Human brain single nucleus cell type enrichments in neurodegenerative diseases

Authors
Chelsea X. Alvarado[1,2,3], Cory A. Weller[1,2,3], Nicholas Johnson[1,2,3], Hampton L. Leonard[1,2,3], Andrew B. Singleton[1,2], Xylena Reed[1,2], Cornelis Blauewendraat[1,2], Mike Nalls[1,2,3]

[1] Center for Alzheimer's and Related Dementias (CARD), National Institute on Aging and National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD, USA
[2] Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA
[3] Data Tecnica International LLC, Washington, DC, USA

Correspondence
Dr. Mike Nalls
Center for Alzheimer's and Related Dementias
National Institute on Aging, Intramural Research Program at the National Institutes of Health
Building T44
Bethesda, MD 20892, USA
E-Mail: mike@datatecnica.com

Keywords
Neurodegeneration, single cell expression, MAGMA, colocalization, Alzheimer's Disease, Parkinson's Disease, Lewy Body Dementia, Amyotrophic Lateral Sclerosis, Frontotemporal Dementia Lobar Degeneration, Progressive supranuclear palsy

Abstract

Single cell RNA sequencing has opened a window into clarifying the complex underpinnings of disease, particularly in quantifying the relevance of tissue- and cell-type-specific gene expression. To identify the cell types and genes important to therapeutic target development across the neurodegenerative disease spectrum, we leveraged genome-wide association studies, recent single cell sequencing data, and bulk expression studies in a diverse series of brain region tissues. We were able to identify significant immune-related cell types in the brain across three major neurodegenerative diseases: Alzheimer's Disease, Amyotrophic Lateral Sclerosis, and Parkinson's Diseases. Subsequently, we identified the major role of 30 fine-mapped loci implicating seven genes in multiple neurodegenerative diseases and their pathogenesis.

Introduction

Neurodegenerative diseases (NDD) encompass diseases characterized by progressive degeneration of cell types including neurons and glia in the central nervous system and/or peripheral nervous system. NDDs vary in their anatomic vulnerabilities and main affected neuropathologies, resulting in specific cell type dysregulation which may or may not be shared between NDDs.[1,2] The complexity of biological processes and pathways involved in NDD pathogenesis has stymied progress in the understanding of disease and treatment development. While the expression of microglia in one cell type of the brain may be dysregulated, within the same disease state, other cell types may be regulated normally. Technology such as bulk RNA sequencing (RNA-seq) would be unlikely to identify these differences in expression since RNA-seq averages the expression of transcripts from all cells present in a tissue sample.

High throughput single-cell, or single-nucleus, mRNA sequencing (scRNA-seq, or snRNA-seq) technology provides a new window into understanding the functionally complex interactions between cell populations and across tissues. While bulk RNA-seq provides an assessment of expression averaged across a population of sampled cells, snRNA-seq allows for nuanced insight at the cell-type level. Instead of identifying gene set enrichment across an entire tissue, one can identify enrichment of gene expression partitioned by cell type, or changes in the composition of cell types themselves. uncertainty still limits our progress in better understanding the biological underpinnings of NDDs. In this report we leverage population-scale genome-wide association study (GWAS) data in conjunction with snRNA-seq in the brain to gain more specific insights into potential cell-type-specific mechanisms of risk within and across neurodegenerative diseases.

Methods

Single nucleus RNA-seq expression data

Data from adult human single nucleus brain tissues, consisting of the forebrain, midbrain, and hindbrain, was obtained from Siletti et al.[3] The data was obtained by sequencing dissections from three healthy post-mortem donors. The subset of data used in this study consisted of snRNA-seq data. It included each of the 461 super clusters defined by Siletti et al. as being represented by the islands identified through a two-dimensional representation calculated by t-distributed stochastic neighbor embedding (t-SNE).[3] Additionally, we calculated enrichment at the scRNA-seq data at the level of annotated cell type classes, using auto-annotation data provided by Siletti et al. in the supplementary materials of their manuscript.

Genome-wide association study summary statistics

We included data from six NDD GWAS: Amyotrophic lateral sclerosis (ALS) from van Rheenen et al.[4]; Alzheimer's disease (AD) from Bellenguez, et al.[5]; Frontotemporal lobar degeneration (FTLD) from Pottier, et al.[6]; Lewy body dementia (LBD) from Chia, et al.[7]; Parkinson's disease (PD) from Nalls, et al.[8]; and Progressive supranuclear palsy (PSP) from Höglinger, et al.[9] Data for ALS, AD, LBD, and PSP were

obtained from GWAS Catalog (https://www.ebi.ac.uk/gwas) and data for FTLD and PD were obtained directly from the respective authors.

