A public resource of single cell transcriptomes and multiscale networks from persons with and without Alzheimer's disease

Qi Wang[1], Jerry Antone[2], Eric Alsop[2], Rebecca Reiman[2], Cory Funk[3], Jaroslav Bendl[4], Joel T Dudley[1], Winnie S Liang[2], Timothy L Karr[1], Panos Roussos[4], David A Bennett[5], Philip L De Jager[6], Geidy E Serrano[7], Thomas G Beach[7], Kendall Van Keuren-Jensen[2], Diego Mastroeni[1], Eric M Reiman[8X], Benjamin P Readhead[1X*]

Author affiliations:

[1]ASU-Banner Neurodegenerative Disease Research Center, Arizona State University, Tempe, AZ, 85281, USA

[2]Division of Neurogenomics, The Translational Genomics Research Institute, Phoenix, AZ, 85004, USA

[3]Institute for Systems Biology, Seattle, WA, 98109, USA

[4]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, 10029, USA

[5]Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, 60612, USA

[6]Department of Neurology, Center for Translational and Computational Neuroimmunology, Columbia University Irving Medical Center, New York, NY, 10032, USA

[7]Civin Laboratory for Neuropathology, Banner Sun Health Research Institute, Sun City, AZ, 85351, USA

[8]Banner Alzheimer's Institute, Phoenix, AZ, 85006, USA

[X] These authors contributed equally

[*] Correspondence: Benjamin P. Readhead

ASU-Banner Neurodegenerative Disease Research Center, Arizona State University, Tempe, AZ, 85281, USA

E-mail: ben.readhead@asu.edu

Abstract

The emergence of technologies that can support high-throughput profiling of single cell transcriptomes offers to revolutionize the study of brain tissue from persons with and without Alzheimer's disease (AD). Integration of these data with additional complementary multiomics data such as genetics, proteomics and clinical data provides powerful opportunities to link observed cell subpopulations and molecular network features within a broader disease-relevant context. We report here single nucleus RNA sequencing (snRNA-seq) profiles generated from superior frontal gyrus cortical tissue samples from 101 exceptionally well characterized, aged subjects from the Banner Brain and Body Donation Program in combination with whole genome sequences. We report findings that link common AD risk variants with *CR1* expression in oligodendrocytes as well as alterations in peripheral hematological lab parameters, with these observations replicated in an independent, prospective cohort study of ageing and dementia. We also observed an AD-associated CD83(+) microglial subtype with unique molecular networks that encompass many known regulators of AD-relevant microglial biology, and which are associated with immunoglobulin IgG4 production in the transverse colon. These findings illustrate the power of multi-tissue molecular profiling to contextualize snRNA-seq brain

transcriptomics and reveal novel disease biology. The transcriptomic, genetic, phenotypic, and network data resources described within this study are available for access and utilization by the scientific community.

Introduction

Single cell sequencing technologies such as snRNA-seq in combination with steady development of analytical methods have greatly advanced our understanding of complex human diseases in the past decade[3]. These techniques have been employed to study Alzheimer's disease (AD), a devastating neurodegenerative disease characterized by the development of brain neuropathologies including neuritic plaques and neurofibrillary tangles leading to impaired cognition, with the goal of interpreting dynamic molecular processes within and across cell types[4-13]. Delineating cell type specific changes and dysregulation in AD at the single cell level is crucial for deciphering the molecular mechanisms underpinning the onset and progression of AD, and thus enabling the discovery of novel drug targets and development of effective therapeutic strategies[14].

Findings from large scale genetic studies of AD risk have convincingly implicated microglial biology as a critical causal component of AD onset and progression, including important roles in amyloid clearance[15] and immune response in the presence of tau pathology[16]. These responses involve a specific transcriptional state referred to as Activation Response Microglia (ARM)[17], Disease-Associated Microglia (DAM)[18], or MicroGlial neuroDegenerative phenotype (MGnD)[19], which demonstrate activation signature genes that overlap considerably with AD risk genes identified in genome-wide association studies[20] (GWAS). These activation signatures, however, have not been fully captured in several recently reported snRNA-seq studies of microglia in frozen human AD postmortem brain tissues[10-12]. It has been reported that the expressions of these genes are depleted in nuclei compared with cellular expression (as reflected in snRNA-seq versus single cell RNA-seq) assayed from human microglia[21]. Further investigations are thus warranted to explore the utility of snRNA-seq for studying microglia activation, particularly in the context of AD.

In this study we generated snRNA-seq profiles from 481,840 nuclei collected from postmortem superior frontal gyrus (SFG) cortical tissue samples from 101 aged subjects with excellent clinical and postmortem neuropathological characterizations from the Arizona Study of Aging and Neurodegenerative Disorders/Brain and Body Donation Program (BBDP)[22]. By integrating whole genome sequencing (WGS) data, we report findings that link common AD risk variants with *CR1* expression in oligodendrocytes as well as alterations in peripheral hematocrit levels. We also applied multiscale network modelling approaches to learn the gene regulatory networks that characterize AD-associated cell subpopulations. Our findings have revealed a specific CD83(+) microglial subtype with unique molecular networks that encompass many known regulators of AD-relevant microglial biology, and which are associated with novel immunoglobulin production in the transverse colon. These findings demonstrate the power of multi-tissue molecular profiling to contextualize single nucleus brain transcriptomics and thus illuminate novel disease biology. The transcriptomic, genetic, phenotypic, and network data resources described within this study are available for access and utilization by the scientific community.

Results

A public resource of single cell transcriptome and other associated molecular data

We developed a shared resource of snRNA-seq data from SFG along with WGS profiles, from exceptionally high-quality brain tissue from 101 well characterized brain donors with and without the clinical and neuropathological features of AD. Detailed demographic, clinical, and postmortem neuropathological data from this cohort are reported in **Table 1A**. We used the National Institute on Aging and Alzheimer's Association

(NIA-AA) AD criteria[23] from neuropathological characterization to dichotomize the subjects, to facilitate targeted comparison in this study.

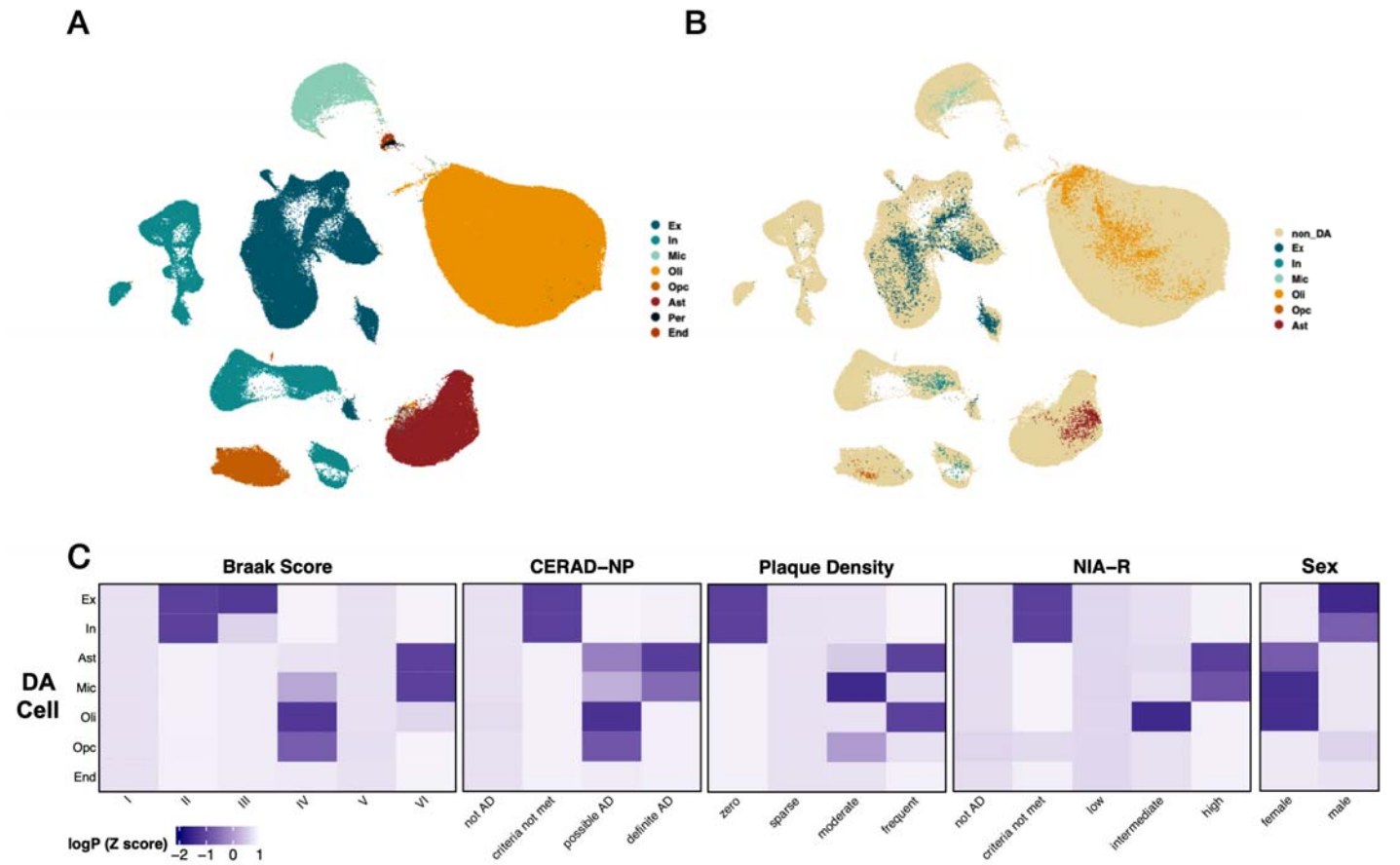


Figure 1: (A) UMAP plot of SFG snRNA-seq data from 481,840 cells collected from 101 subjects (66 AD, 35 Controls) with **(B)** differentially abundant (DA) cell types observed in all major cell types, (C) with varying associations with clinicopathological AD traits. DA: Differentially Abundant, SFG: Superior Frontal Gyrus, Ex: Excitatory Neurons, In: Inhibitory Neurons, Mic: Microglia, Oli: Oligodendrocytes, Opc: Oligodendrocyte Progenitor Cells, Ast: Astrocytes

Differentially abundant cell types in Superior Frontal Gyrus of Alzheimer's disease

We performed Chromium 10x snRNA-seq on postmortem, SFG brain tissue samples from 101 donors (AD n = 66, Aged Controls n = 35). We applied a set of rigorous quality control criteria to exclude low quality nuclei or doublets from each sample, then integrated all nuclei into a single data object (see Methods). After cell filtering we retained a total of 481,840 nuclei which were used for all downstream analyses. Unsupervised clustering by UMAP identified 24 cell clusters, which were then annotated with cell type via mapping against a reference dataset[11] and transferring cell labels (**Figure 1A, Supplementary Figure 1A, Supplementary Table 1**).

We then analyzed the data in a supervised, targeted approach, applying DAseq,[24] a multiscale approach for detecting cell subpopulations with significant differential abundance (DA) between groups of interest (AD vs. Aged Controls). A DA population is defined as a cell subpopulation present at statistically higher or lower frequencies in one condition compared with another (AD vs Aged Control for this study). This approach supports rich, complex analyses within the context of observed cell subpopulations at the cohort level, in a manner that is not biased by inter-individual differences in cell type fractions. We identified a total of 9,345

cells from 11 distinct DA subpopulations across all major cell type classes represented, ranging from 131 to

4,505 cells in each population (**Supplementary Table 1**). The DA subpopulations were then collapsed into

