescues inflammation-mediated neurodegeneration. *Front Neurosci*. 2015;9:196.
80. Merino JJ, Muñetón-Gómez V, Álvarez M-I, Toledano-Díaz A. Effects of CX3CR1 and Fractalkine Chemokines in Amyloid Beta Clearance and p-Tau Accumulation in Alzheimer's Disease (AD) Rodent Models: Is Fractalkine a Systemic Biomarker for AD? *Curr Alzheimer Res*. 2016;13:403–12.

81. Bolós M, Llorens-Martín M, Perea JR, Jurado-Arjona J, Rábano A, Hernández F, et al. Absence of CX3CR1 impairs the internalization of Tau by microglia. *Mol Neurodegener.* 2017;12:59.
82. Guedes JR, Lao T, Cardoso AL, El Khoury J. Roles of Microglial and Monocyte Chemokines and Their Receptors in Regulating Alzheimer's Disease-Associated Amyloid- β and Tau Pathologies. *Front Neurol.* 2018;9:549.
83. Fan Q, He W, Gayen M, Benoit MR, Luo X, Hu X, et al. Activated CX3CL1/Smad2 Signals Prevent Neuronal Loss and Alzheimer's Tau Pathology-Mediated Cognitive Dysfunction. *J Neurosci.* 2020;40:1133–44.
84. Puntambekar SS, Moutinho M, Lin PB-C, Jadhav V, Tumbleson-Brink D, Balaji A, et al. CX3CR1 deficiency aggravates amyloid driven neuronal pathology and cognitive decline in Alzheimer's disease. *Mol Neurodegener.* 2022;17:47.
85. Murai N, Mitalipova M, Jaenisch R. Functional analysis of CX3CR1 in human induced pluripotent stem (iPS) cell-derived microglia-like cells. *Eur J Neurosci.* 2020;52:3667–78.
86. Deczkowska A, Keren-Shaul H, Weiner A, Colonna M, Schwartz M, Amit I. Disease-Associated Microglia: A Universal Immune Sensor of Neurodegeneration. *Cell.* 2018;173:1073–81.
87. Ishizawa K, Dickson DW. Microglial activation parallels system degeneration in progressive supranuclear palsy and corticobasal degeneration. *J Neuropathol Exp Neurol.* 2001;60:647–57.

88. Yoshiyama Y, Higuchi M, Zhang B, Huang S-M, Iwata N, Saido TC, et al. Synapse loss and microglial activation precede tangles in a P301S tauopathy mouse model. *Neuron*. 2007;53:337–51.
89. Bellucci A, Bugiani O, Ghetti B, Spillantini MG. Presence of reactive microglia and neuroinflammatory mediators in a case of frontotemporal dementia with P301S mutation. *Neurodegener Dis. S. Karger AG*; 2011;8:221–9.
90. Asai H, Ikezu S, Woodbury ME, Yonemoto GMS, Cui L, Ikezu T. Accelerated neurodegeneration and neuroinflammation in transgenic mice expressing P301L tau mutant and tau-tubulin kinase 1. *Am J Pathol*. 2014;184:808–18.
91. Schröder JB, Pawlowski M, Meyer Zu Hörste G, Gross CC, Wiendl H, Meuth SG, et al. Immune Cell Activation in the Cerebrospinal Fluid of Patients With Parkinson’s Disease. *Front Neurol*. 2018;9:1081.
92. Balendra R, Isaacs AM. C9orf72-mediated ALS and FTD: multiple pathways to disease. *Nat Rev Neurol*. 2018;14:544–58.
93. Chang D, Nalls MA, Hallgrímsdóttir IB, Hunkapiller J, van der Brug M, Cai F, et al. A meta-analysis of genome-wide association studies identifies 17 new Parkinson’s disease risk loci. *Nat Genet*. 2017;49:1511–6.
94. Blauwendraat C, Nalls MA, Singleton AB. The genetic architecture of Parkinson’s disease. *Lancet Neurol*. 2020;19:170–8.
95. Villani A-C, Satija R, Reynolds G, Sarkizova S, Shekhar K, Fletcher J, et al. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science [Internet]*. 2017;356. Available from: <http://dx.doi.org/10.1126/science.aah4573>

96. Calzetti F, Tamassia N, Micheletti A, Finotti G, Bianchetto-Aguilera F, Cassatella MA. Human dendritic cell subset 4 (DC4) correlates to a subset of CD14CD16 monocytes. *J. Allergy Clin. Immunol.* 2018. p. 2276–9.e3.
97. Collin M, Bigley V. Human dendritic cell subsets: an update. *Immunology.* 2018;154:3–20.
98. Korenfeld D, Roussak K, Dinkel S, Vogel TP, Pollack H, Levy J, et al. STAT3 Gain-of-Function Mutations Underlie Deficiency in Human Nonclassical CD16 Monocytes and CD141 Dendritic Cells. *J Immunol.* 2021;207:2423–32.
99. Zhang Y, Fung ITH, Sankar P, Chen X, Robison LS, Ye L, et al. Depletion of NK Cells Improves Cognitive Function in the Alzheimer Disease Mouse Model. *The Journal of Immunology.* American Association of Immunologists; 2020;205:502–10.
100. Garofalo S, Coccozza G, Porzia A, Inghilleri M, Raspa M, Scavizzi F, et al. Natural killer cells modulate motor neuron-immune cell cross talk in models of Amyotrophic Lateral Sclerosis. *Nat Commun.* Nature Publishing Group; 2020;11:1–16.
101. Kaur G, Trowsdale J, Fugger L. Natural killer cells and their receptors in multiple sclerosis. *Brain.* 2013;136:2657–76.
102. Hertwig L, Hamann I, Romero-Suarez S, Millward JM, Pietrek R, Chanvillard C, et al. CX3CR1-dependent recruitment of mature NK cells into the central nervous system contributes to control autoimmune neuroinflammation. *Eur J Immunol.* 2016;46:1984–96.
103. Leuzy A, Chiotis K, Lemoine L, Gillberg P-G, Almkvist O, Rodriguez-Vieitez E, et al. Tau PET imaging in neurodegenerative tauopathies—still a challenge. *Mol Psychiatry.* Nature Publishing Group; 2019;24:1112–34.

104. Peet BT, Spina S, Mundada N, La Joie R. Neuroimaging in Frontotemporal Dementia: Heterogeneity and Relationships with Underlying Neuropathology. *Neurotherapeutics*. 2021;18:728–52.
105. Zetterberg H, Bendlin BB. Biomarkers for Alzheimer’s disease-preparing for a new era of disease-modifying therapies. *Mol Psychiatry*. 2021;26:296–308.
106. Gendron TF, Heckman MG, White LJ, Veire AM, Pedraza O, Burch AR, et al. Comprehensive cross-sectional and longitudinal analyses of plasma neurofilament light across FTD spectrum disorders. *Cell Rep Med*. 2022;3:100607.
107. Rexach JE, Polioudakis D, Yin A, Swarup V, Chang TS, Nguyen T, et al. Tau Pathology Drives Dementia Risk-Associated Gene Networks toward Chronic Inflammatory States and Immunosuppression. *Cell Rep*. 2020;33:108398.

Chapter 3 Paper 2

Expansion of highly interferon-responsive T cells in early-onset Alzheimer's disease

Contributing Authors: Daniel W. Sirkis*, Caroline Warly Solsberg*, Taylor P. Johnson, Luke W. Bonham, Alexis P. Oddi, Ethan G. Geier, Bruce L. Miller, Gil D. Rabinovici, and Jennifer S. Yokoyama**

*Equal contribution

**Correspondence

3.2.1 Abstract

3.2.1.1 Introduction

Altered immune signatures are emerging as a central theme in neurodegenerative disease, yet little is known about immune responses in early-onset Alzheimer's disease (EOAD).

3.2.1.2. Methods

We examined single-cell RNA-sequencing (scRNA-seq) data from peripheral blood mononuclear cells (PBMCs) and droplet digital (dd)PCR data from CD4 T cells from participants with EOAD and clinically normal controls.

3.2.1.3 Results

We analyzed ~182,000 PBMCs by scRNA-seq and discovered increased interferon signaling-associated gene (ISAG) expression and striking expansion of antiviral-like ISAG^{hi} T cells in EOAD. We isolated CD4 T cells from additional EOAD cases and confirmed increased expression of ISAG^{hi} marker genes. Publicly available cerebrospinal fluid leukocyte scRNA-seq data from late-onset mild cognitive impairment and AD also revealed increased expression of interferon-response genes.

3.2.1.4 Discussion

ISAG^{hi} T cells, apparently primed for antiviral activity, are expanded in EOAD. Additional research into these cells and the role of heightened peripheral IFN signaling in neurodegeneration is warranted.

3.2.2 Background

Approximately 5–10% of the ~7 million Americans living with Alzheimer's disease (AD) [1] experience symptom onset prior to age 65 [2]. In this early-onset form of AD (EOAD), affected individuals are more likely to experience an aggressive clinical course, have an atypical clinical

syndrome, encounter delays in diagnosis, and experience unique social disruptions due to their relatively young age [2]. The vast majority ($\geq 90\%$) of EOAD cases are not inherited in an autosomal-dominant manner, and for these individuals, we understand relatively little about the genetic and other biological factors underpinning disease risk.