MetaBrain eQTL summary statistics

We included expression quantitative trait loci (eQTL) summary statistics from MetaBrain, a large scale eQTL meta-analysis de Klein et al.[10] For colocalization analysis, we included SNPs with a reported p-value no greater than $1x10^{-4}$ for the QTL component of the study.

Cell Type Enrichment Analyses

We conducted cell type enrichment analyses using the R package MAGMA.[11,12] Cell typing for each adult human brain snRNA-seq data and disease GWAS combination. Multiple testing correction (Bonferroni method) allows for efficient identification of enriched cell types. Required input data for conducting MAGMA analysis includes formatted GWAS summary statistics and a CellTypeDataset object (CTD). Preprocessing of disease specific GWAS summary statistics and the creation of the CTD object, which holds cell specificity data, was performed using the R packages *MungeSumstats*[13] and *ECWE*[14] respectively. Quality control and munging of input data was conducted using methods available through the used R packages (see supplementary methods).

Colocalization

We conducted Bayesian colocalization analysis using the R package *coloc*[15] for all pairwise combinations of putatively significant autosomal NDD GWAS ($p \leq 5x10^{-8}$) and MetaBrain eQTLs ($p \leq 1x10^{-4}$). To account for the possibility of a single SNP influencing expression at multiple loci, we iterated over significant probes (see supplementary methods). We report associations with a posterior probability of at least 90%. A summary of data used and numbers of hits is included in **Supplementary Table S1**.

Gene Expression Summarization

We summarized expression ranks for genes of interest within the single cell adult human brain transcriptome data set adult_human_20221007.loom from Siletti et al.[3] Using custom R scripts we converted feature counts into TPM (transcripts per million). For a given sample, feature counts were divided by maximum nonredundant intron-removed exon lengths to correct for differences in gene length. Values were then multiplied by a sample-specific constant (10^6 / T, where T is the sum of length-normalized counts) such that the resulting unitless vector sums to one million. We extracted exon lengths based on annotations from the GTF file used to originally annotate the single cell data (gb_pri_annot.gtf). We calculated the expression percentile rank for genes of interest using the empirical cumulative distribution function and then calculated the mean and median expression percentile rank (EPR) value for each gene for each tested cell type. In order to ease interpretation, we binned the EPR values into 3 classes: off (EPR < 10), low (10 < EPR < 90) or high (EPR > 90).

Online methods

Additional methods details can be found in the Supplementary Materials section as well as in the Data and Code Availability section. All analyses comprising this workflow are summarized in **Figure 1**.