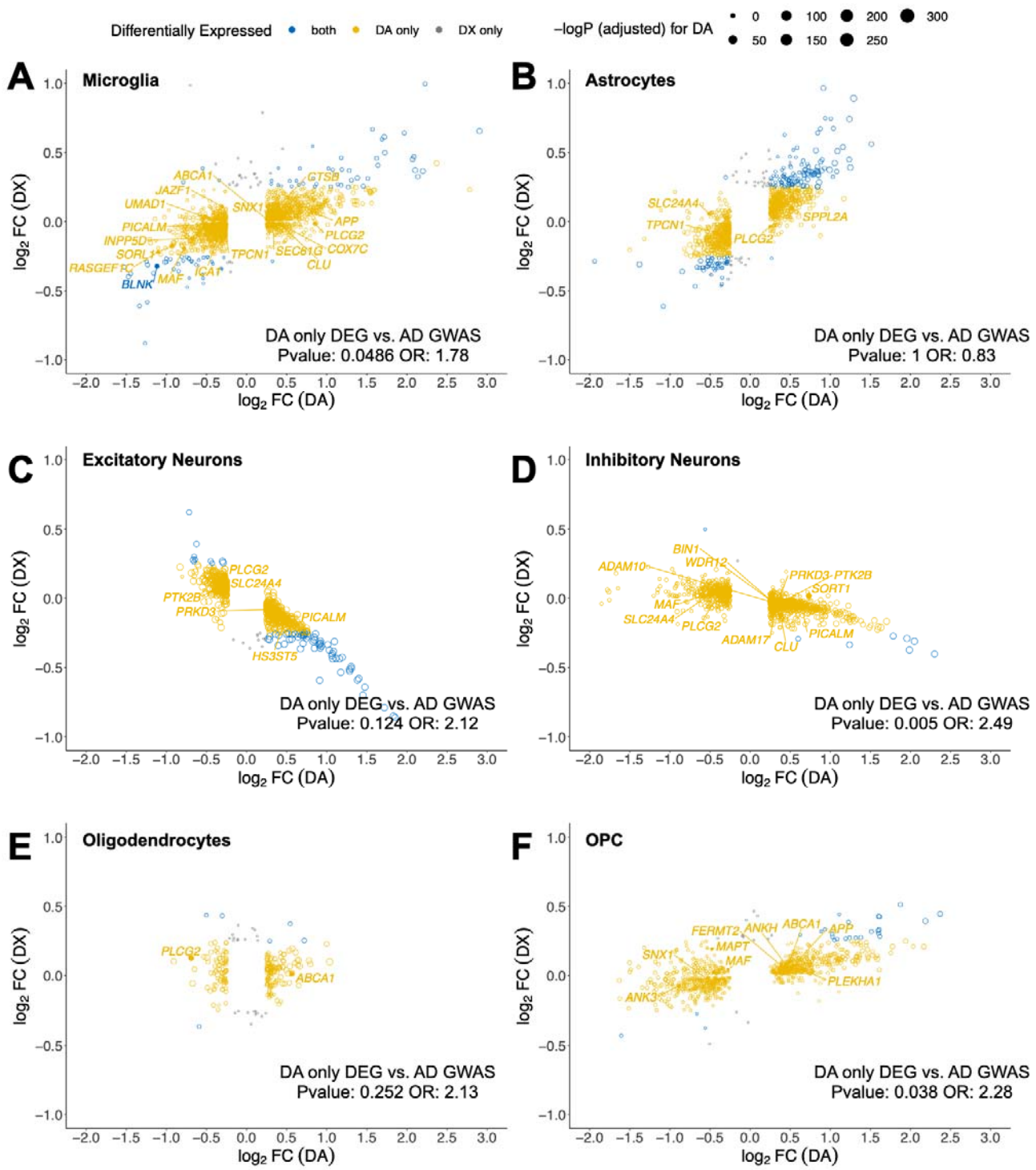
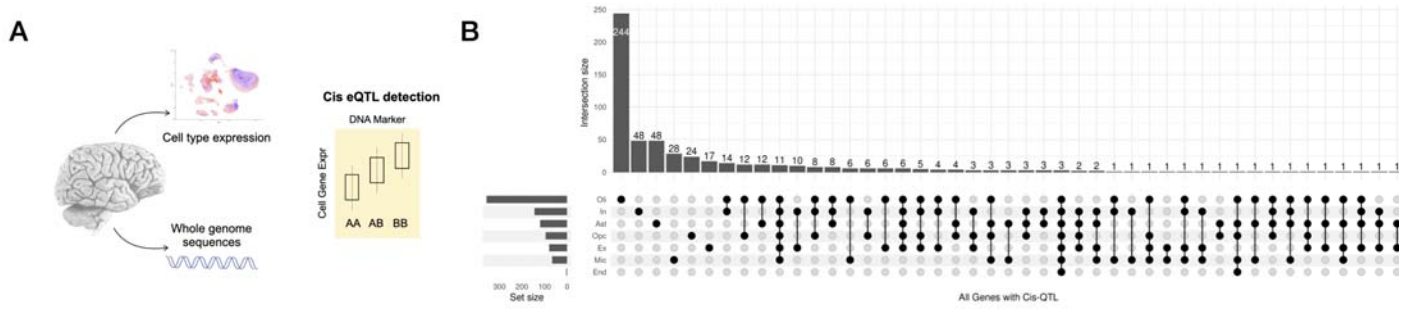


Figure 2. Log2 fold change for differentially expressed genes (DEGs) obtained by two different comparisons obtained for major cell populations. DA: Differentially abundant nuclei vs non differentially abundant nuclei; DX: AD nuclei vs age-matched control nuclei. Circle size is proportional to the -log of the adjusted pvalue observed in the DA contrast. AD-associated[2] signature genes are labelled. P-value and Odds Ratio (OR) are shown for overlap between genes identified as DEG only using the DA approach (orange) and a set of AD GWAS associated candidate genes[2] (two-sided Fisher's Exact Test). OPC: Oligodendrocyte Precursor Cells

their respective cell type clusters and annotated as the DA cluster for each cell type (**Figure 1B,**



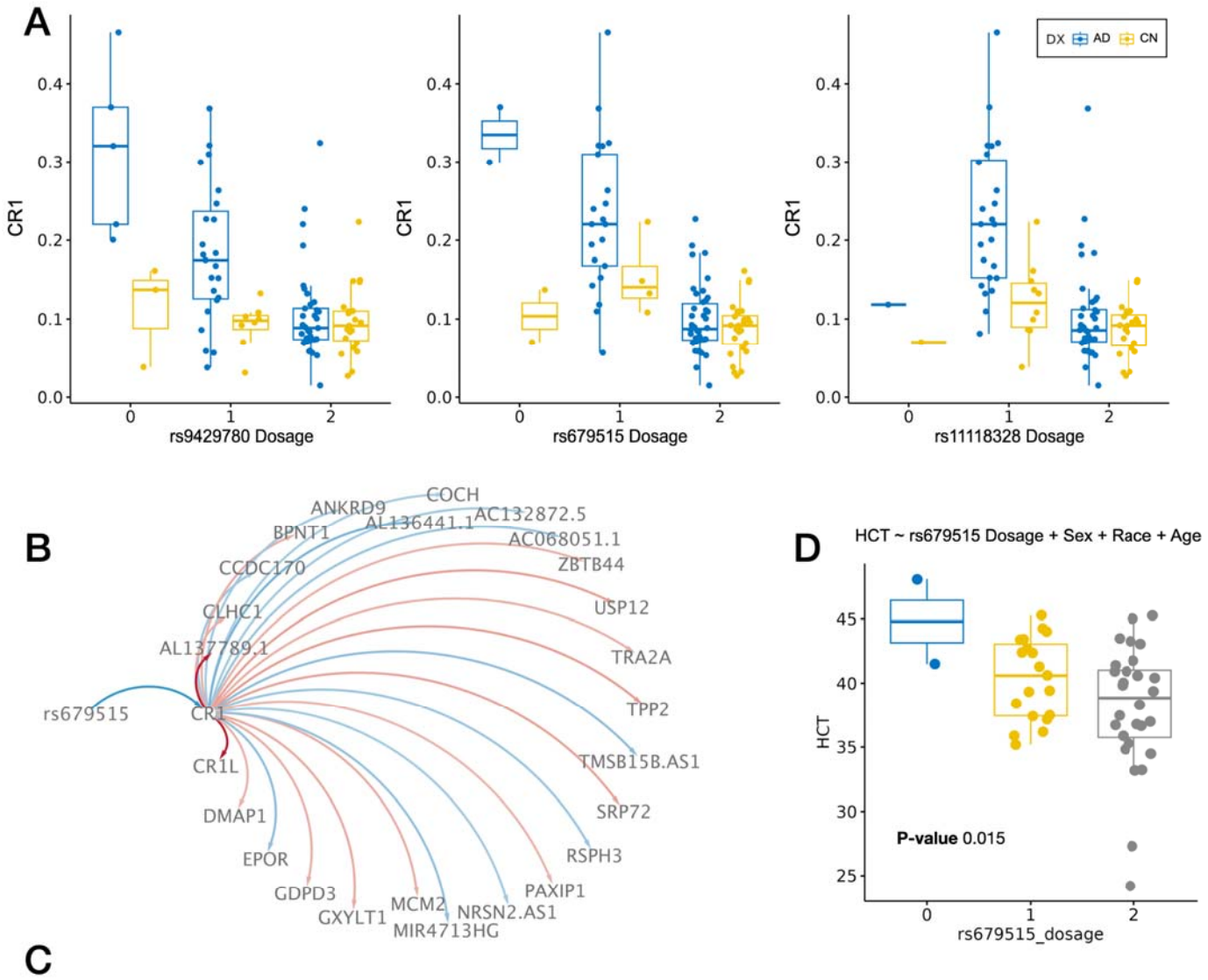
C

Cell type	SNP	eQTL Gene	eQTL Beta	eQTL FDR	GWAS Traits Assoc with SNP
Oli	1:207677194	CR1	0.08	2.22E-05	Alzheimer's disease
Ast	1:27893147	RPA2	-0.06	6.42E-06	HDL cholesterol
Mic, Oli	11:47654618	NUP160, MTCH2	-0.14, -0.02	2.96E-04, 2.26E-3	Cortical thickness, Neuroticism, Depressed Affect
Mic	11:85941783	PICALM	0.17	7.03E-20	Educational Attainment
In	12:56752285	PRIM1	0.06	1.16E-04	Menopause|Fat-free mass|Type 2 DM|RBC Count
Oli	12:6451407	CD27-AS1	-0.03	2.46E-03	Multiple sclerosis
Ast, Opc	15:45167366	AC051619.5	-0.21, -0.09	3.72E-04, 1.86E-04	Vertex-wise sulcal depth
Oli	16:89782151	SPIRE2	-0.01	3.18E-04	C-reactive protein levels
In	16:89793345	GAS8	-0.14	1.23E-04	Homocysteine levels
Ast	17:46786047	LRRC37A2	0.10	4.37E-04	White matter microstructure
Ast	17:46788073	LRRC37A2	-0.12	6.87E-05	Cortical surface area | General cognitive ability | Neuroticism
Ast, Oli	18:12945311	CEP192|	0.06, 0.06	3.26E-02, 7.27e-07	Intelligence
Opc	18:77247169	GALR1	-0.19	4.96E-06	Theta electroencephalogram power
Ast, Oli, Oli, Oli	20:25349123	NINL, PYGB, NINL, ABHD12	0.08, 0.04, 0.03, -0.03	4.51E-04, 4.07E-10, 3.84E-06, 1.38E-02	Phosphatidylinositol (18:0_20:4) levels
Mic	6:32644620	HLA-DQB1	0.11	3.24E-05	Epstein Barr virus nuclear antigen 1 IgG levels

Figure 3 (A) Integration of whole genome sequences with cell type expression to detect cell type eQTL associations **(B)** with shared and cell type specific distribution **(C)** revealing multiple cell type eQTL linking with DNA loci that have been implicated in diverse neurological and neurodegenerative traits via GWAS data[1]. eQTL: Expression Quantitative Trait Loci, Ex: Excitatory Neurons, In: Inhibitory Neurons, Mic: Microglia, Oli: Oligodendrocytes, Opc: Oligodendrocyte Progenitor Cells, Ast: Astrocytes

Supplementary Figure 1B, Supplementary Table 1). Top gene expression markers which discriminate DA subpopulations from non-DA cells of the same class are shown (**Supplementary Table 2**). We then examined whether individual DA cell subpopulations are associated with clinical and neuropathological traits of interest to AD (**Figure 1C**). We observed that DA subpopulations over represented among AD subjects (compared with Aged Controls) are primarily composed of glial cells and those over represented among Aged Control subjects (compared with AD) are primarily composed of neuronal subtypes (**Supplementary Table 1**) , consistent with the marked gliosis[25] and neuronal loss[26] that characterizes AD. Furthermore, the differential association of DA subpopulations with varying stages of neuropathological severity is consistent with considerable cell type heterogeneity across the spectrum of AD.

We also observed that the identification of DA subpopulations offered advantages in detecting AD-linked genes beyond what we could identify through a more conventional differential expression analysis based on a cell-type specific comparison of AD vs. Aged Control. For each major brain cell type, we generated differentially expressed genes (**Figure 2**, DEG, $|\log_2 FC| > 0.25$ and Bonferroni adjusted $P < 0.05$) identified via a comparison of DA cells against non-DA cells (DA, **Figure 2**, X-axes) as well as a comparison of AD cells against Aged Control cells (DX, **Figure 2**, Y-axes). For all cell types, the range of DEG $\log_2 FC$ was larger for the DA based comparisons, with a commensurately larger set of DEG identified exclusively using that comparison (**Supplementary Table 2**). This included DEG that were detected solely within a comparison of DA against non-DA cells ('DA only') as well as comparatively fewer DEG detected only within a comparison of AD against Aged Control for each major cell type ('DX only'). We also observed a significant enrichment of AD GWAS risk loci prioritized genes[2] among the DA only genes in microglia, inhibitory neurons and OPC. Overall, these findings are supportive of the potential for DA based cell clustering to demarcate biologically informative cell groupings that might be used to frame further analyses.