Recent reports using single-cell RNA-sequencing (scRNA-seq) have highlighted changes in peripheral blood and cerebrospinal fluid (CSF) leukocyte populations in AD [3], Lewy body dementia [4], familial tauopathy [5], and during aging [6]. To our knowledge, however, a global, unbiased scRNA-seq analysis of peripheral blood mononuclear cells (PBMCs) in EOAD has not been reported. Using scRNA-seq, we now find evidence for marked expansion of a small population of recently characterized CD4 T cells expressing very high levels of interferon (IFN) signaling-associated genes (ISAG^{hi} T cells) in EOAD. Remarkably, a CD4 T-cell subtype that appears to be highly similar to ISAG^{hi} T cells—with a similar antiviral gene expression signature—is potently expanded in the CSF in the context of viral encephalitis [7], suggesting that EOAD-expanded ISAG^{hi} T cells have antiviral properties. Adding to the weight of evidence for augmented peripheral IFN signaling in EOAD, we also observe expansion of a rare natural killer (NK) cell population previously associated with heightened IFN signaling [8].

Beyond changes in cell-type abundance, we report global upregulation of IFN signaling genes across additional lymphoid and myeloid PBMC types in EOAD. In addition, by analyzing a publicly available scRNA-seq dataset of CSF leukocytes derived primarily from individuals with mild cognitive impairment (MCI) and late-onset AD (LOAD) [3], we find striking upregulation of the same IFN signaling pathways in CD4 T cells in late-onset disease. These findings suggest at least partially conserved IFN responses between EOAD and LOAD. Finally, we find consistent upregulation in the hippocampus of a suite of IFN response genes in a mouse model of familial

EOAD. Collectively, our findings indicate that dysregulation of IFN-related genes extends from the peripheral blood and CSF to the brain in AD and suggest a novel role for a population of unusual, antiviral-like T cells in EOAD.

3.2.3 Methods

3.2.3.1 Overview

After obtaining informed consent, PBMCs from study participants (Table 1) at the University of California, San Francisco Memory and Aging Center (MAC) were analyzed by scRNA-seq essentially as described [5]. Raw sequencing reads were aligned to GRCh38-2020-A and feature-barcode matrices generated using Cell Ranger (v7.1.0) with intronic reads excluded. Cluster proportions were determined for individual samples by dividing the number of cells in a given cluster by the total number of cells in clusters representing all PBMCs, all T cells, or all CD4 T cells (after quality control) for each individual. Statistical differences in cell-type abundances by diagnosis were assessed via linear modeling, controlling for age and sex. Additional details, including bioinformatic and experimental methods, are described in the Supplementary Methods document.

3.2.3.2 Clinical assessment

Study participants underwent a multistep screening prior to an in-person clinical assessment at the MAC that included a neurologic exam, detailed cognitive assessment, medical history, and family history for neurodegenerative disease [9]. Study partners were interviewed about the participant's functional abilities. A multidisciplinary team consisting of a neurologist, a neuropsychologist, and a nurse reviewed all participant clinical information and established diagnoses for cases according to consensus criteria for AD [10–12] or frontal AD [13]. All EOAD cases were diagnosed with probable AD and had at least one positive biomarker consistent with AD. In particular, 10/13 cases

had evidence of amyloid and tau positivity (obtained via PET imaging or assessment of CSF amyloid- β_{42} and phospho-tau181 levels) in addition to neurodegeneration (via structural MRI; A+T+N+), while 3/13 cases had evidence of neurodegeneration (N+). The mean (SD) age of first abnormal diagnosis for all participants with EOAD was 58.5 (3.0) and the range was 54–62. All control participants had a normal neurologic exam, and all controls except one in the ddPCR study had a global Clinical Dementia Rating (CDR) [14] scale score of 0; the remaining control participant, who was diagnosed as clinically normal, had a CDR score of 0.5 due to subjective memory complaint. This participant also reported depressive symptoms. Sensitivity analysis of the ddPCR data after exclusion of this individual indicated that the results remained similar or unchanged. All participants screened negative for disease-causing pathogenic variants in established genes for AD and frontotemporal lobar degeneration, which also causes early-onset dementia.

3.2.4 RESULTS

3.2.4.1 Identification of an expanded T-cell subtype in EOAD

After QC filtering, clustering of ~182,000 PBMCs generated 19 primary clusters consisting of all expected PBMC types. Comparison of relative cluster abundance in EOAD cases vs. controls revealed a single cluster (cluster 15) that was robustly expanded in EOAD (**Figure 3.2.1A**). Expression of marker genes indicates that cluster 15 is a subtype of CD4 T cell (**Figure 3.2.S1A**). Quantification of cluster 15 abundance relative to either all PBMCs, all T cells, or all CD4 T cells revealed significant expansion in EOAD that was driven primarily by females (**Figure 3.2.1B and C**). To determine what type of CD4 T cell cluster 15 represents, we subsetted all T cells and reclustered them separately from all other cell types. Reclustering revealed this cell type in sub-cluster 11, which expresses uniquely high levels of IFN-signaling genes *MX1* and *IFI6* relative to all other T cells (**Figure 3.2.1D**). As expected, sub-cluster 11 was also significantly expanded in EOAD relative to controls (**Figure 3.2.1D**).

3.2.4.2 Characterization of the expanded cell type as ISAG^{hi} T cells

What is the precise identity of this subset of CD4 T cells? Recent literature using scRNA-seq to analyze human leukocyte populations has revealed two poorly understood cell types: ISAG^{hi} T cells, detected in peripheral blood [15], and antiviral CD4 T cells, detected in CSF [7]. Antiviral CD4 T cells were so named due to their marker gene expression and robust expansion in the CSF in the context of viral encephalitis [7]. Comparison of all marker genes for our sub-cluster 11 to the top 200 marker genes for CSF antiviral CD4 T cells revealed highly statistically significant overlap ($P = 6.5 \times 10^{-14}$; supplementary Table S1) [16]. Moreover, all of the 12 most-significant marker genes originally reported for ISAG^{hi} T cells [15] are also top marker genes of our sub-cluster 11 and of antiviral CD4 T cells. Therefore, from here on we refer to the EOAD-expanded CD4 T cells as ISAG^{hi} T cells.

3.2.4.3 Analysis of ISAG^{hi} T-cell abundance in additional samples and datasets

ISAG^{hi} T-cell abundance was consistent across scRNA-seq batches (**Figure 3.2.S1B**) and was not driven by *APOE* $\epsilon 4$ status (**Figure 3.2.S1C**). Moreover, although our control samples came from participants with a younger mean age (**Table 3.2.1**), there was no relationship between age and ISAG^{hi} T-cell abundance (**Figure 3.2.S1D**). To increase the sample size of our scRNA-seq dataset, we included seven additional control PBMC samples previously characterized by scRNA-seq [5]. We found that the expansion of ISAG^{hi} T cells relative to PBMCs and all T cells remained significant after addition of these independent controls, despite a single outlier control sample with very high levels of ISAG^{hi} T cells (**Figure 3.2.S2**). We recently reported a reduction in peripheral non-classical monocytes in familial tauopathy [5]. Comparing the familial tauopathy and EOAD datasets, we found that non-classical monocytes are not reduced in EOAD, and ISAG^{hi} T cells are

not expanded in familial tauopathy (**Figure 3.2.S3**), suggesting distinct peripheral immune responses in sporadic EOAD and primary familial tauopathy.

3.2.4.4 Expansion of proliferating natural killer cells in EOAD

Previous single-cell analyses have revealed additional PBMC types temporally associated with heightened type I IFN signaling. In particular, a rare NK cell subpopulation that expresses markers of proliferation has been shown to significantly expand after vaccination with an experimental HIV vaccine [8]. This expansion coincides with heightened type I IFN signaling in myeloid cells [8], which we also observe in EOAD (see below). After mapping the EOAD PBMC dataset onto a large, well-characterized multimodal PBMC CITE-seq dataset [8], we identified the proliferating NK cell cluster and, remarkably, observed significant expansion of this rare subpopulation in EOAD, specifically in female cases (**Figure 3.2.S4A**). In addition, differential expression analysis confirmed significant enrichment for gene ontology (GO) and reactome terms related to IFN signaling and antiviral response within the primary NK cell cluster in EOAD (**Figure 3.2.S4B**). These findings provide additional corroborative evidence, via a population of innate lymphoid cells, that EOAD is characterized by heightened peripheral IFN signaling.

3.2.4.5 Differential expression analysis of PBMC subsets in EOAD

Differential expression analysis revealed a high number of differentially expressed genes (DEGs) in classical and non-classical monocytes in EOAD, relative to controls (**Figure 3.2.S5A**, supplementary Table S2). Remarkably, we found that, on average, ~19% of the significantly upregulated genes across all clusters (excluding those with fewer than 10 upregulated DEGs) were also ISAG^{hi} T-cell marker genes (i.e., IFN response genes; **Figure 3.2.S5A**). GO analysis of the genes upregulated in CD4 T-cell clusters and myeloid cell clusters revealed significant enrichment for terms such as ‘IFN α/β signaling’ and ‘response to virus’ (**Figure 3.2.S5B and C**). In EOAD,

we therefore observe both expansion of a CD4 T-cell subpopulation expressing very high levels of genes associated with IFN signaling and concomitant upregulation of many of the same genes across additional lymphoid and myeloid cell types.