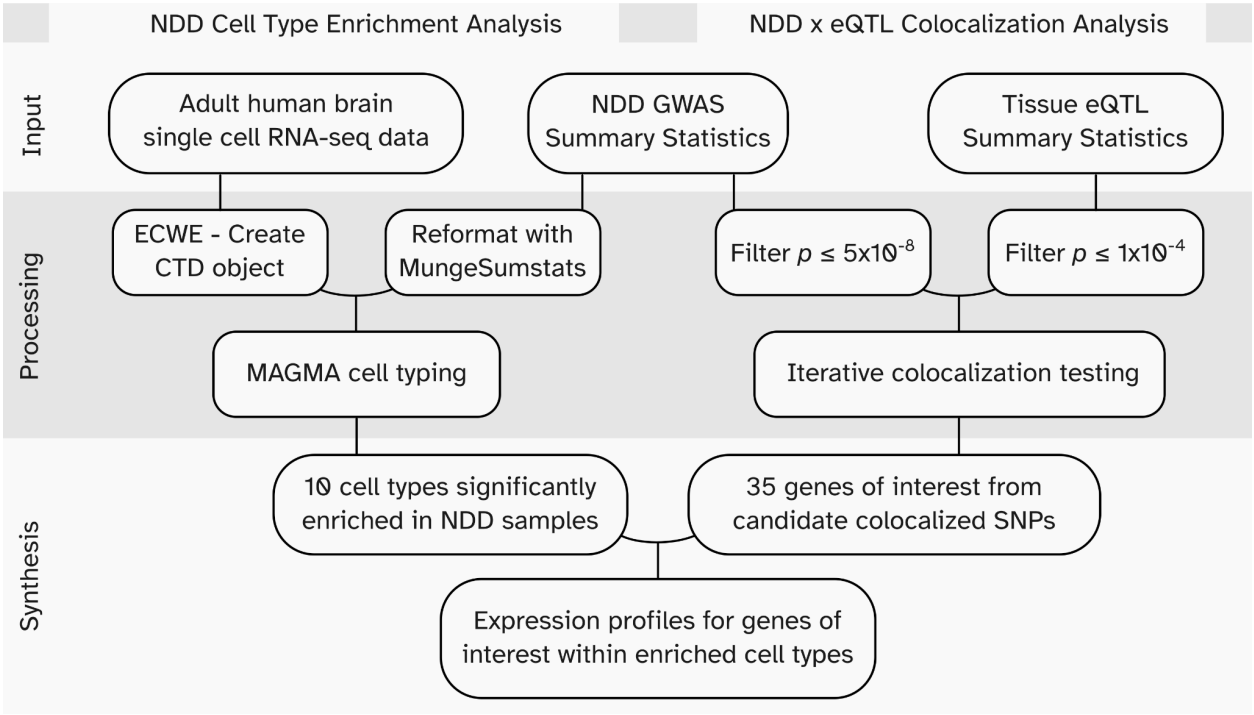

Figure 1: Workflow diagram. **NDD**: Neurodegenerative Disease. **GWAS**: Genome-Wide Association Study. **eQTL**: expression Quantitative Trait Loci. **ECWE**:Expression Weighted Cell Type Enrichment.[14]

Results

Cell type enrichments identified in AD, ALS, and PD

We identified significant cell type enrichments in the adult human brain for three out of six tested diseases: AD, ALS, and PD (**Table 1; Supplementary Table 2**). In general, we found that linear regression style enrichment analysis identified more significant enriched cell types than the top 10% enrichment style ($n_{linear} = 30$, $n_{Top\ 10\%} = 5$). MAGMA. Celltyping documentation does state that using the linear regression enrichment mode results in more significant results due to overlapping cell type signatures.

Disease	Celltype	Genes	Beta	SE	P	Annotation Level	Enrichment Mode
AD	MAC	16170	3.83E-03	0.00068234	9.82E-09	Class	Linear
AD	Microglia	16170	3.98E-03	0.000742	4.19E-08	Supercluster	Linear
AD	MONO	16170	3.36E-03	0.00064412	9.25E-08	Class	Linear
AD	Microglia	1585	1.28E-01	0.027081	1.19E-06	Supercluster	Top 10%
PD	MONO	16178	0.0031777	0.00067611	1.31E-06	Class	Linear
AD	MGL	16170	3.56E-03	0.00075854	1.35E-06	Class	Linear
PD	TCELL	16178	0.0032981	0.00071908	2.27E-06	Class	Linear
PD	CHRP ENDO	16178	0.0028723	0.00065541	5.92E-06	Class	Linear
PD	MAC	16178	0.0031491	0.00071891	5.97E-06	Class	Linear
PD	NK	16178	0.002998	0.00069001	7.02E-06	Class	Linear
PD	Thalamic excitatory	16178	0.0032777	0.00078414	1.47E-05	Supercluster	Linear
PD	PER	16178	0.0030213	0.00072459	1.54E-05	Class	Linear
PD	CGE interneuron	16178	0.0032385	0.00078432	1.83E-05	Supercluster	Linear
PD	LAMP5 LHX6 and Chandelier	16178	0.0032582	0.00079491	2.09E-05	Supercluster	Linear
PD	Midbrain derived inhibitory	16178	0.0032348	0.00079837	2.56E-05	Supercluster	Linear
PD	Eccentric medium spiny neuron	16178	0.0031299	0.00077946	2.98E-05	Supercluster	Linear
PD	VSMC	16178	0.0028347	0.00071616	3.80E-05	Class	Linear
PD	OLIGO	16178	0.0033272	0.00084917	4.48E-05	Class	Linear
AD	NK	16170	2.51E-03	0.00065687	6.65E-05	Class	Linear
PD	Hippocampal dentate gyrus	16178	0.0029692	0.00078409	7.66E-05	Supercluster	Linear
PD	OPC	16178	0.0031794	0.00085042	9.29E-05	Class	Linear
PD	Upper rhombic lip	16178	0.0030704	0.00082409	9.78E-05	Supercluster	Linear
AD	MGL	1080	1.16E-01	0.031771	1.27E-04	Class	Top 10%
AD	NK	856	1.30E-01	0.035634	1.35E-04	Class	Top 10%
AD	MAC	1073	1.16E-01	0.031778	1.37E-04	Class	Top 10%
PD	ENDO	16178	0.002692	0.00074809	1.61E-04	Class	Linear
PD	Hippocampal CA1 3	16178	0.0027432	0.00080036	3.06E-04	Supercluster	Linear
PD	MGE interneuron	16178	0.0026543	0.00078158	3.43E-04	Supercluster	Linear
PD	Deep layer intratelencephalic	16178	0.0026773	0.00079243	3.65E-04	Supercluster	Linear