C

Gene Set Library	Gene Set	FDR	Fold change	Genes in overlap
Descartes Cell Types and Tissue 2021	Erythroblasts in Kidney	1.22E-03	20.54	ANKRD9, CR1L, EPOR
Descartes Cell Types and Tissue 2021	Erythroblasts in Intestine	1.61E-03	18.65	ANKRD9, CR1L, EPOR
MGI Mammalian Phenotype Level 4 2021	abnormal erythrocyte morphology MP:0002447	1.02E-02	22.97	EPOR, MCM2
Rare Diseases AutoRIF ARCHS4 Predictions	Stomatocytosis I	1.70E-04	18.70	ANKRD9, CR1L, EPOR, USP12
Rare Diseases GeneRIF ARCHS4 Predictions	Pyropoikilocytosis hereditary	2.31E-04	17.27	ANKRD9, CR1L, EPOR, USP12
Rare Diseases GeneRIF ARCHS4 Predictions	Hereditary spherocytosis	1.93E-03	17.53	ANKRD9, CR1L, EPOR
Rare Diseases GeneRIF ARCHS4 Predictions	Erythropoietic protoporphyria	3.76E-03	13.93	CR1L, EPOR, USP12
Rare Diseases AutoRIF ARCHS4 Predictions	Hemoglobin sickle-beta thalassemia	2.47E-02	14.58	CR1L, EPOR

Figure 4 (A) AD specific eQTL association between AD GWAS loci dosage and *CR1* expression within oligodendrocytes **(B)** Causal inference network illustrating conditional relationships between AD GWAS locus rs679515, *CR1* expression and downstream molecular networks including impacts on EPO receptor expression. **(C)** Molecular functional enrichments of the rs679515 / *CR1* network revealed themes of erythrocyte biology and hematological traits. **(D)** Antemortem hematocrit levels from 49 AD subjects within the current study, stratified according to rs679515 dosage. Enrichments based on Fisher's Exact Test. HCT: hematocrit, FDR: False Discovery Rate. Major allele dosages are shown.

Cell type resolution of expression Quantitative Trait Loci reveals novel linkage between regulatory variants and neurological traits

We integrated genotype data of 99 subjects based on WGS from these 101 subjects with individual cell type expression profiles to determine cell type specific expression quantitative trait loci (eQTL, **Figure 3A**) which can then be directly investigated as well as used as inputs into downstream analyses including causal network modelling. We adopted a conservative approach to the detection of eQTL using snRNA-seq data which includes scran normalization, mean aggregation of expression across specific cells from one subject; incorporation of principal components as covariates in the associated linear mixed modelling and accounting for multiple testing by using the conditional false discovery rate, leveraging the recently published eQTL summary statistics from the AMP-AD meta-analysis of bulk RNA-seq from cortical tissue as an external reference set[27]. We classified cis eQTL associations with an FDR <= 0.05 as significant, detecting a total of 1,185 associations from 560 genes (**Figure 3B**). Consistent with what have been observed[28,29], the majority of the eQTL associated genes (eGenes) were detected only in a single cell type (n = 409), supporting the utility of snRNA-seq in resolving regulatory relationships that may not be evident at the level of bulk tissue transcriptomics (**Supplementary Table 3**).

Expression QTL relationships have been determined to be enriched for disease risk genetics[30] and offer a means to contextualize and identify mediating mechanisms that link genetic variation with disease phenotypes. We thus annotated each eQTL locus with any published associations published by the NHGRI-EBI GWAS Catalog[1]. We observed many instances of eQTL loci which have been previously published in association with traits that link with AD and associated risk factors and comorbidities, such as cortical thickness, educational attainment, Type 2 diabetes mellitus, homocysteine levels, and HDL cholesterol levels (**Figure 3C**).

Alzheimer's GWAS locus rs679515 impacts on *CR1* networks in oligodendrocytes, known erythrocyte regulators and peripheral blood hematocrit

We observed that reported two AD risk loci, rs11118328[31] (chr1:207677194) and rs9429780[2] (chr1: 1:207493845) are both eQTLs for Complement Receptor 1 (*CR1*) expression in oligodendrocytes. Recently, Fujita et al[28] reported snRNA-seq from 424 dorsolateral prefrontal cortex samples including cell-subtype specific eQTL relationships in the ROSMAP cohort[32]. Using this data (ROSMAP DLPFC snRNA-seq) we confirmed the association between both loci and *CR1* expression in oligodendrocytes (**Supplementary Table 4**). Fujita et al[28] also reported the nearby most significant AD risk locus rs679515 (chr1:207577223, Major allele: C, Minor allele: T, Reference allele: T) as an eQTL in oligodendrocytes, although this variant had not emerged in our analysis. Examination of our eQTL results revealed that rs679515 had been removed following linkage disequilibrium-based pruning with rs9429780. We then confirmed that rs679515 was also an eQTL for *CR1* in oligodendrocytes in our dataset by a direct association test, in agreement with Fujita et al[28] (**Supplementary Table 4**), where minor allele T (also the AD risk allele) is associated with higher CR1 expression. Conditional eQTL analyses were consistent with the strongest independent eQTL association being present for rs679515, with some independent signal also observed for rs9429780, though not rs11118328. Further examination of the association between these three variants and oligodendrocyte expression of *CR1* revealed that this association was driven by the association within AD subjects (**Figure 4A**, **Supplementary Table 4**), indicating an AD-specific interaction (AD diagnosis/dosage interaction pvalue = 2.1e-5) between allele dosage and *CR1* expression within the SFG samples. We did not observe a significant AD-specific interaction within the ROSMAP DLPFC cohort, with significant associations between rs679515 and CR1 expression observed within both AD and control subjects, (**Supplementary Figure 2A**, **Supplementary Table 5**). Whether the difference in detection of an AD-specific interaction between rs679515 and *CR1* expression is explained by brain region-specific differences, or differences between cohorts remains to be investigated.

Given the potential for cell type specific eQTL associations to illuminate the biological context of risk associated variants, we hypothesized that altered *CR1* expression mediates some fraction of the risk effect of

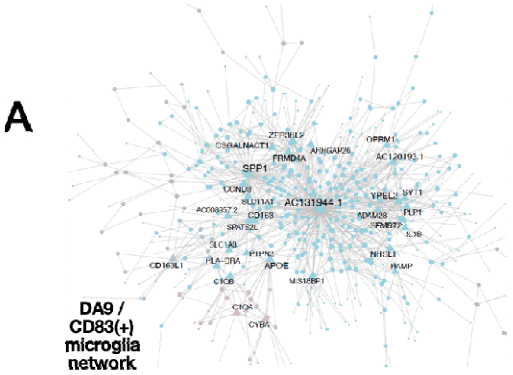
this allele. We therefore constructed a targeted gene regulatory network aimed at identifying genes that are downstream of *CR1*, conditioned on the relationship with rs679515, using a causal inference testing approach[47] applied to the 66 AD samples. We identified 25 significantly associated downstream genes (FDR < 0.05, **Figure 4B, Supplementary Table 4**) including Complement C3b/C4b Receptor 1 Like (*CR1L*) and Erythropoietin Receptor (*EPOR*). Molecular functional enrichments of the rs679515 / *CR1* network revealed themes of erythrocyte biology and hematological traits (**Figure 4C, Supplementary Table 4**), including enrichments for genes linked with erythrocyte signalling and morphology. Given the inclusion of *EPOR* and enrichments implicating hematological traits, we hypothesized that this may also manifest as differences in blood laboratory parameters within our study population. We performed a chart review of antemortem hematocrit values from 49 AD subjects within our study, collected during a time of comparatively stable health. Linear regression revealed a significant association (P-value: 0.015) between rs679515 major allele dosage and hematocrit (**Figure 4D**) while adjusting for sex, race and age at collection, where the major allele is associated with lower hematocrit in AD patients.

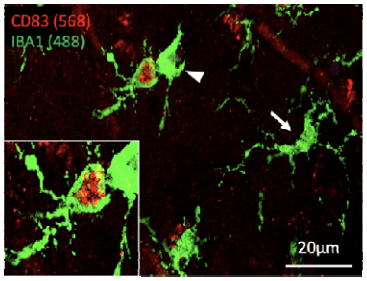
We then sought to determine whether rs679515 is also associated with antemortem hematological parameters (including hematocrit) within the ROSMAP cohort. We took the last available measurement within three years before death for each individual and examined the association of measurements with the rs679515 major allele dosage. We observed significant associations with several hematological traits, including hematocrit, hemoglobin concentration (Hb), red blood cell count (RBC), and anemia diagnosis (**Supplementary Table 5**). These hematological traits were also significantly associated with a variety of neuropathologies, particularly neurofibrillary tangle burden. Interestingly, we observed a neuropathology-associated inversion of the association between rs679515 major allele dosage and these hematological traits (**Supplementary Figure 2B**, **Supplementary Table 5)**, whereby major allele dosage was positively associated with hematocrit/RBC/Hb (and inversely associated with presence of anemia) in subjects with low NFT burden (Braak stage: 0-II), and inversely associated with hematocrit/RBC/Hb (and positively associated with presence of anemia) in subjects with high NFT burden (Braak stage: V-VI). Taken together, these findings suggest that rs679515 influences molecular networks across multiple tissues, with impacts on hematological traits that occur an AD/neuropathology-specific manner.

Multiscale network analysis of AD associated cell types reveals novel molecular drivers in CD83(+) microglia

We constructed cell type specific gene regulatory networks using two complementary network modelling approaches including a probabilistic causal network methodology (RIMBANet) and Multiscale Embedded Gene Expression Network Analysis (MEGENA) reviewed here[33]. We focused our initial efforts on the investigation of networks constructed from DA microglia subpopulation DA9 (**Figure 5**), which revealed directed networks containing many known microglial-relevant transcripts including secreted glycoprotein and Type 1 immunity regulator *SPP1*, *TREM2* and complement component 1q genes (*C1QA*, *C1QB*, *C1QC*). To better characterize the subpopulation, we compared the gene signatures present in DA9 to microglia signatures reported in the scientific literature[34]. Due to the large number of nuclei sampled in our study, many genes that are significantly differentially expressed in DA9 when compared with other nuclei in microglia demonstrate small log-fold changes (**Supplementary Table 2**). We observed downregulation of many homeostatic microglial genes, such as *TMEM119*, *P2RY12*, *P2RY13*, *CX3CR1* and *SELPLG*. Conversely, we observed upregulation of many of DAM signatures genes, such as *SPP1*, *APOE*, *CD9*, *CLEC7A*, *FTH1* and *FLT1* (**Supplementary Figure 3**). Although *TREM2* and *TYROBP* are expressed in the DA9 microglia, they are not differentially expressed when compared with other microglia cells or stratified according to AD / Aged Control status. We also noted overlap between DA9 differential expression and several dystrophic[35] and other microglia signatures[10,11] reported in the scientific literature (**Supplementary Figure 3**). We observed cell surface marker *CD83* as the most strongly overexpressed transcript (**Figure 5B, Supplementary Table 6**) among DA9 microglia. Herein we refer to DA9 microglia as CD83(+) microglia for brevity, however we note that these cells are characterized by several marker genes beyond *CD83* (**Supplementary Table 6)**, and that our data includes other microglia that express

CD83, but which are not members of the same DA subpopulation. We detected CD83(+) microglia in 47% of AD (n=31 of 66) and 25% of Aged Control (n=9 of 35) subjects profiled within the Banner SFG snRNA-seq study.





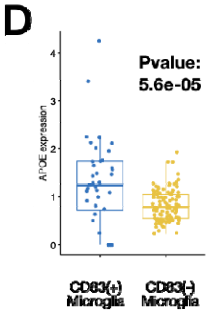
C

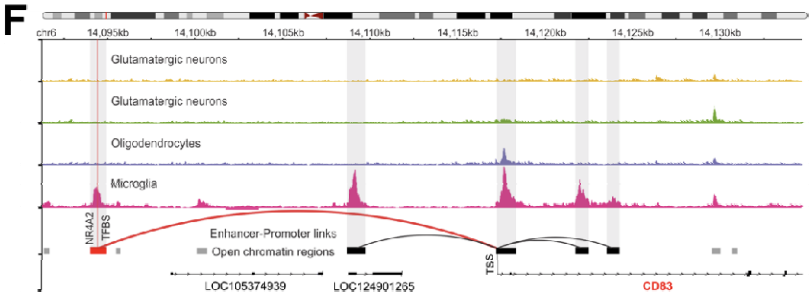
Network hubs gained in CD83(+) microglia subtype

	Differential connectivity CD83(+) vs CD83(-)	Pvalue
CD163L1	1.61	0.02
APOE	6.14	0.03
SPATS2L	5.88	0.05
HAMP	5.38	0.05
CYBA	2.97	0.09