3.2.4.6 Validation of upregulated ISAG^{hi} T-cell marker genes via ddPCR

To validate our primary scRNA-seq findings, we magnetically isolated CD4 T cells from an additional cohort of EOAD cases and control participants. A droplet digital (dd)PCR-based validation assay indicated highly efficient isolation of CD4 T cells (**Figure 3.2.S6**). Reasoning that increased expression of specific ISAG^{hi} T-cell marker genes from isolated CD4 T cells would be consistent with expansion of ISAG^{hi} T cells as well as ISAG upregulation, we performed ddPCR for ISAG^{hi} marker genes *MX1* and *IFI6* (**Figure 3.2.2A**). Cases and controls in the ddPCR cohort had similar average ages (**Table 3.2.1**; see also supplementary Methods document), excluding age as an explanatory factor. Strikingly, ddPCR confirmed increased *MX1* and *IFI6* expression in CD4 T cells from EOAD cases (**Figure 3.2.2B**). Increased *MX1* was observed across two independent ddPCR batches and was driven by females (**Figure 3.2.2C**).

3.2.4.7 Secondary analysis of CSF leukocytes in late-onset MCI/AD

Secondary analysis of a well-known CSF leukocyte dataset [3] revealed that ISAG^{hi}-like T cells, although detected, were not expanded in the CSF in late-onset MCI/AD (**Figure 3.2.3A through C**). Strikingly, however, differential expression analysis revealed robust upregulation of *MX1* within CSF ISAG^{hi}-like T cells in MCI/AD (**Figure 3.2.3D**). Moreover, functional enrichment analysis of the genes upregulated in MCI/AD (relative to healthy controls) across all CSF CD4 T cells revealed highly significant enrichment for terms such as ‘IFN α/β signaling’ and ‘response to virus’ (**Figure 3.2.3E**). In addition, analysis of upregulated DEGs from individual CSF clusters revealed similar enrichment for IFN signaling terms across multiple CD4 T-cell clusters as well as

two innate immune clusters (**Figure 3.2.3F**). Collectively, these results suggest that, although *expansion* of ISAG^{hi} T cells may be specific to EOAD, upregulation of the same IFN signaling pathways in CSF CD4 T cells is conserved in late-onset MCI/AD.

3.2.4.8 Dysregulation of IFN signaling genes in a mouse model of AD

To assess the relevance of heightened type I IFN signaling in the brain, we asked whether ISAG^{hi} T-cell marker genes are dysregulated in the hippocampus and cortex in the *APP^{swe} x PSEN1.M146V* (TASTPM) mouse model [17] of familial EOAD. In the TASTPM model, we observed marked upregulation of many ISAG^{hi} marker genes, particularly in the hippocampus (**Figure 3.2.S7A and B**). These results confirm the relevance of dysregulated type I IFN signaling in the brain in a model of EOAD-like amyloidosis.

3.2.5 Discussion

In this study, we found evidence for a unique peripheral immune signature in EOAD. Our findings complement and expand upon existing evidence of diverse T-cell signatures in other forms of AD [3,18], additional neurodegenerative diseases [4], and during aging [6]. Our study is limited by the relatively small sample sizes that are characteristic of scRNA-seq experiments. Future studies in diverse EOAD cohorts from additional recruitment sites will be needed to confirm the broad relevance of our findings to EOAD. In light of recent findings suggesting that (i) herpes zoster vaccination may be causally associated with reduced dementia risk in women [19]; (ii) viral encephalitis exposure markedly increases risk for AD [20]; and (iii) the ISAG^{hi} T cells increased in EOAD bear striking resemblance to antiviral CD4 T cells expanded in CSF in viral encephalitis [7], our findings raise the intriguing possibility that AD is characterized by an IFN-driven, antiviral-like T-cell response in both peripheral blood and CSF.

Recent work in chronic graft-versus-host disease (cGVHD) has revealed a population of naive CD4 T cells expressing an IFN response signature including *IFI6*, *ISG15*, *IFI44L*, *STAT1*, and *MXI* [21], genes which are all top markers of the ISAG^{hi} T cells described here and elsewhere [15], and of antiviral CD4 T cells expanded in viral encephalitis [7]. This population is strikingly enriched on day 100 post-transplant in patients who ultimately developed cGVHD, relative to those who remained tolerant. This finding strongly suggests that ISAG^{hi}-like CD4 T cells contribute to tissue destruction in the context of cGVHD and may therefore also possess pathogenicity in other contexts, including AD. Future functional studies will be required to directly assess the pathogenic potential of ISAG^{hi} T cells in neurodegenerative disease models.

Complementing these insights, transcriptomic analyses have demonstrated heightened expression of genes linked to IFN signaling associated with the development of post-traumatic stress disorder (PTSD) [22, 23]. Intriguingly, upon examination of detailed clinical notes for our additional control scRNA-seq samples, we found that the outlier with markedly higher levels of ISAG^{hi} T cells (**Figure 3.2.S2**) had been previously diagnosed with PTSD. These findings suggest that the relevance of ISAG^{hi} T cells may extend beyond neurodegenerative disease to neuropsychiatric conditions as well.

In addition to our findings involving T cells, we also found evidence for heightened IFN signaling in myeloid cells and NK cells. Indeed, augmented type I IFN signaling in myeloid cells has previously been associated with expansion of proliferating NK cells [8]. Although these observations were originally reported in the context of vaccination, we observed both of these phenomena in EOAD. Given that type I IFN signaling has been shown to promote NK cell expansion and survival [24], our findings here not only support these prior observations but also

provide additional evidence for heightened peripheral IFN signaling in EOAD in PBMC types beyond T cells.

Our findings build upon prior research that has found increasing evidence for heightened T-cell infiltration into the brain in AD [3, 25] and AD models [25, 26]. In addition, recent work in AD, primary tauopathy, and related model systems has implicated dysregulated IFN signaling pathways in microglia [25, 27–32] and brain barrier tissue [33], indicating that IFN signaling is implicated not only in peripheral and CSF immune cells—as shown here—but also at the blood–CSF barrier and in brain-resident myeloid cells. Indeed, the heightened expression of ISAG^{hi} marker genes in the brain in the TASTPM model that we confirmed here may be mediated primarily by changes in microglial gene expression. Collectively, our novel findings, coupled with this prior body of work, suggest the importance of heightened IFN signaling in PBMCs, CSF immune cells, brain border tissue, and brain parenchyma, which may be mediated by distinct cellular populations in each compartment. Future work should focus on identifying the functional and compartment-specific roles of these IFN-responsive cells in neurodegenerative disease.

3.2.6 Article Information

3.2.6.1 Acknowledgments

The cartoon in **Figure 3.2.2** was created using BioRender.com.

3.2.6.2 Sources of Funding

J.S.Y. receives funding from NIH-NIA R01AG062588, R01AG057234, P30AG062422, P01AG019724, and U19AG079774; NIH-NINDS U54NS123985; NIH-NIDA 75N95022C00031; the Rainwater Charitable Foundation; the Bluefield Project to Cure Frontotemporal Dementia; the Alzheimer’s Association; the Global Brain Health Institute; the French Foundation; and the Mary Oakley Foundation. C.W.S. is supported in part by the NIH Intramural Center for Alzheimer’s and

Related Dementias (CARD), project NIH-NIA ZIAAG000534. G.D.R. receives research funding from Avid Radiopharmaceuticals, GE Healthcare, Life Molecular Imaging, and Genentech. The content of this publication is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

3.2.6.3 Disclosures

J.S.Y. serves on the scientific advisory board for the Epstein Family Alzheimer's Research Collaboration. G.D.R. has received consulting fees from Alector, Eli Lilly, Genentech, Roche, and Merck. He receives fees for serving on a DSMB for Johnson & Johnson. B.L.M. serves as medical advisor for the French Foundation; serves as scientific advisor for the Larry L. Hillblom Foundation, the Association for Frontotemporal Degeneration, the NIHR Cambridge Biomedical Research Centre and its subunit, the Biomedical Research Unit in Dementia, and the Buck Institute for Research on Aging; serves as external advisor for the University of Washington ADRC, Stanford University ADRC, Arizona Alzheimer's Disease Center, and Massachusetts ADRC; and serves on the external advisory committee for the University of Southern California P01 Urban Air Pollution and Alzheimer's Disease: Risk, Heterogeneity and Mechanisms.

3.2.6.4 Consent Statement

All participants or their surrogates provided written informed consent prior to study participation; all aspects of the studies described here were approved by the UCSF institutional review board.