PD	Oligodendrocyte precursor	16178	0.0028421	0.00084173	3.68E-04	Supercluster	Linear
PD	Mammillary body	16178	0.0027659	0.00082975	4.30E-04	Supercluster	Linear
PD	Upper layer intratelencephalic	16178	0.0026419	0.00079874	4.72E-04	Supercluster	Linear
PD	Committed oligodendrocyte precursor	16178	0.0024936	0.00077006	6.03E-04	Supercluster	Linear
AD	MONO	972	1.05E-01	0.03294	7.06E-04	Class	Top 10%
PD	Bergmann glia	16178	0.0024558	0.00078146	8.39E-04	Supercluster	Linear
ALS	MONO	16094	1.78E-03	0.00056656	0.00085158	Class	Linear

Table 1: Significant Celltype enrichments across AD, ALS, and PD. Table provides information on each significantly enriched cell type after Bonferonni correction ($p < 9.2\times10^{-4}$). Results are provided for supercluster and class levels as well as differing enrichment modes. **CHRP ENDO**: Choroid Plexus and Endothelial. **ENDO**: Endothelial. **MAC**:Macrophage. **MGL**: Microglia. **MONO**: Monocyte. **NK**: Natural Killer cell. **OLIGO**: Oligodendrocyte. **OPC**: Oligodendrocyte precursor cell. **PER**: Pericyte. **TCELL**: T cell. **VSMC**: Vascular smooth muscle cell.

In PD, we ran MAGMA cell typing analysis on three variations/subsets of Nalls et al. meta-GWAS[8]: the complete meta-GWAS; meta-GWAS excluding 23andMe data; and meta-GWAS excluding 23andMe and UK Biobank data. In the first two variations tested, no cell types at the supercluster or class level reached significant enrichment after Bonferroni correction ($p < 9.2\times10^{-4}$). The only GWAS variation that identified significant cell type enrichments was when the full PD meta-GWAS was used in the cell typing analysis. We identified 14 enriched cell types when using linear regression analysis at the supercluster level. At the class level, we identified 10 enriched cell types (**Supplementary Table S3**). The top enriched cell type at the supercluster and class level are *thalamic excitatory* and *monocytes* ($p_{\text{thalmic excitatory}} = 1.467\times10^{-5}$, $\beta_{\text{thalmic\_excitatory}} = 3.278\times10^{-3}$, $p_{\text{monocytes}} = 1.314\times10^{-6}$, $\beta_{\text{monocytes}} = 3.178\times10^{-3}$), with monocytes being the most significantly enriched cell type in PD. Genes enriched in the monocyte cell type include *FCN1, CLEC12A, S100A4, TNFRSF1B, IFI30, LYZ, CYTIP, FGR, LILRB2, KYNU (***Supplementary Tables S4-5**).

In AD, we found microglia to be the only significantly enriched cell type at the supercluster level in both linear and top 10% analyses ($p < 9.2\times10^{-4}$; $p_{\text{linear}} = 4.193\times10^{-8}$, $\beta_{\text{linear}} = 3.978\times10^{-3}$, β $p_{\text{Top 10\%}} = 1.195\times10^{-6}$, $\beta_{\text{Top 10\%}} = 0.1278$) for risk loci. At the class level, four cell types were found to be significant after Bonferroni correction and in both analyses. Significantly enriched cell types identified include macrophages, monocytes, microglia, and natural killer cells (**Supplementary Table S6**). AD was the only disease to have significant enrichments when using the *Top 10%* enrichment analyses.