E

Gene Set Library	Gene Set	FDR	Fold change	Genes in overlap
Jensen COMPARTMENTS	Ferritin complex	6.02E-07	4.37	APOE, BMP6, CD163, DECR1, FTH1, FTL, HAMP, HSPD1, HSPE1, IFIH1, IL1B, NAMPT, NCOA4, SLC11A1, SLC40A1, TNFRSF1B
Reactome 2022	Binding And Uptake Of Ligands By Scavenger Receptors R-HSA-2173782	4.47E-08	9.25	APOE, CD163, DLC1, FTH1, FTL, HSP90AA1, HSPH1
GO Molecular Function 2021	chemokine receptor activity (GO:0004950)	3.11E-05	15.86	CCR1, CMKLR1, CX3CR1, CXCR4
GO Biological Process 2021	response to unfolded protein (GO:0006986)	1.38E-04	5.29	BAG3, DNAJB1, HSP90AA1, HSPB1, HSPD1, HSPE1, HSPH1, SERPINH1
BioPlanet 2019	BDNF signaling pathway	1.44E-04	2.95	CDH2, DLC1, DNAJB1, DOCK4, GBE1, HSPB1, NFIB, NR4A1, NR4A2, NR4A3, P4HA1, PPIE, SERPINH1, SH3BP5, SPP1, STIP1
GO Biological Process 2021	modulation of chemical synaptic transmission (GO:0050804)	3.61E-04	4.20	CDH2, CNTNAP4, CX3CR1, DGKI, IL1B, NLGN4Y, SYN3, SYT1, TMEM108
Human Phenotype Ontology	Abnormality of iron homeostasis (HP:0011031)	3.74E-04	7.93	FTH1, FTL, HAMP, PUS1, SLC40A1
GO Biological Process 2021	complement activation, classical pathway (GO:0006958)	4.97E-04	15.86	C1QA, C1QB, C1QC
Reactome 2022	Classical Antibody-Mediated Complement Activation R-HSA-173623	4.97E-04	15.86	C1QA, C1QB, C1QC
Reactome 2022	Creation Of C4 And C2 Activators R-HSA-166786	4.97E-04	15.86	C1QA, C1QB, C1QC
GO Biological Process 2021	synapse pruning (GO:0098883)	9.33E-04	9.06	C1QA, C1QB, C1QC, CX3CR1
SynGO 2022	Integral Component Of Postsynaptic Membrane (GO:0099055) CC	1.61E-03	5.01	ADGRL2, CACNA1E, CDH2, GRIA1, KCNJ2, OPRM1
GO Biological Process 2021	humoral immune response mediated by circulating immunoglobulin (GO:0002455)	1.89E-03	11.89	C1QA, C1QB, C1QC
GO Biological Process 2021	positive regulation of lipid transport (GO:0032370)	1.89E-03	11.89	ABCA3, APOE, SPP1
Human Phenotype Ontology	Abnormality of complement system (HP:0005339)	1.89E-03	11.89	C1QA, C1QB, C1QC





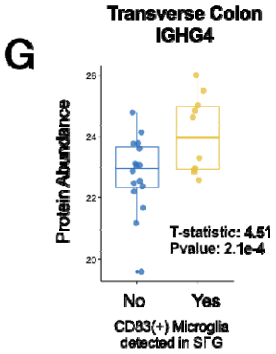


Figure 5 (A) Multiscale gene coexpression network constructed from differentially abundant CD83(+) microglial subpopulation reveals novel antisense transcripts functioning as network hubs. **(B)** Photomicrograph of CD83(+) microglia within Layer II of superior frontal gyrus in a Braak IV AD subject. Morphologically, CD83 reactive microglia appear ameboid (see insert) compared to unreactive microglia (arrow). **(C)** APOE gained as a network hub and **(D)** at increased expression in CD83(+) microglia compared with CD83(-) microglia. **(E)** Representative molecular functional enrichments among CD83(+) microglia hub genes. **(F)** Epigenomic landscape around CD83 gene in major brain cell types. Normalized genomics tracks demonstrate the complex cell-specific regulation of CD83 via enhancer-promoter links. Microglia tracks have overlaying layers for signal generated from AD case samples and controls samples separately. **(G)** Increased abundance of Immunoglobulin Heavy Chain protein IGHG4 protein in the transverse colon as a function of presence of CD83(+) microglia in superior frontal gyrus, while controlling for AD, Age, Sex and PMI (Samples N=26).

Key driver analyses revealed the most influential network node to be an antisense transcript *AC131944.1* (ENSG00000286618.1), which to our knowledge, has not been described in the context of microglial biology or AD, but which is encoded antisense to *SPP1*. In addition, we observed *APOE* as a key network driver in CD83(+) microglia, whereas we did not observe it as a network driver in non-CD83(+) microglia using a permutation-based network comparison approach (**Figure 5C**). In addition to its gain of network influence in CD83(+) microglia, we also observed greater expression of *APOE* in these cells compared with CD83(-) microglia (**Figure 5D**).

To understand whether CD83(+) microglia are transcriptionally adapted to specific biological pathways, we performed a gene set enrichment analysis upon the set of 308 hub genes (genes with a MEGENA connectivity score of ≥ 4, **Figure 5E, Supplementary Table 7**) computed against a background of 3083 genes that were robustly expressed among the full set of detected microglia (> 5% of the cells) within the study. We observed enrichments suggestive of complement activation, ferritin and iron processing, antibody immune response and lipid processing.

CD83(+) microglia are also detected in an independent cohort of single nucleus RNA-seq collected from dorsolateral prefrontal cortex

To evaluate whether CD83(+) microglia are also present in additional snRNA-seq data sets, we again accessed the ROSMAP DLPFC snRNA-seq[28] and performed DAseq using the same workflow we had applied to the Banner SFG snRNA-seq. We identified an AD-associated CD83(+) microglial subpopulation which expressed marker genes that extensively overlap the Banner SFG snRNA-seq defined CD83(+) microglial marker genes. In a comparison of CD83(+) microglia and non-CD83(+) microglia, we identified 312 DEG ($|\log_2 FC| > 0.5$, Bonferroni corrected $p < 0.05$) within the ROSMAP DLPFC snRNA-seq and 597 within the Banner SFG snRNA-seq, with an overlap of 139 genes (P-value: 4.04e-120, Odds ratio: 28.2, Fisher's Exact Test, **Supplementary Table 8**, **Supplementary Figure 4**) with shared over expression of many genes that had emerged during the SFG network analysis, including *CD83*, *AC131944.1*, *SPP1*, *APOE* and *ATP1B3*. We observed these DLPFC-derived CD83(+) microglia in 25% of AD subjects and 4% of Aged Control subjects.

CD83(+) microglia are modestly associated with increased multi-regional neuritic plaque burden

We next sought to understand whether there may be any technical, demographic, clinical or neuropathological differences between these groups. Within the Banner SFG snRNA-seq, we did not observe any difference in age of death, sex, ethnicity, post-mortem interval (PMI), neuritic plaque density, neurofibrillary tangle burden, clinical dementia rating or ApoE4 carriage rates between AD subjects with and without CD83(+) microglia (**Supplementary Table 9**). We also inspected whole body autopsy results available for 68 (AD n=45, Aged Control n=23) of the 101 Banner SFG subjects, which allowed us to examine potential associations with diverse disease comorbidities. We did not identify any difference in rates of Type 2 diabetes mellitus, hypertension, obesity, atrial fibrillation, coronary stenosis, cardiomegaly, atherosclerosis, smoking, or chronic obstructive pulmonary disease (**Supplementary Table 9**).

Given that CD83 is a marker of mature dendritic cells, with complex, bidirectional interactions between CD83 and diverse pathogens[36-38], we hypothesized that CD83(+) AD subjects may differ from CD83(-) AD subjects on the basis of a microbial or immunological perturbation which could be occurring peripherally, including during end-stage dementia. Aspiration pneumonia and subsequent sepsis is a common cause of death in subjects that die from AD[39], with autopsy-confirmed pneumonia present in 54% of subjects that died with AD and 49.6% of all subjects within the BBDP[40]. We did not observe associations between autopsy confirmed pneumonia and presence of CD83(+) microglia among AD subjects (P-value: 0.37, Odds ratio: 0.57, Fisher's Exact Test).

We then examined potential clinicopathological associations within the ROSMAP DLPFC snRNA-seq samples, and observed a modest increase in global AD pathology burden (a quantitative summary across five brain

regions of neuritic plaque, diffuse plaque and neurofibrillary tangle density) among AD subjects with CD83(+) microglia (n=68) compared with CD83(-) AD (n=203) subjects (T-statistic: 2.63 FDR: 3.9e-2), which was driven primarily by an increase in neuritic plaque density (T-statistic: 2.68 Pvalue: 3.9e-2). These significant associations were also observed in a complementary linear modelling approach looking across all 424 samples, while adjusting for age of death, sex and AD diagnosis (**Supplementary Table 9**). Subsampling of the ROSMAP DLPFC snRNA-seq cohort to sets of 66 AD subjects resulted in our detection of increased (FDR < 0.05) global AD pathology burden in only 12% of subsamplings and detection of increased neuritic plaques in 10% of subsamplings (Random subsamplings, n=1000). These findings are consistent with an improved power to detect a modest clinicopathological association with CD83(+) microglia in AD subjects within the larger ROSMAP DLPFC snRNA-seq data, rather than necessarily implying a discordance across brain regions or cohorts.

Transcriptional regulatory model of *CD83* in microglia implicates transcription factor *NR4A2*

To better understand the potential transcriptional context that may mediate the transition to CD83(+) microglia, we constructed a transcriptional regulatory model for *CD83* as learned from independent microglial transcriptomic data. We extracted expression levels of *CD83* and all candidate transcription factors from a previously published bulk, sorted microglial RNA-seq dataset[41]. These data were then integrated with putative transcription factor binding sites for *CD83* and Transcriptional Regulatory Network Analysis[42] (Trena) was used to identify which transcription factors best predict CD83 expression. The strongest candidate regulator of *CD83* expression to emerge was steroid-thyroid hormone-retinoid superfamily member Nuclear Receptor Subfamily 4 Group A Member 2 (*NR4A2*). Subsequent investigation of *NR4A2* expression in CD83(+) microglia within our data revealed significant upregulation compared with CD83(-) microglia ($\log_2$FC=1.17, FDR=5.42e-61) and examination of our CD83(+) microglial networks revealed direct connectivity between *CD83* and *NR4A2* (**Supplementary Figure 5**).

To further substantiate our hypothesis that *NR4A2* plays a pivotal role as a transcriptional regulator of *CD83* in microglia, we conducted a systematic assessment of transcription factor (TF) activity within regulatory regions of *CD83*. This analysis was performed using previously published microglial ATAC-seq data[41] and included comparisons with oligodendrocytes, GABAergic neurons, and glutamatergic neurons[43]. Initially, we employed an activity-by-contact approach[44] to predict regulatory regions, followed by the TOBIAS footprinting method[45] to estimate TF binding within those regions. Among the five regulatory regions associated with *CD83*, we identified the presence of bound *NR4A2* motif in the microglia-specific open chromatin region located 22kb upstream of *CD83* (**Figure 5F**). Together, these observations support the hypothesis that *NR4A2* is a significant transcriptional regulator of *CD83* within microglia.

CD83(+) microglia are associated with increased immunoglobulin IgG4 in the transverse colon

We then hypothesized that a potential microbial perturbation could be reflected in the proteome of the transverse colon (TC), the anatomical site with the highest microbial diversity[46]. We generated mass spectrometry proteomics data from frozen TC samples from a subset of 26 subjects with SFG snRNA-seq. We examined whether any proteins are differentially abundant as a function of the presence of CD83(+) microglia within the SFG, while adjusting for AD status, age of death, sex and PMI. The most differentially abundant protein in the transverse colon of subjects with CD83(+) microglia was IGHG4 (Immunoglobulin Heavy Constant Gamma 4, **Figure 4A**, T-statistic:4.5 P-value: 2.1e-4, **Supplementary Table 10**) which forms the constant region of the IgG4 antibody heavy chain. This observation was suggestive of an increased IgG4 tissue response in the TC wall of subjects with CD83(+) microglia, and more broadly, consistent with a potential microbial interaction (whether direct or indirect) between components of the gut microbiome and the presence of CD83(+) microglia.

Discussion

In this study we introduce a publicly available, postmortem multiomic data set generated from 101 subjects enrolled in the BBDP[22]. Through the integration of snRNA-seq from SFG, coupled with WGS data and extensive neuropathological and antemortem characterizations, we report a powerful resource for the exploration of cell specific transcriptomics and gene networks in the context of aging and AD.

We adopted an approach for identifying cell subtypes that are differentially abundant between AD and Aged Control subjects. We observed that this method was associated with an improved power to detect DEGs versus a conventional comparison of AD against Aged Controls, with these genes demonstrating a statistical enrichment for AD GWAS associated genes within microglia, inhibitory neurons and OPC populations.

We found that integration of WGS data with snRNA-seq data yielded substantial novel eQTL associations, including in minor cell fractions such as microglia, and thus which are unlikely to have been detected in bulk transcriptomics. Many eQTL loci were also previously linked with diverse neurological and AD associated traits in published genetic studies, and thus represent opportunities to contextualize complex genetic-phenotypic associations through cell specific gene regulatory relationships. In agreement with Fujita et al[28], we observed that AD GWAS locus rs679515 (as well as rs9429780 and rs11118328) is an eQTL for *CR1* in oligodendrocytes and also with apparent impacts upon hematological molecular networks and peripheral blood hematocrit, RBC count, hemoglobin concentration and anemia diagnosis. Interestingly, these associations appears to occur in a neuropathology-specific manner.