3.2.7 Figures and Tables

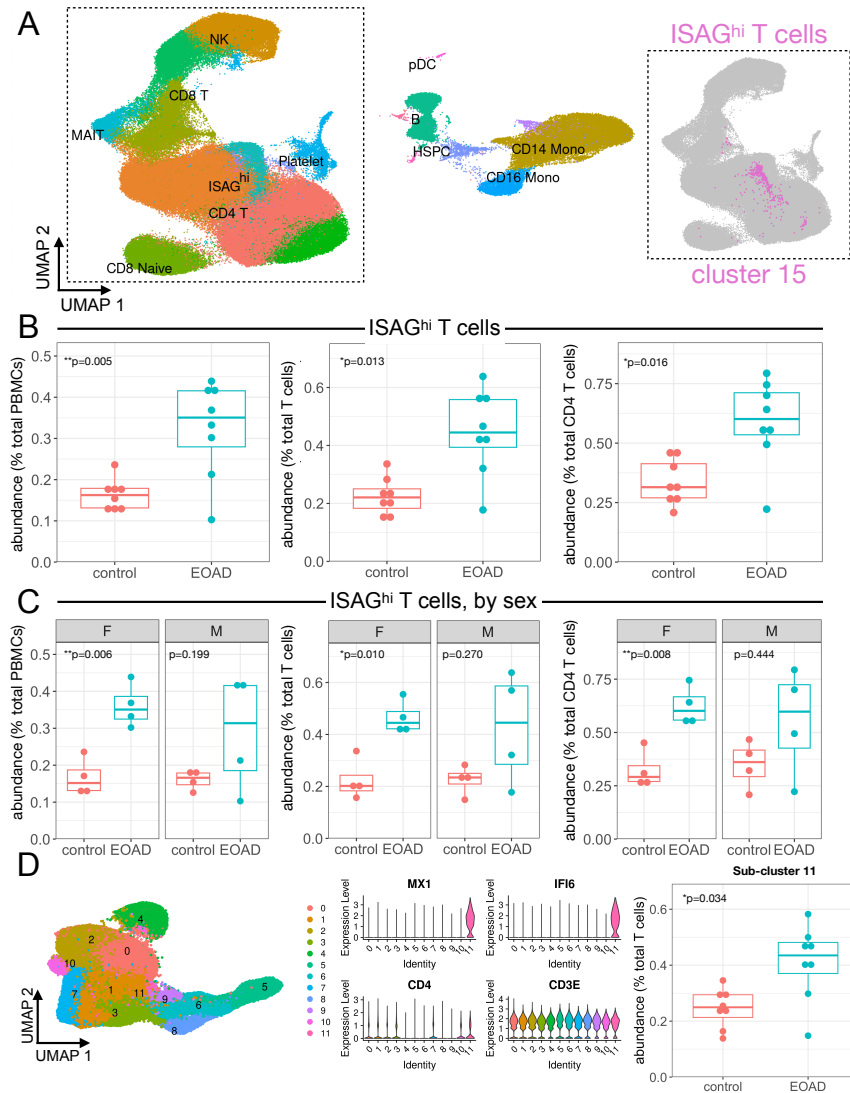


Figure 3.2.1 Expansion of ISAG^{hi} T cells in EOAD characterized by scRNA-seq
 A) Uniform manifold approximation and projection (UMAP) plot of ~182,000 PBMCs from EOAD cases and cognitively normal controls, colored by cluster identity. Major cell types are labeled within the plot. The inset (right) shows the primary T-cell grouping displayed in gray, with the ISAG^{hi} T-cell cluster shown in magenta. B) ISAG^{hi} T-cell abundance is quantified relative to all PBMCs (left; $P = 0.005$), all T cells (middle, $P = 0.013$), and all CD4 T cells (right; $P = 0.016$). C) Stratifying by sex, ISAG^{hi} T-cell relative abundance is significantly increased in EOAD only in females, expressed as a percentage of PBMCs (left, $P = 0.006$), T cells (middle, $P = 0.01$), and CD4 T cells (right, $P = 0.008$). D) Reclustering of all T cells (left) generates a T-cell subcluster (11) representing ISAG^{hi} T cells, which express high levels of marker genes *MX1* and *IFI6*, in addition to T-cell markers *CD4* and *CD3E* (middle). Quantification of the ISAG^{hi} subcluster relative to all T cells again indicates a significant increase in EOAD cases (right, $P = 0.034$).

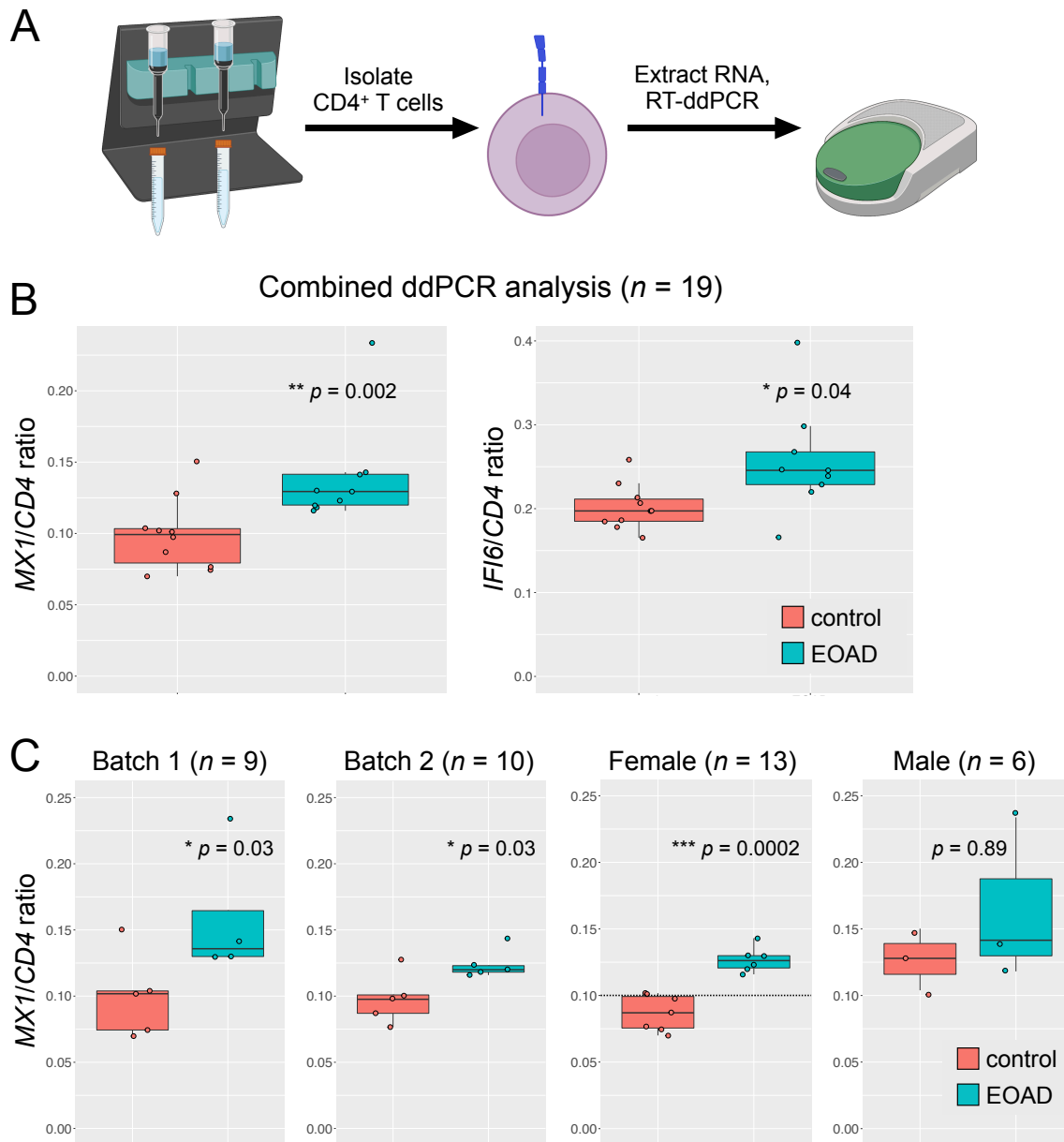


Figure 3.2.2 ISAG^{hi} T-cell marker gene expression is increased in CD4 T cells in EOAD
 A) CD4 T cells were magnetically isolated from PBMCs and RNA was extracted; gene expression was determined by RT-ddPCR. B) Expression of *MXI* and *IFI6* was significantly increased in CD4 T cells from EOAD cases relative to cognitively normal controls ($P = 0.002$ and $P = 0.04$, respectively). C) *MXI* expression was significantly increased in two independent RT-ddPCR batches ($P = 0.03$, both batches). The increase in *MXI* expression observed in EOAD was driven by females ($P = 0.0002$). *CD4* was used as a reference gene.