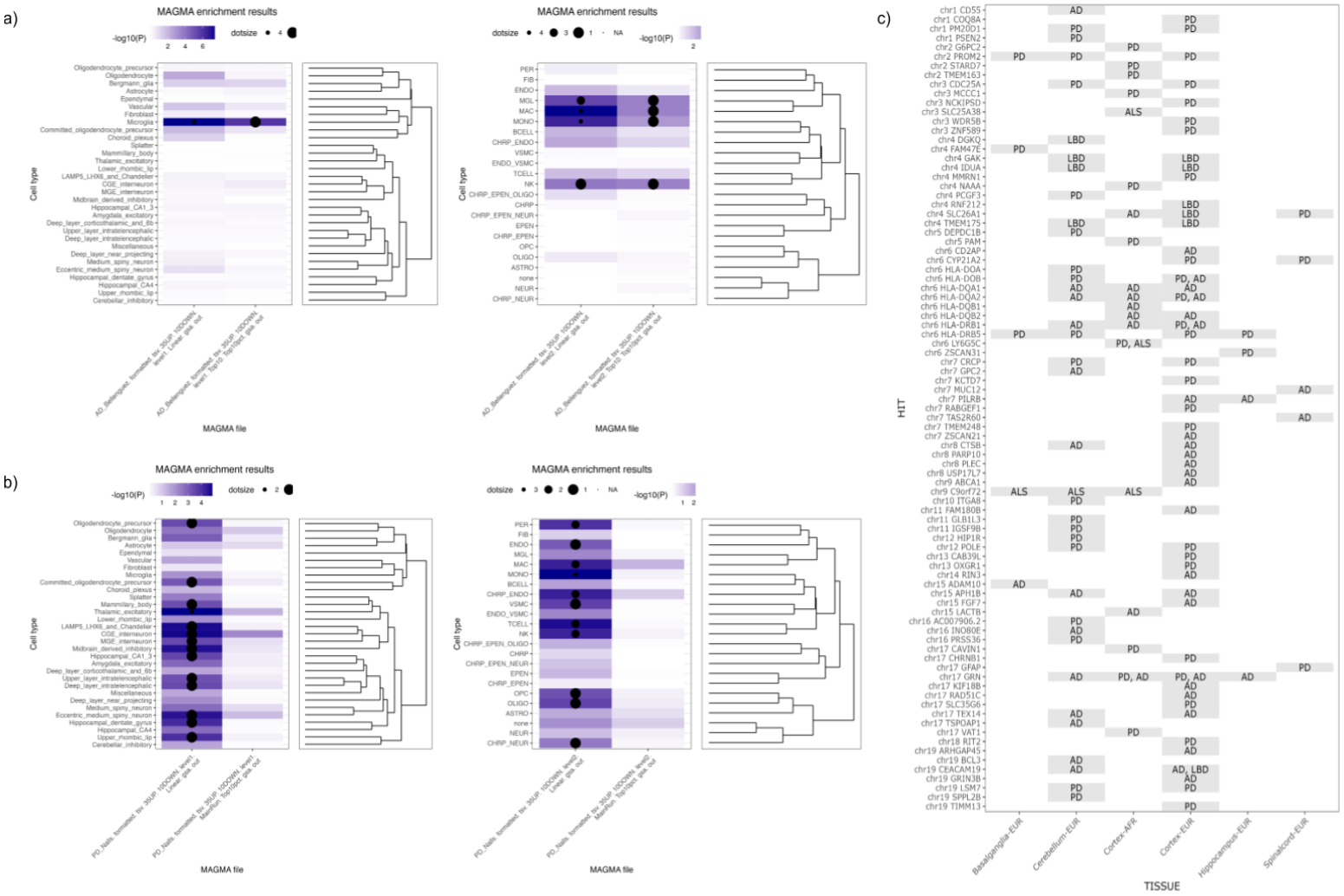

Figure 2: Graphical summaries of cell type enrichment and colocalization results. A. Tile plot comparison of significant cell type enrichments, by disease at each level for both MAGMA.Celltyping analysis. A Bonferonni significance line is provided on the bar chart portion of the image. **B.** Comparison of significant cell type enrichments, with celltype dendrogram, by disease at the class level for both MAGMA.Celltyping analysis. **C.** Summary of fine-mapped loci per disease. AD = Alzheimer's Disease, ALS= Amyotrophic Lateral Sclerosis, PD = Parkinson's Disease.

In ALS, we identified one significant cell type enrichment using linear regression enrichment analysis and at the class level annotations. *Monocytes* were the only cell type to reach significance (Bonferroni-corrected threshold of $p < 9.2\times10^{-4}$); $p_{Monocytes} = 8.516\times10^{-4}$, $\beta_{monocytes} = 1.778\times10^{-3}$; **Supplementary Table S7**). We did not detect significant cell type enrichments in FTLD, LBD, and PSP. No cell types at either the supercluster or class level reached significance after MAGMA-implemented Bonferroni correction ($p < 9.2\times10^{-4}$; **Supplementary Table S2, Supplementary Tables S8-10**).