CR1 has been associated with AD via several polymorphisms[47-50] though findings have been variable across datasets[51] and clarifying whether the pathogenic mechanism occurs in the central nervous system (CNS) or periphery remains elusive. *CR1* encodes a transmembrane glycoprotein receptor for multiple ligands (including complement components *C3b* and *C4b*) and is abundantly expressed on erythrocytes and leukocytes[52], though also on tissue-resident macrophages and dendritic cells and on multiple brain cell types, including microglia, astrocytes[53], and oligodendrocytes[54]. Peripherally, C3b-opsonized immune complexes (including amyloid aggregates) bind to *CR1* on erythrocytes and undergo eventual phagocytic clearance in the liver and spleen[52], representing a major hypothesized mechanism that links *CR1* genetic variants with AD risk[55,56]. Despite this notion, plasma *CR1* levels have not been reported to be altered in AD, though are modestly increased in carriers of several *CR1*-linked AD GWAS risk variants[57]. We observed the rs679515 / *CR1* eQTL association only in oligodendrocytes, with network findings that were sufficiently reminiscent of canonical erythrocyte regulators to motivate our investigation of effects upon peripheral hematocrit. Overall, our findings align with a potentially multi-tissue impact of rs679515 / *CR1*, although whether the oligodendrocyte rs679515 / *CR1* network actually reflects a key AD pathomechanism or is just conserved enough as to suggest the relevant peripheral mechanism remains unresolved. Notably, the GWAS loci for *CR1* have also been independently found to be co-localized in oligodendrocytes in AD, by an integrative analysis of multiple snRNA-seq datasets from several different brain regions[29]. We also note recent reports by Yang et al[58] on AD-relevant activity of *CR1* within perivascular macrophages (PVM). Given complex reciprocal interactions between oligodendrocytes and the neurovascular unit[59,60], simultaneous evaluation of *CR1* and AD GWAS loci in PVM and oligodendrocytes may illuminate a shared mechanism that could explain the observed CNS as well as hematological effects in a parsimonious manner. Further investigation of whether rs679515 is associated with alterations of peripheral *CR1* may also represent a fruitful path for future study.

We observed an AD associated, CD83(+) microglial subtype which was present in 47% of neuropathologically defined AD subjects within the Banner SFG snRNA-seq and 25% of the ROSMAP DLPFC snRNA-seq data. Gene regulatory network reconstruction based on these microglia identified a gain in network connectivity of *APOE* and the presence of multiple microglial nodes of interest, including *TREM2* and complement *C1Q* complex genes, with the most influential network drivers comprised of comparatively understudied noncoding RNA gene *AC131944.1* / ENSG00000286618.1, which is encoded antisense to critical microglial regulator *SPP1*[61,62], with a positive regulatory relationship between *AC131944.1* and *SPP1* implied by our CD83(+)

microglial networks. Functional enrichments of CD83(+) microglial genes highlighted biological themes of humoral immunity, complement activity, lipid processing and iron transport. A transcriptional model constructed from independent microglial transcriptomic data revealed *NR4A2* as the most likely transcription factor contributing to *CD83* expression within the CD83(+) microglia. *NR4A2* (also known as *NURR1*) has been well described in dopaminergic neurons, particularly in the context of Parkinson's Disease[63], though is also responsive to LPS activation in cultured microglia[64] and linked with cognitive and neuropathological changes in the 5XFAD mouse model of AD[65,66]. Previous studies have reported CD83(+) microglia as AD-associated[9] and recent work also links *CD83* as a marker of microglia engaged in myelin debris phagocytosis as well as a potential modulator of autoimmune neuroinflammation during certain disease states[67]. These observations are also consistent with reports of microglial and astrocytic *NR4A2* activity restraining neuroinflammation induced cell death[68].

The presence of CD83(+) microglia within the SFG of subjects with AD was not differentially associated with PMI, demographic factors, major disease comorbidities or terminal pneumonia infection, though it was associated with a significant increase in multi-regional neuritic plaque density. We also observed a significantly increased abundance of IGHG4 protein within the transverse colon of these subjects that is suggestive of increased IgG4 immunoglobulin. IgG4 is one of four main IgG immunoglobulin subclasses, and the least prevalent among healthy adults[69] (approximately 5% of IgG) with important structural differences driving functional differences including reduced triggering of antibody-dependent cell-mediated phagocytosis and complement activation[70]. Although IgG4 is generally regarded as exerting a relatively anti-inflammatory effect relative to other IgG subclasses, it is still capable of driving phagocytosis of opsonized microbial antigens[71] and mediating complement activation at higher antigen and antibody concentrations[72]. Despite the emergence of IGHG4 as the most differentially abundant protein, we did also observe a significant upregulation of IGHG3 and IGHG2 (which would be expected to induce more aggressive antipathogen responses), which is overall suggestive of a complex, adaptive immune response present in the transverse colon of subjects with CD83(+) microglia in the SFG. It is intriguing to note that gene set enrichments generated on CD83(+) microglial hub genes were also reflective of biology underpinning humoral immunity.

Our study has several limitations, including some inherent to snRNA-seq, such as the absence of expression for transcripts that are not abundantly present in the nucleus, and thus represents a major limitation in the scope of our transcriptomic findings. We also note that our primary findings are based on SFG samples, which was selected for relevance to AD pathobiology[73], however the extent to which our findings might apply to additional brain regions is not yet clear. Despite this, the shared observations of *CR1* eQTL and hematological associations within oligodendrocytes and CD83(+) microglia within the ROSMAP DLPFC snRNA-seq data is encouraging that these findings may apply to additional regions. In addition, gene regulatory network modelling approaches applied to snRNA-seq data are a rapidly evolving frontier, and best standards and practices are yet to emerge[33].

Our study demonstrates the power of combining multiomic, multi-tissue and systems biology approaches to illuminate novel disease biology. These efforts were enabled by several key aspects of the Banner Sun Health Research Institute's Brain & Body Donation Program, including exceptional tissue quality, detailed antemortem (e.g. hematocrit laboratory values) and postmortem characterizations and the availability of tissue from additional extra-CNS anatomical sites (e.g. transverse colon). An additional advantage of our study is the comparatively large sample size of 101 subjects, which supported well-powered, cell-specific eQTL analyses and the application of subsequent network modelling approaches, coupled with the opportunity to examine independent, complementary snRNA-seq data[28] to replicate and extend key findings. In addition to the specific findings described within this study, we encourage access to the associated transcriptomic, genetic, phenotypic and multiscale network data by interested investigators and hope that these resources can prove useful for the scientific community.

Acknowledgments

The authors are grateful for the generous support of the NOMIS Foundation. BR is supported in part by NIA grants U01AG061835 and R21AG063068. QW is supported in part by NIA grants U01AG061835. BR is grateful for funding support received from The Benter Foundation. BR is grateful for funding support received from the Global Lyme Alliance. BR and DM are grateful for funding support received from the Arizona Alzheimer's Consortium. QW and EMR are grateful for additional funding from NIA grant P30AG072980. CF is supported by NIA grants 5U01AG046139-10 and R01AG062514. The Brain and Body Donation Program has been supported by the National Institute of Neurological Disorders and Stroke (U24 NS072026 National Brain and Tissue Resource for Parkinson's Disease and Related Disorders), the National Institute on Aging (P30AG19610 and P30AG072980, Arizona Alzheimer's Disease Center), the Arizona Department of Health Services (contract 211002, Arizona Alzheimer's Research Center), the Arizona Biomedical Research Commission (contracts 4001, 0011, 05-901 and 1001 to the Arizona Parkinson's Disease Consortium) and the Michael J. Fox Foundation for Parkinson's Research. We are grateful to the Banner Sun Health Research Institute Brain and Body Donation Program of Sun City, Arizona for the provision of human biological materials. BBDP resources can be requested at http://www.brainandbodydonationprogram.org.

ROSMAP is supported by National Institute on Aging grants P30AG10161, P30AG72975, R01AG15819, R01AG17917, U01AG46152, U01AG61356. ROSMAP resources can be requested at https://www.radc.rush.edu.

Authors' contributions

EMR, WL, DM, KVKJ, BR, TGB, GES designed the study. BR led the computational analyses. QW and BR performed the computational analyses. QW and BR wrote the manuscript. DM performed histochemical studies of CD83(+) microglia, generated transverse colon proteomics and provided key interpretations of major findings within the study. TGB and GES advised on subject and sample selection and provided biological samples for evaluation. TGB and GES also performed chart review of hematocrit values within the Banner study. KVKJ, WL, JA, EA, and RR performed snRNA-seq data generation and data analysis. CF generated the transcriptional regulatory model for NR4A2. JB and PR performed the epigenomic landscape analysis of CD83 regulatory regions within microglia. TLK assisted with the generation of proteomics data from the transverse colon. DAB and PLD provided data from ROSMAP and scientific inputs. All authors reviewed and edited the manuscript for important intellectual content.

Declarations of Interests

The authors declare that they have no relevant conflicts.

Data Availability

All the raw and processed WGS, snRNA-seq and meta data are available from synapse with accession syn51753326.

Methods

Brain tissue sample processing

Subjects were all volunteers in the Arizona Study of Aging and Neurodegenerative Disorders (AZSAND), a longitudinal clinicopathological study of aging, cognition and movement in the elderly since 1996 in Sun City, Arizona. Autopsies are performed by the Banner Sun Health Research Institute Brain and Body Donation Program[22] (BBDP; www.brainandbodydonationprogram.org). All subjects sign Institutional Review Board-approved informed consents allowing both clinical assessments during life and several options for brain and/or bodily organ donation after death. Most subjects are clinically characterized with annual standardized test

batteries consisting of general neurological, cognitive and movement disorders components, including the Mini Mental State Examination (MMSE). Subjects for the current study (Table 1; n=101) were chosen by searching the BBDP database for a full spectrum of AD neuropathology, in the absence of other neurodegenerative disease diagnoses.

The complete neuropathological examination was performed using standard AZSAND methods[22]. The neuropathological examination was performed in a standardized manner and consisted of gross and microscopic observations, the latter including assessment of frontal, parietal, temporal and occipital lobes, all major diencephalic nuclei and major subdivisions of the brainstem, cerebellum and spinal cord (the lattermost only for those with whole-body autopsy). Detailed clinical data, postmortem neuropathological data, and demographics of the cohort are described in **Table 1A**.

Chromium 10x snRNA-seq

Superior frontal gyrus tissue (~50mg) was homogenized in 1 ml of Nuclei Lysis buffer [Nuclei EZ Lysis Buffer (Sigma-Aldrich, St. Louis, MO, USA) supplemented with 1x cOmplete™ Protease Inhibitor Cocktail (Sigma-Aldrich, St. Louis, MO, USA) and RNasin Plus (Promega)] 10-15 times using pestle A "loose" followed by pestle B "tight" 10-15 times (DWK Life Sciences, Millville, NJ, USA). Homogenate was passed through a 70 µm 1.5 ml mini strainer (PluriSelect, El Cajon, CA, USA) and centrifuged at 500 rcf for 5 minutes at 4□C. Nuclei pellet was resuspended in an additional 1 ml of Nuclei Lysis buffer and incubated for 5 minutes followed by centrifugation at 500 rcf for 5 minutes at 4□C. 500 ul of 1x wash/ resuspension buffer (1x PBS, BSA 2%, 0.2 U/ul RNasin Plus) is added to the nuclei pellet and incubated for 5 minutes to allow adequate buffer exchange followed by centrifugation at 500 rcf for 5 minutes at 4□C, repeated once more and resuspended in 500 ul of 1x wash/ resuspension buffer. Resuspended nuclei were incubated with 1-2 drops of NucBlue Live ReadyProbes Reagent (ThermoFisher Scientific) and immediately sorted using the DAPI channel on the Sony SH800S (Sony Biotechnology, San Jose, CA, USA) with a 100 µm chip.

Nuclei were sorted for 15,000 events directly into 10x 3' v3.1 RT Reagent Master Mix and immediately processed with the 10x Genomics Chromium Next GEM Single Cell 3' v3.1 (Dual Index) kit (10x Genomics, Pleasanton, CA). To minimize batch effects, each 10x chip contained samples from all disease and control groups. Samples were loaded, cDNA amplified, and library constructed following the manufacturer's protocol. Library quality control (QC) was based on Agilent Tapestation 4200 HS D1000 screentapes (Agilent Technologies, Waldbronn, Germany). Multiplexed library pool was based on HS D1000 and Kapa Library Quantification Kit for Illumina platforms (Kapa BioSystems, Boston, MA) and sequenced at shallow depths on Illumina's iSeq 100 v2 flow cell for 28x10x10x90 cycles for estimated reads per cell. After demultiplexing, libraries were rebalanced based on reads per cell. Normalized pool QC was based on Agilent Tapestation 4200 HS D1000 and Kapa Library Quantification Kit for Illumina platforms and high-depth sequenced on Illumina's NovaSeq 6000 S4 v1.5 flow cell for 28x10x10x90 cycles.

snRNA-seq data preprocessing, filtering, integration, clustering, and cell type annotation

Gene counts were obtained by aligning reads to the hg38 reference genome (GRCh38.p5, GCA_000001405.20) using CellRanger (v6.0.2) (10x Genomics)[74], with all default parameters adding the "--include-introns" flag. All the subsequent preprocessing were accomplished in the standard Seurat (v4.0) workflow[75] on each subject individually, with a set of rigorous quality control measures. Briefly, nuclei were removed with gene counts less than 200, having > 5% mitochondrial, > 5% ribosomal, or > 0.1% hemoglobin counts. Nuclei with either too high or low total UMI counts were also removed, by fitting a sigmoid function to the distribution of $\log_{10}$UMI and setting the cutoff to the ln4 ∗ scale factor away from the inflection point. Nuclei with $\log_{10}$GenesPerUMI < 0.8 were also removed. Finally, doublets were identified by DoubletFinder (v.3.0)[76] and removed. Each Seurat object was then normalized using the SCTransform[77] protocol. We regressed out, during the normalization, the

number of genes, the number of UMIs, the percentage of mitochondrial genes, and cell cycle scores. Principal components (PC) were calculated using the first 3,000 variable genes, and Uniform Manifold Approximation and Projection (UMAP)[78] analysis was performed with the top 30 PCs.