CSF leukocytes from Gate et al., 2020

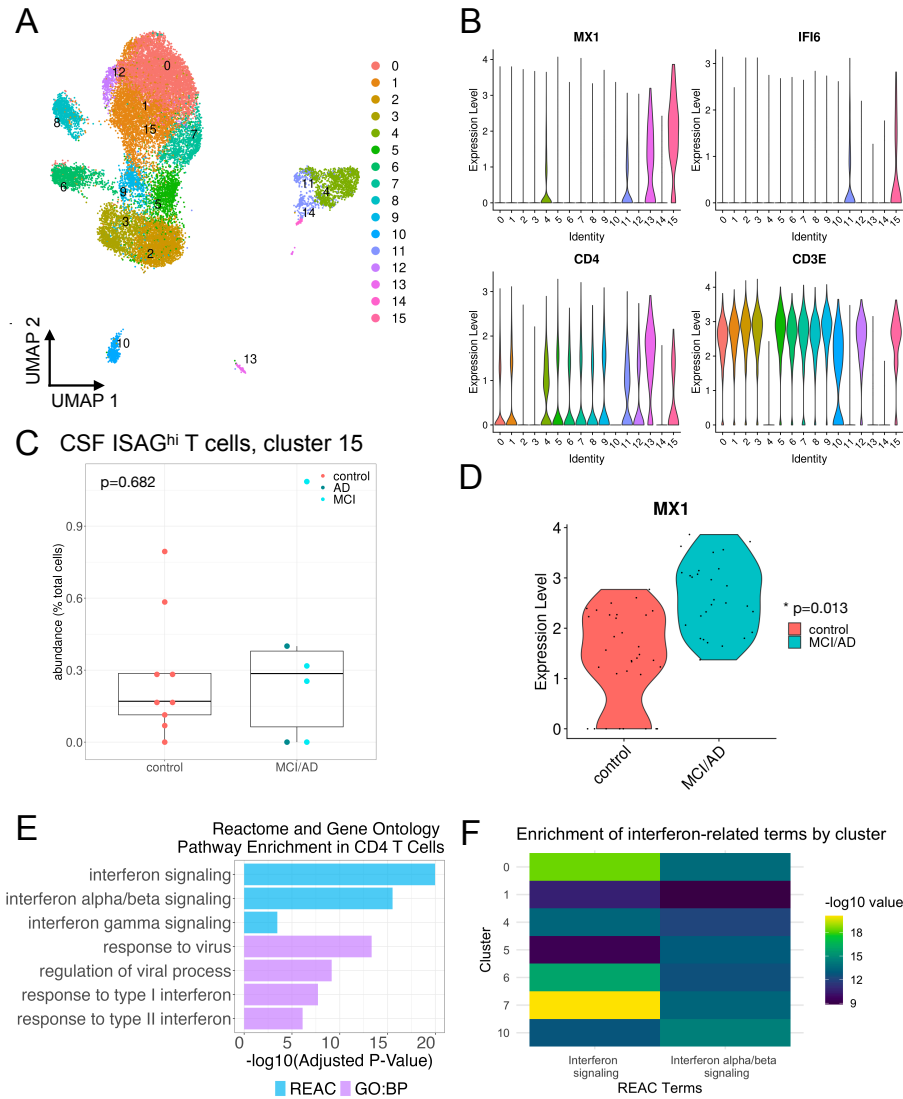


Figure 3.2.3 Heightened IFN response signatures in CSF T cells in MCI and AD

Publicly available data from Gate et al., 2020 [3] were downloaded from GEO (GSE134577) and analyzed as described. A UMAP plot (A) shows the distribution of CSF immune cells. B) Violin plots show that CSF cluster 15 harbors ISAG^{hi}-like cells expressing high levels of *MX1* and *IFI6* along with *CD4* and *CD3E*. C) Quantification of CSF cluster 15 reveals lack of expansion in late-onset MCI/AD ($P = 0.682$). D) *MX1* expression is significantly increased ($P_{FDR} = 0.014$) in CSF ISAG^{hi}-like T cells in MCI/AD relative to healthy controls. E) Functional enrichment analysis of the genes upregulated ($P_{FDR} < 0.05$ and \log_2 fold-change > 0.1) across all CSF CD4 T-cell clusters (0, 1, 5, 6, 7, 8, 9, 12, and 15) reveals significant enrichment of IFN and antiviral response pathways in MCI/AD. GO biological process (BP) and reactome databases were used. F) Analysis of significantly upregulated genes from individual CSF immune cell clusters revealed significant enrichment for IFN signaling in individual CSF CD4 T-cell clusters (0, 1, 5, 6, and 7) as well as monocyte and NK cell clusters (4 and 10, respectively).

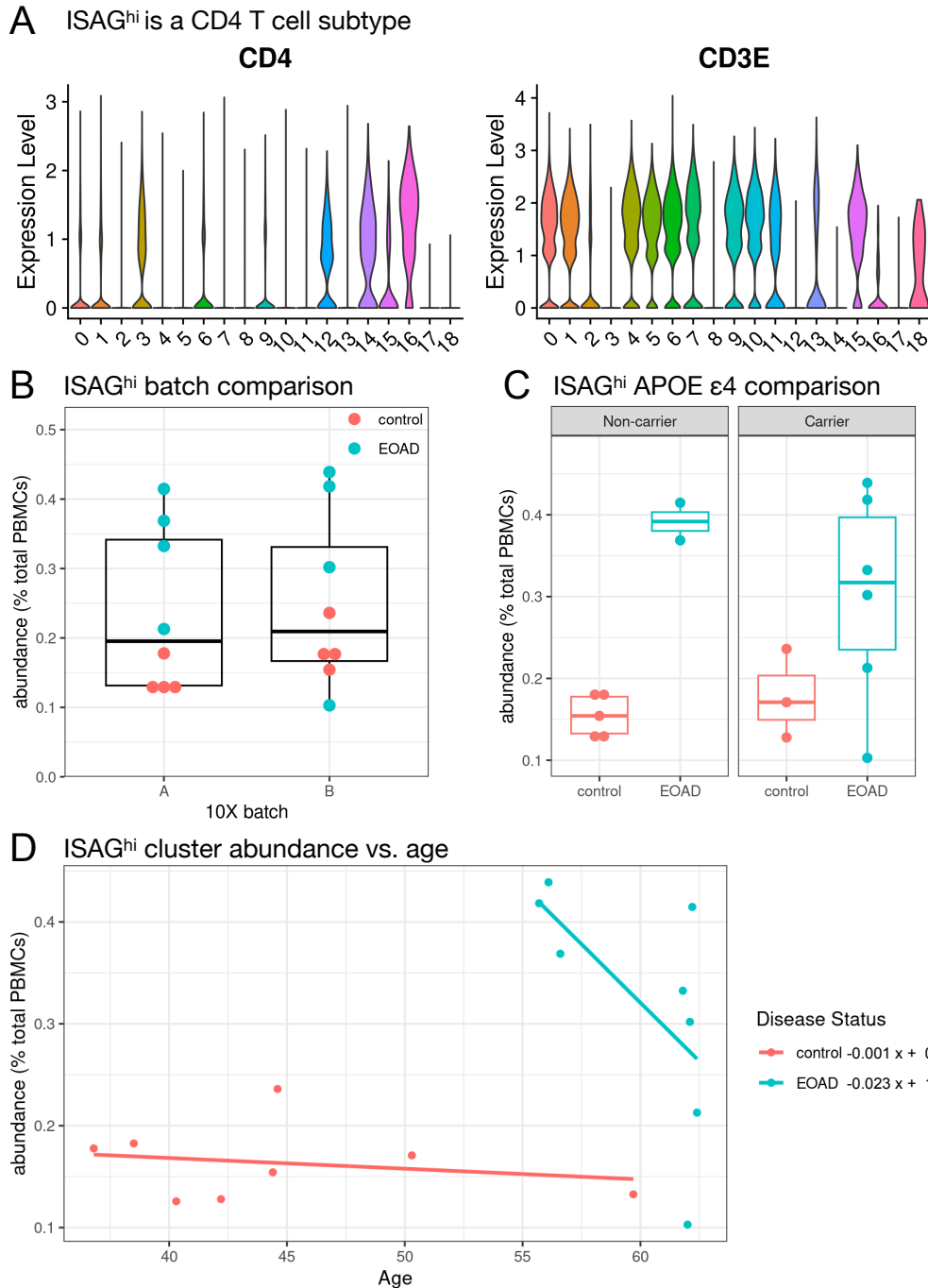


Figure 3.2.S1 Analysis of additional variables' relationships with ISAG^{hi} T cells

A) Cluster 15 displays robust expression of both *CD4* and *CD3E*, indicating that it contains CD4 T cells. B) Estimated relative abundance of ISAG^{hi} T cells (expressed as percentage of total PBMCs) was consistent across two scRNA-seq batches. C) *APOE* ε4 carrier status was not clearly associated with ISAG^{hi} T-cell abundance in EOAD cases or controls. D) Increased age was not associated with increased abundance of ISAG^{hi} T cells in EOAD cases or controls.