Colocalization

Across all diseases tested we fine mapped a total of 205 association signals at posterior probability > 90%. This included 89 unique genes identified as harboring putative causal associations. Of these 205 associations, 20 were centered around the *HLA* region, with colocalized omic associations in the cerebellum and cortex (the latter in multiple ancestry groups) suggesting extremely complicated risk of neuroinflammation in this part of the genome. Interestingly, *GRN* is fine-mapped using colocalized expression QTL (eQTL) signals in the cerebellum for AD and the cortex for PD suggesting related but potentially different mechanisms. The *TMEM175*/*GAK* region shows multiple colocalized signals across multiple diseases (PD and LBD). Of particular interest is that *GAK* and *TMEM175* both are fine-mapped to the same SNP (rs6599388) with eQTL effects in opposite directions in LBD. At the same time, regional signals for the gene *SLC26A1* were also fine-mapped for PD, LBD and AD with the effect in PD being detected in the spinal cord while the other disease QTLs were localized to the cortex.

Variants exhibiting a colocalization posterior probability ≥ 90% are summarized in **Figure 2** and **Table 2.** Extended data is available in **Supplementary Table S11**. 30 loci were fine-mapped to a single gene per disease by leveraging QTL data (see **Table 2** for details[10,16]). Of these, six are known druggable genes including *ABCA1*, *ADAM10*, *CD55*, *FGF7*, *OXGR1* and *POLE* (see **Supplementary Table S12**). Mining additional data on these druggable genes from the omicSynth[16] database (**Supplementary Table S13**), we note that *ABCA1* has putative functional multi-omic associations with AD and PSP in blood and brain tissues. *ADAM10* has a similar pattern of functional inferences in AD. *CD55* is shown to have multiple significant functional inferences in brain tissues mediating risk for AD and LBD. However, *FGF7* does not display any significant functional inferences in any diseases from the database query. Methylation QTLs in blood connect PSP and FTD at *OXGR1* via functional inferences using SMR, while multiple brain, blood and nerve associations connect this gene with PD risk across both expression and methylation QTLs. A similar pattern of disease and tissue associations is seen for *POLE*, although there is no significant neural tissue association for PD and PSP is also connected via blood eQTLs to the same gene. Of the fine mapped loci, 22 (Supplemental Table S16) have also been nominated elsewhere as potential therapeutic targets with likely functional impacts on neurodegenerative disease risk in the context of methylation, expression, protein or chromatin QTLs detailed in the omicSynth web application.[16]