Reference-based integration was employed for integration of the 101 subjects in Seurat, using one male and one female as references. Reciprocal PCA (RPCA) was applied to the first 30 PCs of the 3,000 variable genes to find the anchors before integration by SCTransform. PCA and UMAP analysis were performed on the integrated object. Clustering was done using a resolution of 0.5. We annotated the cell clusters by leveraging the snRNA-seq data published earlier[11] as a reference (syn18485175), by cell type label transfer and projecting query cells onto reference UMAPs.

Single-nucleus expression quantitative trait locus (sn-eQTL) mapping

We ran whole genome sequencing for 103 individuals from the BBDP, 99 of which are overlaps with those in snRNA-seq. Reads were aligned to hg38 (GCA_000001405.15) using the BWA-mem aligner (v0.7.10)[79]. Variants were called using the GATK Best Practices workflow (v4.0.1.2)[80]. Variants with a minor allele frequency < 0.05, with missing call rates > 0.1, or having Hardy-Weinberg equilibrium exact test p-value < 1e-6 were removed from further analysis. Common variants for all the remaining biallelic sites were imputed in Beagle (v5.0)[81] using 1000 Genomes Phase 3 reference genotypes[82].

We followed the recommended workflow as reported recently[83] to map the sn-eQTL in this study. Specifically, for each annotated cell type, normalization was first done in scran (v1.24.0)[84] for all the nuclei within the cluster(s), followed by mean aggregation of the expression profiles across nuclei from each individual. Association tests were performed using linear models in LIMIX (v2.0.3)[85], including age, gender, race, disease status and top 20 principal components as covariates. Cis-eSNPs were reported for those sites located within 250kb of the transcribed region for a gene. The p values were corrected for multiple testing by using the conditional false discovery rate (cFDR) method[86], leveraging the meta-analysis eQTLs from cerebral cortical bulk tissues (syn16984815)[27] to increase power. The meta-analysis eQTLs were lifted up to hg38 by CrossMap (v0.6.3)[87].

Differentially abundant (DA) cell population identification

Cell subpopulations whose abundance differs between the two states (AD and control in our study) were detected by DAseq[24], using the integrated Seurat object. Top 40 PCs of the expression profiles were used for comparison. The values of k used for calculation of score vector with kNN were set 100 to 4,000 with an increment of 500. DA measure value thresholds were set at -0.8 to 0.8. Clustering parameter was set to 0.01. The identified DA clusters were integrated with cell cluster annotation for their respective annotations.

Cluster trait associations

To make an unbiased comparison, we collapsed the nuclei from the same individual in each cluster into one and annotated their phenotypic and neuropathological traits by the metrics from the corresponding individuals. The overrepresentation of individuals in each cluster was assessed as reported[11]. For categorical traits, enrichment was evaluated using the hypergeometric distribution (Fisher's exact test) and FDR correction over all clusters. Enrichment or depletion of quantitative traits was assessed individually by contrasting the average observed value across the individuals of a given cluster with a corresponding null distribution estimated by a scheme of 10,000 permutations. The deviation of the observed value from the random expected distribution was quantified using z-score.

Differential gene expression analysis

Differential expression of genes between conditions was identified by the FinderMarkers function of the Seurat package, using the MAST algorithm[88]. Adjusted p-value is based on Bonferroni correction using all genes in the dataset. For each cell type, two comparisons were made. One is between all the cells from AD subjects and those from control subjects (DX). The other comparison is between all the cells from DA populations and those from non-DA populations (DA). Only genes that are detected in a minimum fraction of 1% cells in either of the two conditions were tested. Lists of DEGs were generated by filtering all genes for absolute $\log_2$fold change > 0.25, adjusted P < 0.05.

MEGENA and RIMBANet gene regulatory network construction

We ran gene network analysis for all the DA cell populations in each annotated cell type, following the recommended workflow[33]. Specifically, we constructed gene co-expression and detected multiscale gene modules using MEGENA[89] and Bayesian networks (BN) using RIMBANet[90]. For comparison, we chose the same number of nuclei in both AD and control cell populations and constructed the comparative networks using the same workflow. The nuclei were chosen based on the shortest Euclidean distances of the top 40 PCs between the expression profiles of DA and target populations. Differential connectivity scores for the nodes in the networks were calculated based on the procedure in the R package dnapath[91] (https://github.com/tgrimes/dnapath) with the p values generated by a scheme of 1,000 permutations.

Construction of rs679515 / CR1 gene regulatory network

We performed causal inference testing[47], to build a causal gene regulatory network focused on *CR1* in post-mortem SFG oligodendrocytes. This approach requires paired gene expression and genotype data for a large number of samples to establish the direction of regulation between *CR1* and its correlated genes. Causal inference testing (CIT) has been well described previously[47]. Briefly, it offers a hypothesis test for whether a molecule (in this case, the expression of *CR1*) is potentially mediating a causal association between a DNA locus (rs679515), and some other quantitative trait (such as the expression of genes correlated with *CR1* and rs679515). Causal relationships can be inferred from a chain of mathematic conditions, requiring that for a given trio of loci (L), a potential causal mediator i.e., *CR1* (G) and a quantitative trait (T), the following conditions must be satisfied to establish that G is a causal mediator of the association between L and T:

(a) L and G are associated
(b) L and T are associated
(c) L is associated with G, given T
(d) L is independent of T, given G

We used the R software package "cit"[47], to perform the causal inference test, calculating a false discovery rate using 1000 test permutations. Trios with a Qvalue < 0.05 were classified as significant, and the associated T genes were considered downstream of *CR1*.

Transcriptional regulatory model of *CD83*

A transcriptional regulatory model of *CD83* was created as follows. Expression levels of *CD83* and all candidate transcription factors came from a previously published bulk, sorted microglial RNA-seq dataset[41] (syn25671134) and bulk, sorted GABAergic, glutamatergic and oligodendrocytes[43] (GEO GSE143666). The putative regulator region of *CD83* was defined using regions identified by GeneHancer[92]. With the regions identified by GeneHancer, FIMO[93] was used to identify all putative transcription factor binding sites. We then generated a list of candidate transcription factors based on the presence of these transcription factor binding sites with the GeneHancer regions. Trena[94], which uses several regression and machine learning approaches to

identify which transcription factors can best predict expression of *CD83*, was used to generate the model. To ascertain the binding status of transcription factors across various brain cell types and in AD case/control subjects, we conducted footprinting analysis with the TOBIAS tool[45]. Building upon the settings employed in our previous study[95], we examined 431 motifs, representing 798 transcription factors (some motifs shared due to high similarity), within the consensus set of open chromatin regions generated from bulk, sorted chromatin accessibility (ATAC-seq) data representing GABAergic neurons, glutamatergic neurons, oligodendrocytes and microglia[41,43]. To contextualize the footprinting findings within the *CD83* regulatory landscape, we computed enhancer-promoter links for *CD83* using the activity-by-contact model[44]. This model integrated data from bulk, sorted chromatin accessibility (ATAC-seq)[41,43] and contact frequency matrices derived from bulk, sorted Hi-C data[41,95].

Transverse colon proteomics

For LC–MS/MS, solubilized proteins were quantified (Thermo Fisher EZQ Protein Quantitation Kit or the Pierce BCA). Proteins were reduced with 50 mM dithiothreitol (Sigma-Aldrich) at 95°C for 10 min and alkylated for 30 min with 40 mM iodoacetamide (Pierce). Proteins were digested using 2.0 µg of MS-grade porcine trypsin (Pierce) and peptides were recovered using S-trap Micro Columns (Protifi) per manufacturer directions. Recovered peptides were dried via speed vac and resuspended in 30 µl of 0.1% formic acid. All data-dependent mass spectra were collected in positive mode using an Orbitrap Fusion Lumos mass spectrometer (Thermo Scientific) coupled with an UltiMate 3000 UHPLC (Thermo Scientific). One µL of the peptide was fractionated using an Easy-Spray LC column (50 cm Å~ 75 µm ID, PepMap C18, 2 µm particles, 100 Å pore size, Thermo Scientific) with an upstream 300 µm Å~ 5 mm trap column. Electrospray potential was set to 1.6 kV and the ion transfer tube temperature to 300°C. The mass spectra were collected using the "Universal" method optimized for peptide analysis provided by Thermo Scientific. Full MS scans (375–1500 m/z range) were acquired in profile mode with the following settings: Orbitrap resolution 120,000 (at 200 m/z), cycle time 3 seconds and mass range "Normal;" RF lens at 30% and the AGC set to "Standard"; maximum ion accumulation set to "Auto;" monoisotopic peak determination (MIPS) at "peptide" and included charge states 2–7; dynamic exclusion at 60 s, mass tolerance 10 ppm, intensity threshold at 5.0e3; MS/MS spectra acquired in a centroid mode using quadrupole isolation at 1.6 (m/z); collision-induced fragmentation (CID) energy at 35%, activation time 10 ms. Spectra were acquired over a 240-min gradient, flow rate 0.250 µl/min as follows: 0–3 min at 2%, 3–75 min at 2%–15%, 75–180 min at 15%–30%, 180–220 min at 30%–35%, 220–225 min at 35%–80% 225–230 at 80% and 230–240 at 80–5%.

Label-free quantification (LFQ) and statistical analysis

Raw spectra were loaded into Proteome Discover 2.4 (Thermo Scientific) and protein abundances were determined using Uniprot (www.uniprot.org) Homo sapiens database (Hsap UP000005640.fasta). Protein abundances were determined using raw files and were searched using the following parameters: Trypsin as an enzyme, maximum missed cleavage site 3, min/max peptide length 6/144, precursor ion (MS1) mass tolerance at 20 ppm, fragment mass tolerance at 0.5 Da, and a minimum of 1 peptide identified. Carbamidomethyl (C) was specified as fixed modification and dynamic modifications set to Acetyl and Met-loss at the N-terminus, and oxidation of Met. A concatenated target/decoy strategy and a false-discovery rate (FDR) set to 1.0% were calculated using Percolator. Accurate mass and retention time of detected ions (features) using the Minora Feature Detector algorithm were then used to determine the area-under-the- curve (AUC) of the selected ion chromatograms of the aligned features across all runs and the relative abundances calculated.

ROSMAP snRNA-seq processing and statistical analysis

The Religious Orders Study and Rush Memory and Aging Project (ROSMAP) are prospective cohort studies of aging and dementia[32]. Participants without known dementia agree to annual clinical evaluation and brain

donation. Both studies were approved by an Institutional Review Board of Rush University Medical Center. All participants signed informed and repository consents, and an Anatomic Gift Act. Pathologic methods and APOE genotyping have been previously reported[96-98]. Clinical, postmortem neuropathological data, and demographics of the cohort profiled by the snRNA-seq study are reported in **Table 1B**. We downloaded all the processed data (aligned UMI count matrices) for ROSMAP snRNA-seq, generated from dorsolateral prefrontal cortex sample, from synapse (syn51123521)[28]. For each processed library batch, a Seurat object was created. Based on the cell annotation file (syn51218314), the object was demultiplexed into the objects for each individual. Eventually all the demultiplexed objects from the same individual were merged into one object in Seurat. We applied the same analysis workflow as above in Seurat for each subject, by filtering out low quality nuclei and doublets, normalization, and clustering.

Integration of all the nuclei was accomplished in multiple steps. We first randomly split the 436 subjects into three batches. For each batch, reference-based integration was employed for integration of all subjects in Seurat, using one male and one female as references. RPCA was applied to the first 30 PCs of the 3,000 variable genes to find the anchors before integration by SCTransform. The resulting three integrated Seurat objects were then integrated together using the atomic sketch integration method[99] by selecting and storing 50,000 representative cells ("atoms") from each dataset in the sampling step. We applied the same workflow as above for dimensionality deduction, clustering, and cell type annotation.

For DA population detection, the NIA-Reagan diagnosis[100] was used to dichotomize the subjects into AD and control. The same workflow was applied to identify the DA population in DAseq[24], except the values of k used for calculation of score vector with kNN were set 100 to 1,000 with an increment of 100 to reduce the computing burden.

CR1 expression in oligodendrocytes was obtained by normalization in scran for all the nuclei within the cluster, followed by mean aggregation of the expression profiles across nuclei from each individual. Genotypes were obtained from WGS profiles (syn11707418). Association tests were performed using linear models in R, including age, gender, race, years of education, disease status and PMI as covariates.