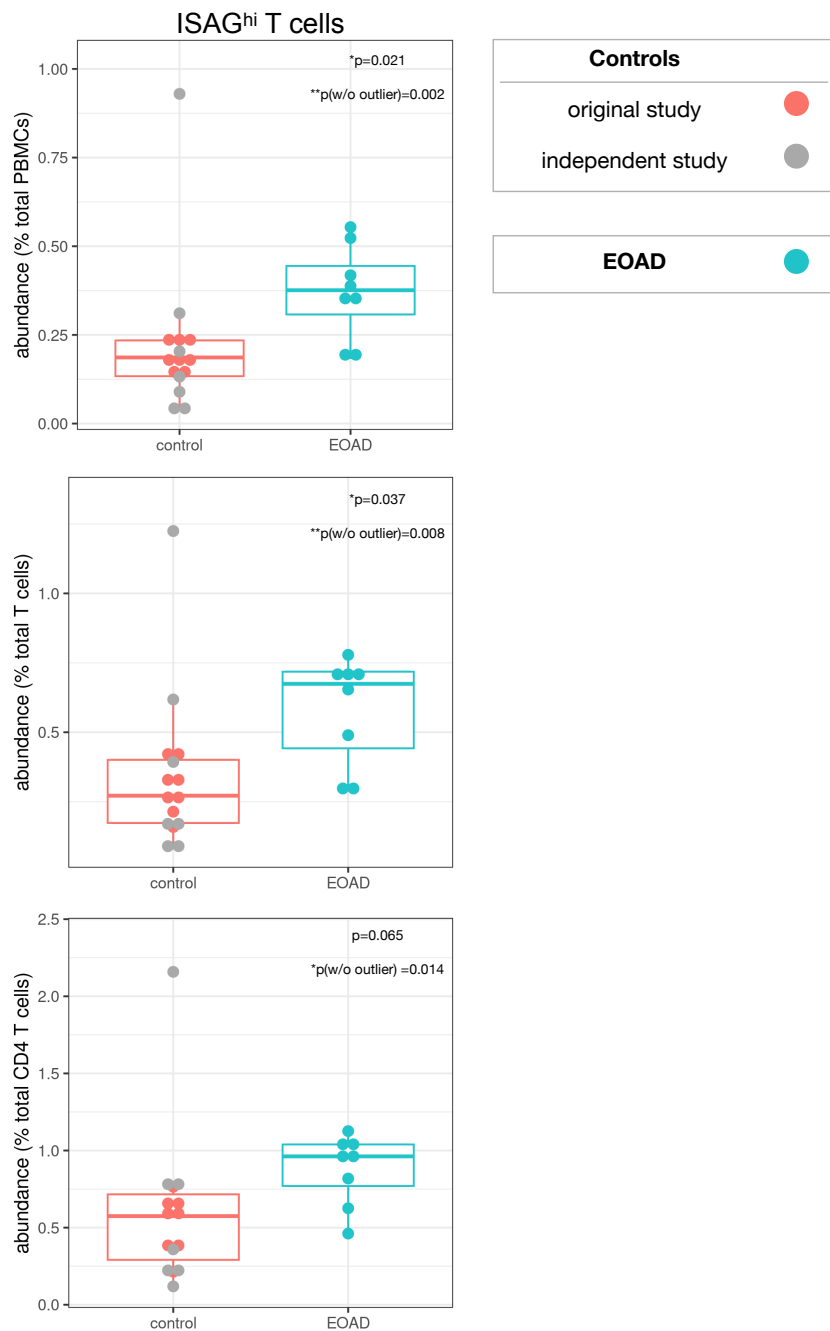


Figure 3.2.S2 Reanalysis after integration of additional cognitively normal control samples. Seven additional cognitively normal control samples from a prior study [1] were integrated together with the 16 samples from the present study. ISAG^{hi} abundance relative to all PBMCs (top), all T cells (middle), and all CD4 T cells (bottom) was quantified. The increased abundance of ISAG^{hi} T cells in EOAD remained significant relative to total PBMCs ($P = 0.021$) and total T cells ($P = 0.037$), while the difference relative to total CD4 T cells remained significant only after removal of the control outlier sample ($P = 0.014$).

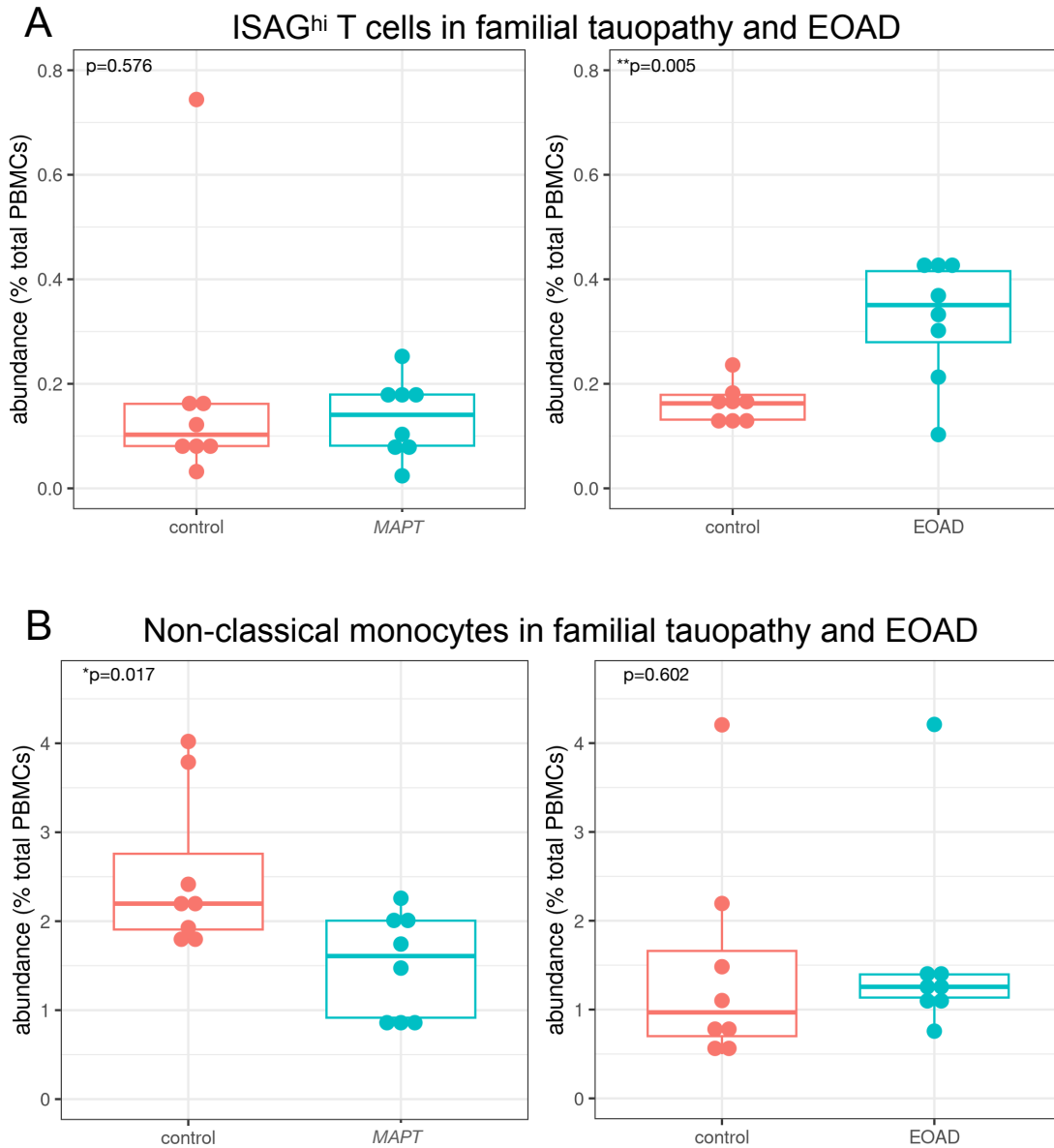


Figure 3.2.S3 Analysis of changes in cell-type abundance in familial tauopathy and EOAD
 A) Analysis of changes in ISAG^{hi} T-cell abundance in the present study and an independent study focusing on familial tauopathy revealed a specific expansion in EOAD ($P = 0.005$). B) Similar analysis of changes in non-classical monocyte abundance revealed that the reduction in non-classical monocytes is specific to familial tauopathy ($P = 0.017$).

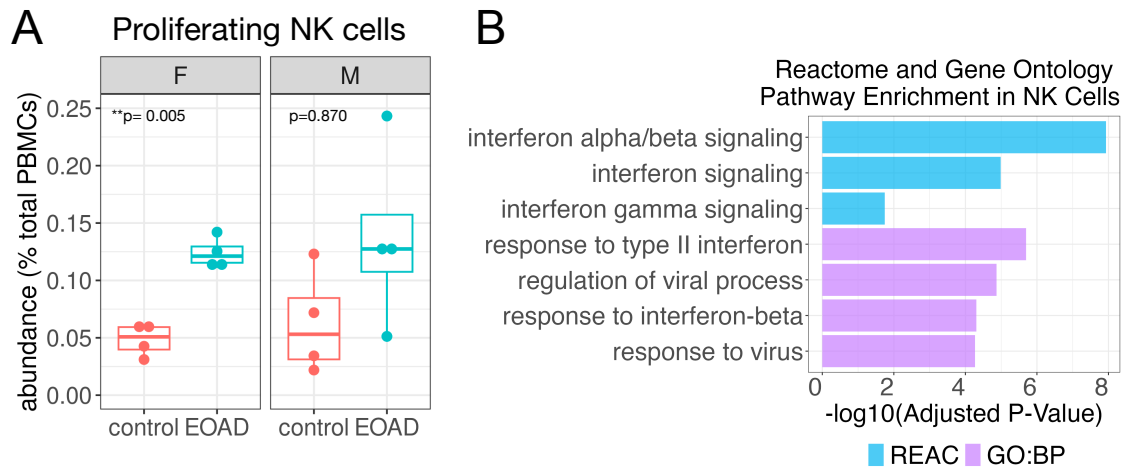


Figure 3.2.S4 Analysis of proliferating NK cells and NK cell IFN signaling in EOAD
 A) Quantitative analysis revealed a significant increase in proliferating NK cell abundance specifically in female EOAD cases compared to controls ($P = 0.005$). B) Significant upregulation of IFN signaling and antiviral response pathways in NK cells from EOAD patients is demonstrated by querying the GO BP and reactome databases with significantly upregulated genes.