CHR	BP	NDD	Tissue-Ancestry	GENEID	Posterior Probability (PP, H4)	QTL Beta	Probe	Probe BP	SNP ID	QTL effect allele	QTL reference allele	QTL effect allele freuqency	Annotation
4	965,703	AD_Bellenguez	Cortex-AFR	SLC26A1	1.000	0.364	ENSG00000145217.14	993440	4:965703:rs4690325:A_G	A	G	0.156	Multiple diseases per gene
4	945,299	LBD_Chia	Cortex-EUR	SLC26A1	1.000	-0.178	ENSG00000145217.14	993440	4:945299:rs6599388:C_T	C	T	0.688	
4	755,832	PD_Nalls	Spinalcord-EUR	SLC26A1	1.000	1.349	ENSG00000145217.14	993440	4:755832:rs77937203:T_G	T	G	0.963	
6	32,636,021	AD_Bellenguez	Cortex-EUR	HLA-DOB	1.000	-0.415	ENSG00000241106.8	32820466	6:32636021:rs1391371:A_T	A	T	0.815	
6	32,973,729	PD_Nalls	Cerebellum-EUR	HLA-DOB	1.000	-0.737	ENSG00000241106.8	32820466	6:32973729:rs116485583:C_T	C	T	0.967	
6	32,636,021	AD_Bellenguez	Cortex-AFR	HLA-DQA2	1.000	-0.951	ENSG00000237541.4	32741391	6:32636021:rs1391371:A_T	A	T	0.868	
6	32,973,729	PD_Nalls	Cortex-EUR	HLA-DQA2	1.000	-0.358	ENSG00000237541.4	32741391	6:32973729:rs116485583:C_T	C	T	0.972	
6	32,636,021	AD_Bellenguez	Cortex-AFR	HLA-DRB1	1.000	0.662	ENSG00000196126.11	32589848	6:32636021:rs1391371:A_T	A	T	0.868	
6	31,636,267	PD_Nalls	Cortex-EUR	HLA-DRB1	1.000	0.765	ENSG00000196126.11	32589848	6:31636267:rs3132453:T_G	T	G	0.066	
6	31,638,615	ALS_vanRheenen	Cortex-AFR	LY6G5C	1.000	-0.788	ENSG00000204428.12	31684040	6:31638615:rs2077492:G_A	G	A	0.656	
6	31,622,577	PD_Nalls	Cortex-AFR	LY6G5C	1.000	0.554	ENSG00000204428.12	31684040	6:31622577:rs3130068:G_A	G	A	0.088	
17	44,352,876	AD_Bellenguez	Cerebellum-EUR	GRN	1.000	0.348	ENSG00000030582.18	44345262	17:44352876:rs5848:C_T	C	T	0.698	
17	44,352,876	PD_Nalls	Cortex-EUR	GRN	1.000	0.329	ENSG00000030582.18	44345262	17:44352876:rs5848:C_T	C	T	0.691	
19	44,544,863	AD_Bellenguez	Cerebellum-EUR	CEACAM19	1.000	-0.579	ENSG00000186567.13	44666958	19:44544863:rs75161053:G_A	G	A	0.943	
19	44,820,881	LBD_Chia	Cortex-EUR	CEACAM19	1.000	-0.144	ENSG00000186567.13	44666958	19:44820881:rs28399637:G_A	G	A	0.699	
1	205,770,138	PD_Nalls	Cortex-EUR	PM20D1	1.000	-0.541	ENSG00000162877.13	205850132	1:205770138:rs708723:C_T	C	T	0.442	Unique gene disease pair per locus
1	207,230,999	AD_Bellenguez	Cerebellum-EUR	CD55	0.997	0.347	ENSG00000196352.15	207321376	1:207230999:rs57003024:T_C	T	C	0.775	
1	226,896,959	PD_Nalls	Cortex-EUR	COQ8A	1.000	0.124	ENSG00000163050.17	226897536	1:226896959:rs2802268:T_G	T	G	0.771	
2	134,680,901	PD_Nalls	Cortex-AFR	TMEM163	1.000	0.337	ENSG00000152128.13	134719000	2:134680901:rs16830908:G_A	G	A	0.615	
2	168,228,349	PD_Nalls	Cortex-AFR	G6PC2	1.000	-0.301	ENSG00000152254.11	168901291	2:168228349:rs10176669:T_C	T	C	0.366	
3	39,451,748	ALS_vanRheenen	Cortex-AFR	SLC25A38	1.000	0.244	ENSG00000144659.13	39383370	3:39451748:rs6765697:C_T	C	T	0.583	
3	122,478,045	PD_Nalls	Cortex-EUR	WDR5B	0.983	-0.228	ENSG00000196981.4	122416062	3:122478045:rs55961674:C_T	C	T	0.829	
3	183,007,683	PD_Nalls	Cortex-AFR	MCCC1	1.000	-0.286	ENSG00000078070.13	183116075	3:183007683:rs11718171:A_G	A	G	0.681	
4	965,703	AD_Bellenguez	Cortex-AFR	SLC26A1	1.000	0.364	ENSG00000145217.14	993440	4:965703:rs4690325:A_G	A	G	0.156	
4	89,378,220	PD_Nalls	Cortex-EUR	MMRN1	1.000	-0.542	ENSG00000138722.10	89879532	4:89378220:rs115793806:A_C	A	C	0.972	
5	60,951,988	PD_Nalls	Cerebellum-EUR	DEPDC1B	1.000	-0.538	ENSG00000035499.13	60689050	5:60951988:rs7722373:G_A	G	A	0.540	
5	102,990,174	PD_Nalls	Cortex-AFR	PAM	1.000	-0.273	ENSG00000145730.20	102753981	5:102990174:rs12656131:A_G	A	G	0.699	