Antemortem blood measurements were obtained from Rush RADC Research Resource Sharing Hub (https://www.radc.rush.edu/requests.htm). We used the last available measurement within three years of death for each individual and performed linear regression of major allele dosage of rs679515 with the traits, including age at measurement, sex, and year of education as covariates. All AD specific effects were modelled by adding an interaction term for each postmortem neuropathological measurement. Semiquantitative measurements (e.g. Braak stage or CERAD score) were treated as quantitative. Quantitative measurements (e.g. amyloid or tangles) were log transformed.

Table 1: Clinical, neuropathological and demographic information for the 101 subjects in this study and 433 subjects from ROSMAP cohort profiled by snRNA-seq. CERAD-NP: CERAD-Neuropsychological (CERAD-NP) test battery; NIA-R: NIA-Reagan diagnosis of AD; NIA-AA: NIA-AA diagnostic guidelines for AD; MCI: mild cognitive impairment.

Banner		AD	Control
Total Subjects		66	35
Expired Age	mean	85.9	83.5
	SD	8.7	8.2
Sex	F	33	17
	M	33	18
Race	white	65	35
	black	1	
PMI	mean	3.5	3.3
	SD	1.8	1
Braak Staging	I		6
	II		11
	III	7	13
	IV	24	5
	V	16	
	VI	19	
Plaque Density	zero		22
	sparse	3	10
	moderate	10	2
	frequent	53	1
CERAD-NP	criteria not met	1	21
	not AD	3	11
	possible	25	3
	definite	37	
NIA-R	not AD		1
	low		1
	intermediate	3	
	high	35	
	criteria not met	28	33
NIA-AA	not AD		21
	low		14
	intermediate	33	
	high	33	
Clinical Diagnosis	No CI	20	28
	MCI	11	7
	dementia	35	
APOE	22		1
	23	9	10
	33	30	20
	24	3	
	34	22	4
	44	2	

ROSMAP		AD	Control
Total Subjects		270	163
Expired Age	mean	90.4	87.0
	SD	6.0	7.5
Sex	F	194	100
	M	76	63
Race	White	270	163
PMI	mean	7.8	7.5
	SD	5.4	4.5
Braak Staging	0		6
	1		24
	2	3	32
	3	61	67
	4	110	34
	5	93	
	6	3	
CERAD	1 (Definite)	130	
	2 (Probable)	139	12
	3 (Possible)	1	42
	4 (No AD)		109
NIA-Reagan	1 (High)	69	
	2 (Intermediate)	201	
	3 (Low)		158
	4 (No AD)		5
Cognitive Diagnosis	1 (No CI)	62	81
	2 (MCI and no other cause of CI)	63	47
	3 (MCI and another cause of CI)	2	3
	4 (AD and no other cause of CI)	125	15
	5 (AD and another cause of CI)	11	10
	6 (Other dementia)	7	7
APOE	22		5
	23	34	25
	33	147	112
	24	6	1
	34	79	19
	44	4	1

Supplementary Data

Supplementary Figure 1: Single cell transcriptomics based on snRNA-seq using 101 human brains. A) Experiment workflow in our study; B) targeted comparison approach to identify cell subpopulations associated with glial activation and neurodegeneration.
Supplementary Figure 2: The relationship between major allele dosage of rs679515 and *CR1* gene expression in oligodendrocytes and antemoterm hematocrit measurement in ROSMAP cohort. A) CR1 expression stratified by diagnosis group. Diagnosis was dichotomized by NIA-Reagan score. B) Hematocrit measurement and the regression line stratified by Braak score.
Supplementary Figure 3: DEGs identified in the DA9 population within microglia. DAM gene signatures are labelled. Only those with absolute $\log_2FC > 0.5$ and adjusted $P < 0.05$ are colored. A) human DAM gene signatures; B) other human microglial gene signatures reported in literature[34].
Supplementary Figure 4: Log2 fold change for DEGs obtained by DA vs non-DA cells in microglia in two datasets (Banner vs ROSMAP) analysed in this study.
Supplementary Figure 5: CD83(+) microglial networks with direct connectivity between CD83 and NR4A2.

Supplementary Table 1: Cell type annotation, cluster labels and DA population distribution. Diagnosis enrichment for each DA cluster.
Supplementary Table 2: DEGs for each cell type by two different comparisons (DA vs non-DA; AD vs Controls). Genes are reported if it's differentially expressed in either comparison.
Supplementary Table 3: eQTL summary statistics in each cell type.
Supplementary Table 4: eQTL analyses for CR1 expression in oligodendrocytes and rs679515 association with hematocrit. chr1:207677194 / CR1 network and molecular functional enrichments of the network.
Supplementary Table 5: Major allele dosage of rs679515 association with CR1 expression in oligodendrocytes and hematological traits in ROSMAP cohort.
Supplementary Table 6: DA9 / CD83(+) microglia signature genes. Only those with absolute $\log_2FC > 0.5$ are reported.
Supplementary Table 7: Gene set enrichment analysis of 308 DA9 / CD83(+) microglia hub genes
Supplementary Table 8: DEGs for the DA vs non-DA microglial cells in ROSMAP snRNA-seq data. Only those with absolute $\log_2FC > 0.5$ are reported.
Supplementary Table 9: Phenotypic associations with CD83(+) microglia
Supplementary Table 10: Differential protein abundance between AD subjects with and without CD83(+) microglia (Transverse colon)

References

1 Sollis, E. *et al.* The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Res* **51**, D977-D985, doi:10.1093/nar/gkac1010 (2023).
2 Bellenguez, C. *et al.* New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nature Genetics* **54**, 412-436, doi:10.1038/s41588-022-01024-z (2022).
3 Vandereyken, K., Sifrim, A., Thienpont, B. & Voet, T. Methods and applications for single-cell and spatial multi-omics. *Nat Rev Genet*, 1-22, doi:10.1038/s41576-023-00580-2 (2023).
4 Zhang, L. *et al.* Single-cell transcriptomic atlas of Alzheimer's disease middle temporal gyrus reveals region, cell type and sex specificity of gene expression with novel genetic risk for MERTK in female. *medRxiv*, doi:10.1101/2023.02.18.23286037 (2023).
5 Anderson, A. G. *et al.* Single nucleus multiomics identifies ZEB1 and MAFB as candidate regulators of Alzheimer's disease-specific cis-regulatory elements. *Cell Genomics*, doi:10.1016/j.xgen.2023.100263 (2023).

6 Yang, A. C. *et al.* A human brain vascular atlas reveals diverse mediators of Alzheimer's risk. *Nature* **603**, 885-892, doi:10.1038/s41586-021-04369-3 (2022).

7 Morabito, S. *et al.* Single-nucleus chromatin accessibility and transcriptomic characterization of Alzheimer's disease. *Nature Genetics* **53**, 1143-1155, doi:10.1038/s41588-021-00894-z (2021).

8 Zhou, Y. *et al.* Human and mouse single-nucleus transcriptomics reveal TREM2-dependent and TREM2-independent cellular responses in Alzheimer's disease. *Nat Med* **26**, 131-142, doi:10.1038/s41591-019-0695-9 (2020).

9 Olah, M. *et al.* Single cell RNA sequencing of human microglia uncovers a subset associated with Alzheimer's disease. *Nat Commun* **11**, 6129, doi:10.1038/s41467-020-19737-2 (2020).

10 Lau, S. F., Cao, H., Fu, A. K. Y. & Ip, N. Y. Single-nucleus transcriptome analysis reveals dysregulation of angiogenic endothelial cells and neuroprotective glia in Alzheimer's disease. *Proc Natl Acad Sci U S A* **117**, 25800-25809, doi:10.1073/pnas.2008762117 (2020).

11 Mathys, H. *et al.* Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332-337, doi:10.1038/s41586-019-1195-2 (2019).

12 Grubman, A. *et al.* A single-cell atlas of entorhinal cortex from individuals with Alzheimer's disease reveals cell-type-specific gene expression regulation. *Nat Neurosci* **22**, 2087-2097, doi:10.1038/s41593-019-0539-4 (2019).

13 Brase, L. *et al.* A landscape of the genetic and cellular heterogeneity in Alzheimer disease. *medRxiv*, 2021.2011.2030.21267072, doi:10.1101/2021.11.30.21267072 (2022).

14 Luquez, T. *et al.* Cell type-specific changes identified by single-cell transcriptomics in Alzheimer's disease. *Genome Med* **14**, 136, doi:10.1186/s13073-022-01136-5 (2022).

15 Hu, J. *et al.* Microglial Piezo1 senses Abeta fibril stiffness to restrict Alzheimer's disease. *Neuron* **111**, 15-29 e18, doi:10.1016/j.neuron.2022.10.021 (2023).

16 Chen, X. *et al.* Microglia-mediated T cell infiltration drives neurodegeneration in tauopathy. *Nature* **615**, 668-677, doi:10.1038/s41586-023-05788-0 (2023).

17 Sala Frigerio, C. *et al.* The Major Risk Factors for Alzheimer's Disease: Age, Sex, and Genes Modulate the Microglia Response to Abeta Plaques. *Cell Rep* **27**, 1293-1306 e1296, doi:10.1016/j.celrep.2019.03.099 (2019).

18 Keren-Shaul, H. *et al.* A Unique Microglia Type Associated with Restricting Development of Alzheimer's Disease. *Cell* **169**, 1276-1290 e1217, doi:10.1016/j.cell.2017.05.018 (2017).

19 Krasemann, S. *et al.* The TREM2-APOE Pathway Drives the Transcriptional Phenotype of Dysfunctional Microglia in Neurodegenerative Diseases. *Immunity* **47**, 566-581 e569, doi:10.1016/j.immuni.2017.08.008 (2017).

20 Efthymiou, A. G. & Goate, A. M. Late onset Alzheimer's disease genetics implicates microglial pathways in disease risk. *Mol Neurodegener* **12**, 43, doi:10.1186/s13024-017-0184-x (2017).

21 Thrupp, N. *et al.* Single-Nucleus RNA-Seq Is Not Suitable for Detection of Microglial Activation Genes in Humans. *Cell Rep* **32**, 108189, doi:10.1016/j.celrep.2020.108189 (2020).

22 Beach, T. G. *et al.* Arizona Study of Aging and Neurodegenerative Disorders and Brain and Body Donation Program. *Neuropathology* **35**, 354-389, doi:10.1111/neup.12189 (2015).

23 Jack, C. R., Jr. *et al.* NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimers Dement* **14**, 535-562, doi:10.1016/j.jalz.2018.02.018 (2018).

24 Zhao, J. *et al.* Detection of differentially abundant cell subpopulations in scRNA-seq data. *Proc Natl Acad Sci U S A* **118**, doi:10.1073/pnas.2100293118 (2021).

25 Bates, K. A., Fonte, J., Robertson, T. A., Martins, R. N. & Harvey, A. R. Chronic gliosis triggers Alzheimer's disease-like processing of amyloid precursor protein. *Neuroscience* **113**, 785-796, doi:10.1016/s0306-4522(02)00230-0 (2002).

26 Price, J. L. *et al.* Neuron number in the entorhinal cortex and CA1 in preclinical Alzheimer disease. *Arch Neurol* **58**, 1395-1402, doi:10.1001/archneur.58.9.1395 (2001).

27 Sieberts, S. K. *et al.* Large eQTL meta-analysis reveals differing patterns between cerebral cortical and cerebellar brain regions. *Sci Data* **7**, 340, doi:10.1038/s41597-020-00642-8 (2020).

28 Fujita, M. *et al.* Cell-subtype specific effects of genetic variation in the aging and Alzheimer cortex. *bioRxiv*, 2022.2011.2007.515446, doi:10.1101/2022.11.07.515446 (2022).

29 Bryois, J. *et al.* Cell-type-specific cis-eQTLs in eight human brain cell types identify novel risk genes for psychiatric and neurological disorders. *Nat Neurosci* **25**, 1104-1112, doi:10.1038/s41593-022-01128-z (2022).

30 Albert, F. W. & Kruglyak, L. The role of regulatory variation in complex traits and disease. *Nat Rev Genet* **16**, 197-212, doi:10.1038/nrg3891 (2015).

31 Marioni, R. E. *et al.* GWAS on family history of Alzheimer's disease. *Transl Psychiatry* **8**, 99, doi:10.1038/s41398-018-0150-6 (2018).

32 Bennett, D. A. *et al.* Religious Orders Study and Rush Memory and Aging Project. *Journal of Alzheimer's Disease* **64**, S161-S189, doi:10.3233/JAD-179939 (2018).

33 Wang, M. *et al.* Guidelines for bioinformatics of single-cell sequencing data analysis in Alzheimer's disease: review, recommendation, implementation and application. *Mol Neurodegener* **17**, 17, doi:10.1186/s13024-022-00517-z (2022).

34 Chen, Y. & Colonna, M. Microglia in Alzheimer's disease at single-cell level. Are there common patterns in humans and mice? *J Exp Med* **218**, doi:10.1084/jem.20202717 (2021).

35 Nguyen, A. T. *et al.* APOE and TREM2 regulate amyloid-responsive microglia in Alzheimer's disease. *Acta Neuropathol* **140**, 477-493, doi:10.1007/s00401-020-02200-3 (2020).