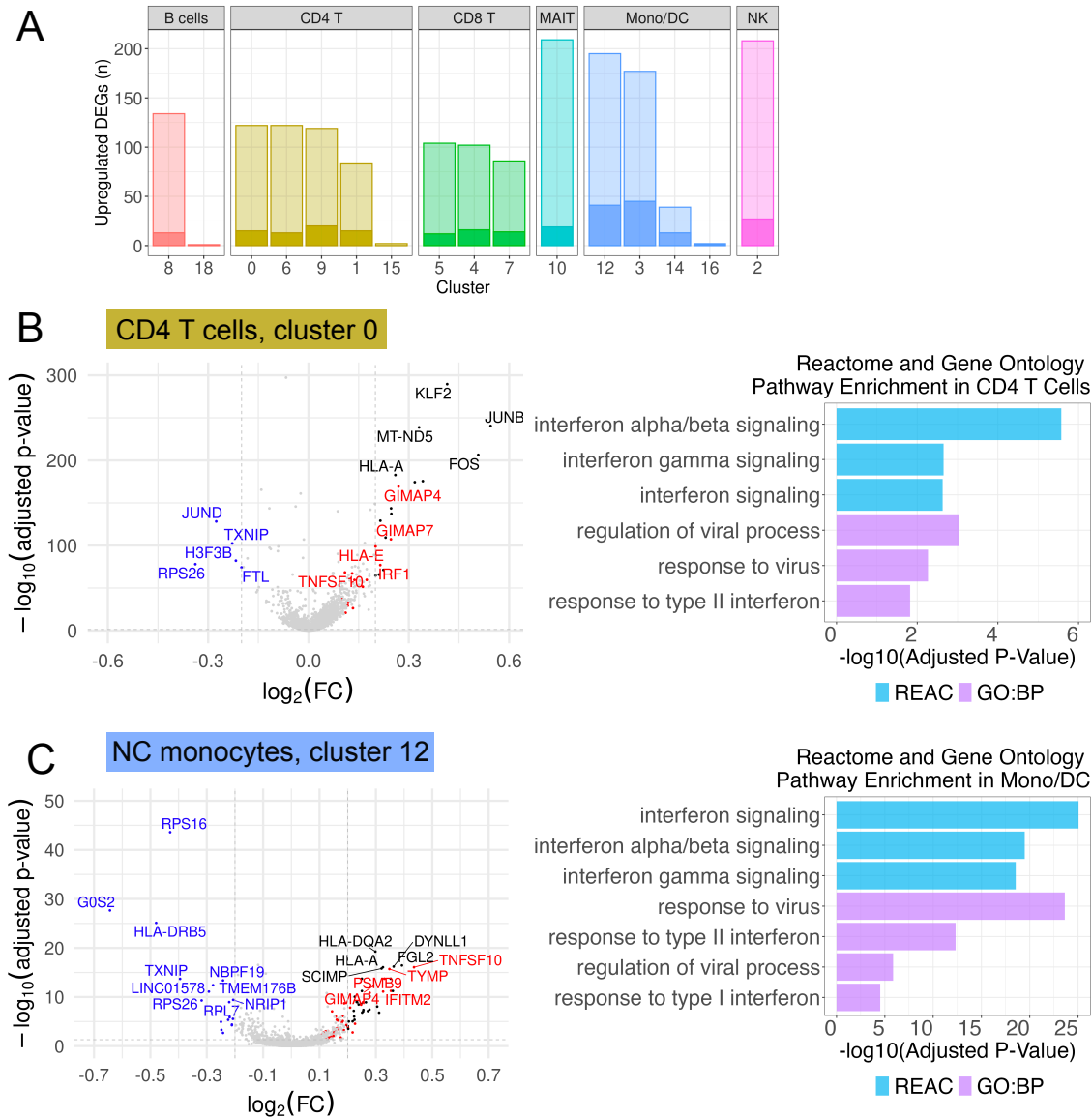


Figure 3.2.S5 Differential expression analysis in additional PBMC types in EOAD

A) Significantly upregulated genes are displayed across clusters, which are organized by PBMC type. The darker portion of each bar indicates the fraction of upregulated genes that are also ISAG^{hi} T-cell marker genes. B,C) Volcano plots for CD4 T cells and NC monocytes (clusters 0 and 12, respectively) display downregulated (blue) and upregulated (black) differentially expressed genes (DEGs); a subset of those with $PFDR < 0.05$ and absolute LFC > 0.25 are labeled on the plots. In addition, a subset of upregulated genes with $PFDR < 0.05$ and LFC > 0.1 that are also ISAG^{hi} T-cell marker genes are labeled in red. A maximum of 5 genes are labeled for each category to improve readability. Functional enrichment analysis of the upregulated genes ($PFDR < 0.05$ and LFC > 0.1) in CD4 T cell clusters (B) and in monocyte/dendritic cell clusters (C). There was significant enrichment of IFN and antiviral response pathways in the CD4 T cell clusters (0, 1, 6, 9, and 15) and the monocyte/dendritic cell clusters (3, 12, 14, and 16) as sourced from the GO BP and reactome databases.

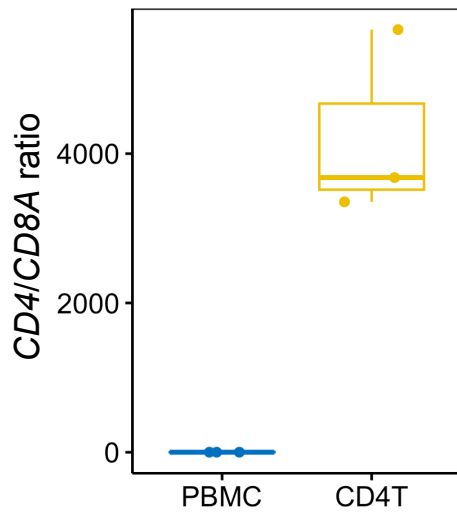


Figure 3.2.S6 ddPCR validation of CD4 T-cell isolation

ddPCR-based quantification of the *CD4/CD8A* ratio from starting PBMC RNA and isolated CD4 T-cell RNA revealed an ~4,800-fold increase in the *CD4/CD8A* ratio upon CD4 T-cell isolation.

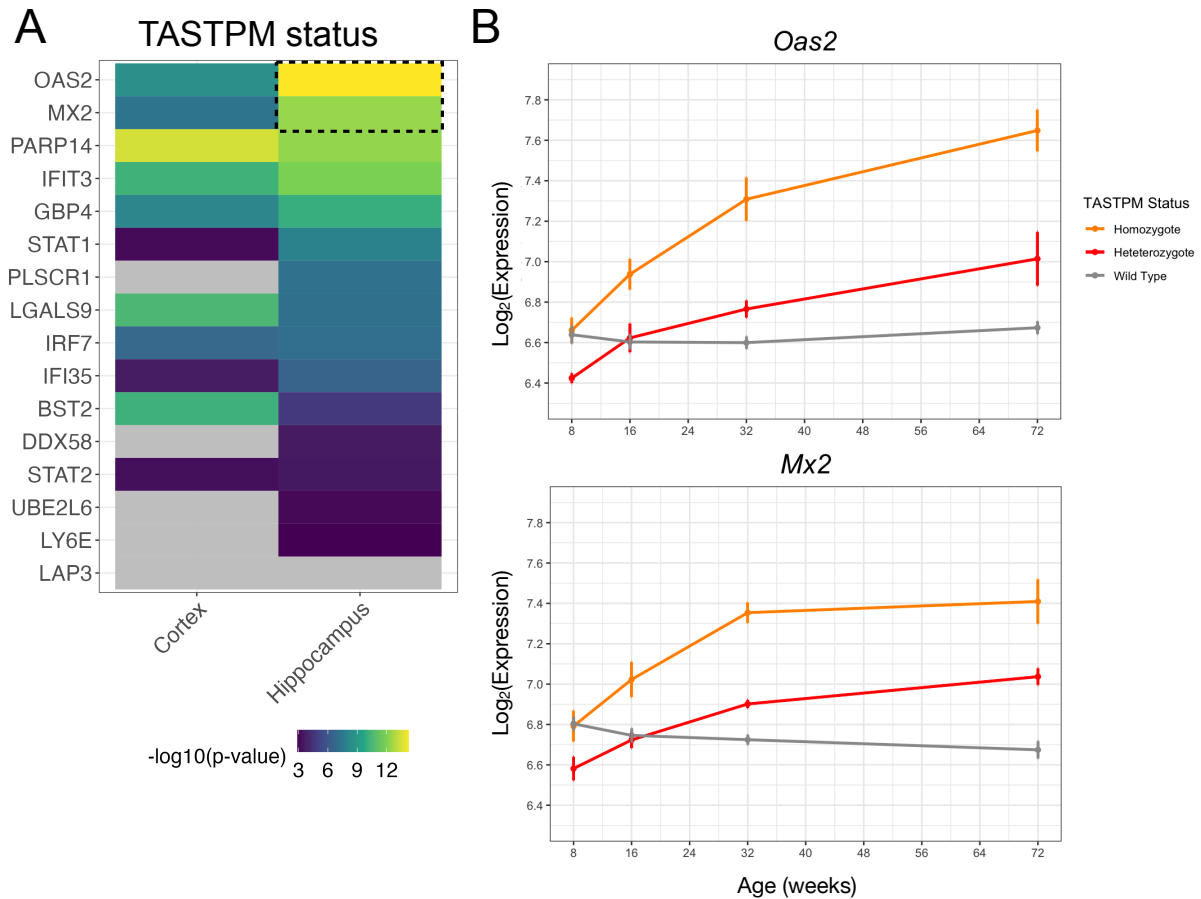


Figure 3.2.S7 Dysregulation of IFN response genes in a mouse model of familial EOAD
 A) All queried genes are displayed by heatmap in descending order of significance for TASTPM status in the hippocampus. Results are also displayed for the cortex. A total of 15 of the 16 genes queried showed significant dysregulation in mouse hippocampus. Genes not reaching significance for a given region are displayed as gray within the heatmap. B) The top two upregulated genes in the hippocampus (*Oas2* and *Mx2*) showed increased expression over time in TASTPM mice relative to WT mice, with TASTPM homozygotes showing greater increases relative to heterozygotes.

Table 3.2.1 Demographic and experimental information for samples used in scRNA-seq and ddPCR studies

	scRNA-seq discovery study		ddPCR validation study	
	Controls	Cases	Controls	Cases
<i>n</i>	8	8	10	9
<i>n</i> per batch (Batch A, Batch B)	4, 4	4, 4	5, 5	4, 5
PBMCs analyzed, <i>n</i>	91,955	90,398	N/A	N/A
CD4 T-cell RIN, mean (SD)	N/A	N/A	9.5 (0.5)	9.5 (0.3)
Sex, <i>n</i> female (%)	4 (50)	4 (50)	7 (70)	6 (66.7)
Age, mean (SD)	44.6 (7.4)	59.9 (3.1)	56.9 (5.0)	58.6 (2.7)
<i>APOE</i> ε4 status <i>n</i> heterozygous, <i>n</i> homozygous	3, 0	5, 1	2, 0	5, 1
Clinical syndrome (<i>n</i>)	clinically normal (8)	AD (6), frontal AD (2)	clinically normal (10)	AD (9)
Global CDR, mean (SD)	0.0 (0.0)	1.1 (0.4)	0.1 (0.2)	0.9 (0.2)

3.2.8 Supplemental Online Content

All supplemental online content can be downloaded directly from the preprint material at BioRxiv.

3.2.9 References

1. 2023 Alzheimer's disease facts and figures. *Alzheimers Dement* 2023;19:1598–695.
2. Sirkis DW, Bonham LW, Johnson TP, La Joie R, Yokoyama JS. Dissecting the clinical heterogeneity of early-onset Alzheimer's disease. *Mol Psychiatry* 2022;27:2674–88.
3. Gate D, Saligrama N, Leventhal O, Yang AC, Unger MS, Middeldorp J, et al. Clonally expanded CD8 T cells patrol the cerebrospinal fluid in Alzheimer's disease. *Nature* 2020; 577:399–404.
4. Gate D, Tapp E, Leventhal O, Shahid M, Nonninger TJ, Yang AC, et al. CD4 T cells contribute to neurodegeneration in Lewy body dementia. *Science* 2021;374:868–74.
5. Sirkis DW, Warly Solsberg C, Johnson TP, Bonham LW, Sturm VE, Lee SE, et al. Single-cell RNA-seq reveals alterations in peripheral CX3CR1 and nonclassical monocytes in familial tauopathy. *Genome Med* 2023;15:53.
6. Piehl N, van Olst L, Ramakrishnan A, Teregulova V, Simonton B, Zhang Z, et al. Cerebrospinal fluid immune dysregulation during healthy brain aging and cognitive impairment. *Cell* 2022;185:5028–39.e13.
7. Heming M, Li X, Räuber S, Mausberg AK, Börsch A-L, Hartlehnert M, et al. Neurological Manifestations of COVID-19 Feature T Cell Exhaustion and Dedifferentiated Monocytes in Cerebrospinal Fluid. *Immunity* 2021;54:164–75.e6.
8. Hao Y, Hao S, Andersen-Nissen E, Mauck WM 3rd, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. *Cell* 2021;184:3573–87.e29.
9. Miller ZA, Mandelli ML, Rankin KP, Henry ML, Babiak MC, Frazier DT, et al. Handedness and language learning disability differentially distribute in progressive aphasia variants. *Brain* 2013;136:3461–73.

10. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* 1984;34:939–44.
11. McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR Jr, Kawas CH, 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. *Alzheimers Dement* 2011;7:263–9.
12. Jack CR Jr, Bennett DA, Blennow K, Carrillo MC, Dunn B, Haeberlein SB, et al. NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimers Dement* 2018;14:535–62.
13. Ossenkoppele R, Pijnenburg YAL, Perry DC, Cohn-Sheehy BI, Scheltens NME, Vogel JW, et al. The behavioural/dysexecutive variant of Alzheimer's disease: clinical, neuroimaging and pathological features. *Brain* 2015;138:2732–49.
14. Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology* 1993;43:2412–4.
15. Wang X, Shen X, Chen S, Liu H, Hong N, Zhong H, et al. Reinvestigation of Classic T Cell Subsets and Identification of Novel Cell Subpopulations by Single-Cell RNA Sequencing. *J Immunol* 2022;208:396–406.
16. Amand J, Fehlmann T, Backes C, Keller A. DynaVenn: web-based computation of the most significant overlap between ordered sets. *BMC Bioinformatics* 2019;20:743.

17. Matarin M, Salih DA, Yasvoina M, Cummings DM, Guelfi S, Liu W, et al. A genome-wide gene-expression analysis and database in transgenic mice during development of amyloid or tau pathology. *Cell Rep* 2015;10:633–44.
18. Chen Y, Colonna M. Spontaneous and induced adaptive immune responses in Alzheimer's disease: new insights into old observations. *Curr Opin Immunol* 2022;77:102233.
19. Eyting M, Xie M, Heß S, Heß S, Geldsetzer P. Causal evidence that herpes zoster vaccination prevents a proportion of dementia cases. *medRxiv* 2023. <https://doi.org/10.1101/2023.05.23.23290253>.
20. Levine KS, Leonard HL, Blauwendraat C, Iwaki H, Johnson N, Bandres-Ciga S, et al. Virus exposure and neurodegenerative disease risk across national biobanks. *Neuron* 2023;111:1086–93.e2.
21. Albanese A, Chen E, Zavistaski J, Betz K, Suessmuth Y, Cagnin L, et al. Single-cell RNA-seq reveals an interferon-driven inflammatory CD4 naïve T cell subpopulation at day 100 in hematopoietic stem cell transplant patients that ultimately develop chronic GvHD. *Blood* 2022;140:270–1.
22. Breen MS, Maihofer AX, Glatt SJ, Tylee DS, Chandler SD, Tsuang MT, et al. Gene networks specific for innate immunity define post-traumatic stress disorder. *Mol Psychiatry* 2015;20:1538–45.
23. Katrinli S, Oliveira NCS, Felger JC, Michopoulos V, Smith AK. The role of the immune system in posttraumatic stress disorder. *Transl Psychiatry* 2022;12. <https://doi.org/10.1038/s41398-022-02094-7>.

24. Madera S, Rapp M, Firth MA, Beilke JN, Lanier LL, Sun JC. Type I IFN promotes NK cell expansion during viral infection by protecting NK cells against fratricide. *J Exp Med* 2016;213:225–33.
25. Chen X, Firulyova M, Manis M, Herz J, Smirnov I, Aladyeva E, et al. Microglia-mediated T cell infiltration drives neurodegeneration in tauopathy. *Nature* 2023;615:668–77.
26. Lee S-H, Rezzonico MG, Friedman BA, Huntley MH, Meilandt WJ, Pandey S, et al. TREM2-independent oligodendrocyte, astrocyte, and T cell responses to tau and amyloid pathology in mouse models of Alzheimer disease. *Cell Rep* 2021;37:110158.
27. Roy ER, Wang B, Wan Y-W, Chiu G, Cole A, Yin Z, et al. Type I interferon response drives neuroinflammation and synapse loss in Alzheimer disease. *J Clin Invest* 2020;130:1912–30.
28. Rexach JE, Polioudakis D, Yin A, Swarup V, Chang TS, Nguyen T, et al. Tau Pathology Drives Dementia Risk-Associated Gene Networks toward Chronic Inflammatory States and Immunosuppression. *Cell Rep* 2020;33:108398.
29. Sayed FA, Kodama L, Fan L, Carling GK, Udeochu JC, Le D, et al. AD-linked R47H-mutation induces disease-enhancing microglial states via AKT hyperactivation. *Sci Transl Med* 2021;13:eabe3947.
30. Wang C, Fan L, Khawaja RR, Liu B, Zhan L, Kodama L, et al. Microglial NF- κ B drives tau spreading and toxicity in a mouse model of tauopathy. *Nat Commun* 2022;13:1969.
31. Udeochu JC, Amin S, Huang Y, Fan L, Torres ERS, Carling GK, et al. Tau activation of microglial cGAS-IFN reduces MEF2C-mediated cognitive resilience. *Nat Neurosci* 2023;26:737–50.

32. Prater KE, Green KJ, Mamde S, Sun W, Cochoit A, Smith CL, et al. Human microglia show unique transcriptional changes in Alzheimer's disease. *Nat Aging* 2023;3:894–907.
33. Mesquita SD, Ferreira AC, Gao F, Coppola G, Geschwind DH, Sousa JC, et al. The choroid plexus transcriptome reveals changes in type I and II interferon responses in a mouse model of Alzheimer's disease. *Brain Behav Immun* 2015;49:2

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:

C6E6352E4577402... Author Signature

2/13/2024
Date