36 Senechal, B., Boruchov, A. M., Reagan, J. L., Hart, D. N. & Young, J. W. Infection of mature monocyte-derived dendritic cells with human cytomegalovirus inhibits stimulation of T-cell proliferation via the release of soluble CD83. *Blood* **103**, 4207-4215, doi:10.1182/blood-2003-12-4350 (2004).

37 Kummer, M. *et al.* Herpes simplex virus type 1 induces CD83 degradation in mature dendritic cells with immediate-early kinetics via the cellular proteasome. *J Virol* **81**, 6326-6338, doi:10.1128/JVI.02327-06 (2007).

38 Majumder, B. *et al.* Human immunodeficiency virus type 1 Vpr impairs dendritic cell maturation and T-cell activation: implications for viral immune escape. *J Virol* **79**, 7990-8003, doi:10.1128/JVI.79.13.7990-8003.2005 (2005).

39 Todd, S., Barr, S. & Passmore, A. P. Cause of death in Alzheimer's disease: a cohort study. *QJM* **106**, 747-753, doi:10.1093/qjmed/hct103 (2013).

40 Beach, T. G. *et al.* Increased Risk of Autopsy-Proven Pneumonia with Sex, Season and Neurodegenerative Disease. *medRxiv*, 2021.2001.2007.21249410, doi:10.1101/2021.01.07.21249410 (2021).

41 Kosoy, R. *et al.* Genetics of the human microglia regulome refines Alzheimer's disease risk loci. *Nat Genet* **54**, 1145-1154, doi:10.1038/s41588-022-01149-1 (2022).

42 trena: Fit transcriptional regulatory networks using gene expression, priors, machine learning. v. R package version 1.21.0 (2022).

43 Hauberg, M. E. *et al.* Common schizophrenia risk variants are enriched in open chromatin regions of human glutamatergic neurons. *Nat Commun* **11**, 5581, doi:10.1038/s41467-020-19319-2 (2020).

44 Fulco, C. P. *et al.* Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. *Nat Genet* **51**, 1664-1669, doi:10.1038/s41588-019-0538-0 (2019).

45 Bentsen, M. *et al.* ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation. *Nat Commun* **11**, 4267, doi:10.1038/s41467-020-18035-1 (2020).

46 Human Microbiome Project, C. Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207-214, doi:10.1038/nature11234 (2012).

47 Corneveaux, J. J. *et al.* Association of CR1, CLU and PICALM with Alzheimer's disease in a cohort of clinically characterized and neuropathologically verified individuals. *Hum Mol Genet* **19**, 3295-3301, doi:10.1093/hmg/ddq221 (2010).

48 Jun, G. *et al.* Meta-analysis confirms CR1, CLU, and PICALM as alzheimer disease risk loci and reveals interactions with APOE genotypes. *Arch Neurol* **67**, 1473-1484, doi:10.1001/archneurol.2010.201 (2010).
49 Carrasquillo, M. M. *et al.* Replication of CLU, CR1, and PICALM associations with alzheimer disease. *Arch Neurol* **67**, 961-964, doi:10.1001/archneurol.2010.147 (2010).
50 Lambert, J. C. *et al.* Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nat Genet* **41**, 1094-1099, doi:10.1038/ng.439 (2009).
51 Van Cauwenberghe, C. *et al.* Complement receptor 1 coding variant p.Ser1610Thr in Alzheimer's disease and related endophenotypes. *Neurobiol Aging* **34**, 2235 e2231-2236, doi:10.1016/j.neurobiolaging.2013.03.008 (2013).
52 Torvell, M. *et al.* Genetic Insights into the Impact of Complement in Alzheimer's Disease. *Genes (Basel)* **12**, doi:10.3390/genes12121990 (2021).
53 Daskoulidou, N. *et al.* Complement receptor 1 is expressed on brain cells and in the human brain. *Glia* **71**, 1522-1535, doi:10.1002/glia.24355 (2023).
54 Brase, L. *et al.* Single-nucleus RNA-sequencing of autosomal dominant Alzheimer disease and risk variant carriers. *Nature Communications* **14**, 2314, doi:10.1038/s41467-023-37437-5 (2023).
55 Johansson, J. U. *et al.* Peripheral complement interactions with amyloid beta peptide in Alzheimer's disease: Polymorphisms, structure, and function of complement receptor 1. *Alzheimers Dement* **14**, 1438-1449, doi:10.1016/j.jalz.2018.04.003 (2018).
56 Rogers, J. *et al.* Peripheral clearance of amyloid beta peptide by complement C3-dependent adherence to erythrocytes. *Neurobiol Aging* **27**, 1733-1739, doi:10.1016/j.neurobiolaging.2005.09.043 (2006).
57 Fonseca, M. I. *et al.* Analysis of the Putative Role of CR1 in Alzheimer's Disease: Genetic Association, Expression and Function. *PLoS One* **11**, e0149792, doi:10.1371/journal.pone.0149792 (2016).
58 Yang, A. C. *et al.* A human brain vascular atlas reveals diverse mediators of Alzheimer's risk. *Nature* **603**, 885-892, doi:10.1038/s41586-021-04369-3 (2022).
59 Pfeiffer, F. Reciprocal Interactions between Oligodendrocyte Precursor Cells and the Neurovascular Unit in Health and Disease. *Cells* **11**, doi:10.3390/cells11121954 (2022).
60 Palhol, J. S. C. *et al.* Direct association with the vascular basement membrane is a frequent feature of myelinating oligodendrocytes in the neocortex. *Fluids Barriers CNS* **20**, 24, doi:10.1186/s12987-023-00425-4 (2023).
61 Qiu, Y. *et al.* Definition of the contribution of an Osteopontin-producing CD11c(+) microglial subset to Alzheimer's disease. *Proc Natl Acad Sci U S A* **120**, e2218915120, doi:10.1073/pnas.2218915120 (2023).
62 Hammond, T. R. *et al.* Single-Cell RNA Sequencing of Microglia throughout the Mouse Lifespan and in the Injured Brain Reveals Complex Cell-State Changes. *Immunity* **50**, 253-271 e256, doi:10.1016/j.immuni.2018.11.004 (2019).
63 Chu, Y. *et al.* Nurr1 in Parkinson's disease and related disorders. *J Comp Neurol* **494**, 495-514, doi:10.1002/cne.20828 (2006).
64 Fan, X. *et al.* Nurr1 expression and its modulation in microglia. *Neuroimmunomodulation* **16**, 162-170, doi:10.1159/000204229 (2009).
65 Moon, M. *et al.* Nurr1 (NR4A2) regulates Alzheimer's disease-related pathogenesis and cognitive function in the 5XFAD mouse model. *Aging Cell* **18**, e12866, doi:10.1111/acel.12866 (2019).
66 Moon, M. *et al.* Correlation between orphan nuclear receptor Nurr1 expression and amyloid deposition in 5XFAD mice, an animal model of Alzheimer's disease. *J Neurochem* **132**, 254-262, doi:10.1111/jnc.12935 (2015).
67 Sinner, P. *et al.* Microglial expression of CD83 governs cellular activation and restrains neuroinflammation in experimental autoimmune encephalomyelitis. *Nat Commun* **14**, 4601, doi:10.1038/s41467-023-40370-2 (2023).
68 Saijo, K. *et al.* A Nurr1/CoREST pathway in microglia and astrocytes protects dopaminergic neurons from inflammation-induced death. *Cell* **137**, 47-59, doi:10.1016/j.cell.2009.01.038 (2009).

69 Aalberse, R. C., Stapel, S. O., Schuurman, J. & Rispens, T. Immunoglobulin G4: an odd antibody. *Clin Exp Allergy* **39**, 469-477, doi:10.1111/j.1365-2222.2009.03207.x (2009).
70 Bruggemann, M. *et al.* Comparison of the effector functions of human immunoglobulins using a matched set of chimeric antibodies. *J Exp Med* **166**, 1351-1361, doi:10.1084/jem.166.5.1351 (1987).
71 Goh, Y. S. *et al.* Human IgG isotypes and activating Fcgamma receptors in the interaction of Salmonella enterica serovar Typhimurium with phagocytic cells. *Immunology* **133**, 74-83, doi:10.1111/j.1365-2567.2011.03411.x (2011).
72 Oskam, N. *et al.* Factors affecting IgG4-mediated complement activation. *Front Immunol* **14**, 1087532, doi:10.3389/fimmu.2023.1087532 (2023).
73 Liang, W. S. *et al.* Altered neuronal gene expression in brain regions differentially affected by Alzheimer's disease: a reference data set. *Physiol Genomics* **33**, 240-256, doi:10.1152/physiolgenomics.00242.2007 (2008).
74 Zheng, G. X. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat Commun* **8**, 14049, doi:10.1038/ncomms14049 (2017).
75 Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587 e3529, doi:10.1016/j.cell.2021.04.048 (2021).
76 McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst* **8**, 329-337 e324, doi:10.1016/j.cels.2019.03.003 (2019).
77 Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol* **20**, 296, doi:10.1186/s13059-019-1874-1 (2019).
78 McInnes, L., Healy, J., Saul, N. & Großberger, L. UMAP: Uniform Manifold Approximation and Projection. *Journal of Open Source Software* **3**, doi:10.21105/joss.00861 (2018).
79 Li, H. (arXiv, 2013).
80 Van der Auwera, G. A. & O'Connor, B. D. *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra*. 1st edition edn, (O'Reilly Media, Inc., 2020).
81 Browning, B. L., Zhou, Y. & Browning, S. R. A One-Penny Imputed Genome from Next-Generation Reference Panels. *Am J Hum Genet* **103**, 338-348, doi:10.1016/j.ajhg.2018.07.015 (2018).
82 Genomes Project, C. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74, doi:10.1038/nature15393 (2015).
83 Cuomo, A. S. E. *et al.* Optimizing expression quantitative trait locus mapping workflows for single-cell studies. *Genome Biol* **22**, 188, doi:10.1186/s13059-021-02407-x (2021).
84 Lun, A. T., McCarthy, D. J. & Marioni, J. C. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Res* **5**, 2122, doi:10.12688/f1000research.9501.2 (2016).
85 Casale, F. P., Rakitsch, B., Lippert, C. & Stegle, O. Efficient set tests for the genetic analysis of correlated traits. *Nat Methods* **12**, 755-758, doi:10.1038/nmeth.3439 (2015).
86 Liley, J. & Wallace, C. Accurate error control in high-dimensional association testing using conditional false discovery rates. *Biom J* **63**, 1096-1130, doi:10.1002/bimj.201900254 (2021).
87 Zhao, H. *et al.* CrossMap: a versatile tool for coordinate conversion between genome assemblies. *Bioinformatics* **30**, 1006-1007, doi:10.1093/bioinformatics/btt730 (2014).
88 Finak, G. *et al.* MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol* **16**, 278, doi:10.1186/s13059-015-0844-5 (2015).
89 Song, W. M. & Zhang, B. Multiscale Embedded Gene Co-expression Network Analysis. *PLoS Comput Biol* **11**, e1004574, doi:10.1371/journal.pcbi.1004574 (2015).
90 Zhu, J. *et al.* Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. *Nat Genet* **40**, 854-861, doi:10.1038/ng.167 (2008).
91 Grimes, T., Potter, S. S. & Datta, S. Integrating gene regulatory pathways into differential network analysis of gene expression data. *Sci Rep* **9**, 5479, doi:10.1038/s41598-019-41918-3 (2019).

92 Fishilevich, S. *et al.* GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database (Oxford)* **2017**, doi:10.1093/database/bax028 (2017).
93 Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017-1018, doi:10.1093/bioinformatics/btr064 (2011).
94 trena: Fit transcriptional regulatory networks using gene expression, priors, machine learning v. R package version 1.21.0 (2022).
95 Bendl, J. *et al.* The three-dimensional landscape of cortical chromatin accessibility in Alzheimer's disease. *Nature Neuroscience*, doi:10.1038/s41593-022-01166-7 (2022).
96 Boyle, P. A. *et al.* Attributable risk of Alzheimer's dementia attributed to age-related neuropathologies. *Ann Neurol* **85**, 114-124, doi:10.1002/ana.25380 (2019).
97 Yu, L. *et al.* Neuropathologic features of TOMM40 '523 variant on late-life cognitive decline. *Alzheimers Dement* **13**, 1380-1388, doi:10.1016/j.jalz.2017.05.002 (2017).
98 Bennett, D. A. *et al.* Apolipoprotein E ε4 allele, AD pathology, and the clinical expression of Alzheimer's disease. *Neurology* **60**, 246-252, doi:10.1212/01.Wnl.0000042478.08543.F7 (2003).
99 Hao, Y. *et al.* Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat Biotechnol*, doi:10.1038/s41587-023-01767-y (2023).
100 Newell, K. L., Hyman, B. T., Growdon, J. H. & Hedley-Whyte, E. T. Application of the National Institute on Aging (NIA)-Reagan Institute Criteria for the Neuropathological Diagnosis of Alzheimer Disease. *Journal of Neuropathology & Experimental Neurology* **58**, 1147-1155, doi:10.1097/00005072-199911000-00004 (1999).