6	27,820,284	PD_Nalls	Hippocampus-EUR	ZSCAN31	0.918	-0.498	ENSG00000235109.7	28356271	6:27820284:rs200502:C_T	C	T	0.759	
6	31,638,615	ALS_vanRheenen	Cortex-AFR	LY6G5C	1.000	-0.788	ENSG00000204428.12	31684040	6:31638615:rs2077492:G_A	G	A	0.656	
6	47,627,419	AD_Bellenguez	Cortex-EUR	CD2AP	0.967	0.171	ENSG00000198087.7	47477789	6:47627419:rs1385742:A_T	A	T	0.351	
7	143,432,985	AD_Bellenguez	Spinalcord-EUR	TAS2R60	1.000	0.550	ENSG00000185899.1	143443453	7:143432985:rs9640384:T_C	T	C	0.644	
9	27,574,517	ALS_vanRheenen	Cortex-AFR	C9orf72	1.000	-0.351	ENSG00000147894.16	27573866	9:27574517:rs11789520:C_T	C	T	0.800	
9	104,903,697	AD_Bellenguez	Cortex-EUR	ABCA1	1.000	-0.291	ENSG00000165029.16	104928155	9:104903697:rs1800978:C_G	C	G	0.873	
10	15,522,754	PD_Nalls	Cerebellum-EUR	ITGA8	1.000	0.507	ENSG00000077943.8	15719922	10:15522754:rs7069969:G_A	G	A	0.122	
11	47,358,789	AD_Bellenguez	Cortex-EUR	FAM180B	1.000	0.319	ENSG00000196666.5	47586693	11:47358789:rs3740688:G_T	G	T	0.453	
12	132,487,182	PD_Nalls	Cortex-EUR	POLE	0.999	0.162	ENSG00000177084.17	132687376	12:132487182:rs11610045:G_A	G	A	0.521	
13	49,365,767	PD_Nalls	Cortex-EUR	CAB39L	1.000	0.558	ENSG00000102547.19	49444064	13:49365767:rs9535211:C_T	C	T	0.760	
13	97,212,767	PD_Nalls	Cortex-EUR	OXGR1	0.972	0.287	ENSG00000165621.8	96994730	13:97212767:rs4771268:T_C	T	C	0.230	
14	92,472,511	AD_Bellenguez	Cortex-EUR	RIN3	0.998	-0.205	ENSG00000100599.16	92513781	14:92472511:rs12590654:G_A	G	A	0.655	
15	49,901,356	AD_Bellenguez	Cortex-EUR	FGF7	1.000	0.133	ENSG00000140285.10	49423178	15:49901356:rs2009833:G_A	G	A	0.640	
15	58,680,796	AD_Bellenguez	Basalganglia-EUR	ADAM10	1.000	0.408	ENSG00000137845.15	58749791	15:58680796:rs11854073:G_A	G	A	0.684	
16	29,971,827	AD_Bellenguez	Cerebellum-EUR	INO80E	0.985	0.714	ENSG00000169592.15	29996064	16:29971827:rs9925102:T_C	T	C	0.542	
16	31,142,825	PD_Nalls	Cerebellum-EUR	PRSS36	0.994	0.812	ENSG00000178226.11	31150066	16:31142825:rs1549299:G_A	G	A	0.267	
16	52,922,415	PD_Nalls	Cerebellum-EUR	AC007906.2	1.000	-0.436	ENSG00000277639.2	53052873	16:52922415:rs113487377:A_G	A	G	0.912	
18	43,185,271	PD_Nalls	Cortex-EUR	RIT2	0.916	-0.199	ENSG00000152214.14	43115691	18:43185271:rs1076971:C_A	C	A	0.847	

Table 2: Summary of colocalized signals. This summary details all associations suggesting a single gene associated with multiple diseases of interest or a single gene per locus at posterior probability above 90%.

Cell type resolution of colocalized genes

We further evaluated gene expression from snRNA-seq used in enrichment analysis. We calculated the mean and median expression percentile rank (EPR) for each gene across cells corresponding to the nominated supercluster cell types identified in our enrichment analyses (**Table 1**) and compared the aggregate mean and median values against the nominated colocalized genes (**Supplementary Figure S2, Supplementary Tables S15-16**). In order to ease interpretation we binned the mean EPR values into three categories based on the mean EPR value for each gene-cell type combination: *off*, *low*, and *high* (see methods).

We identified 10 colocalized genes with *high* median EPR values out of the 14 tested genes. The gene *PAM* has 6 cell type combinations (CGE interneuron, Eccentric medium spiny neuron, MGE interneuron, Hippocampal dentate gyrus, Thalamic excitatory, and Mammillary body) classified as *high* (**Figure 3**). Overall, the Mamillary body cell type had the greatest count of eight *high* EPR genes.

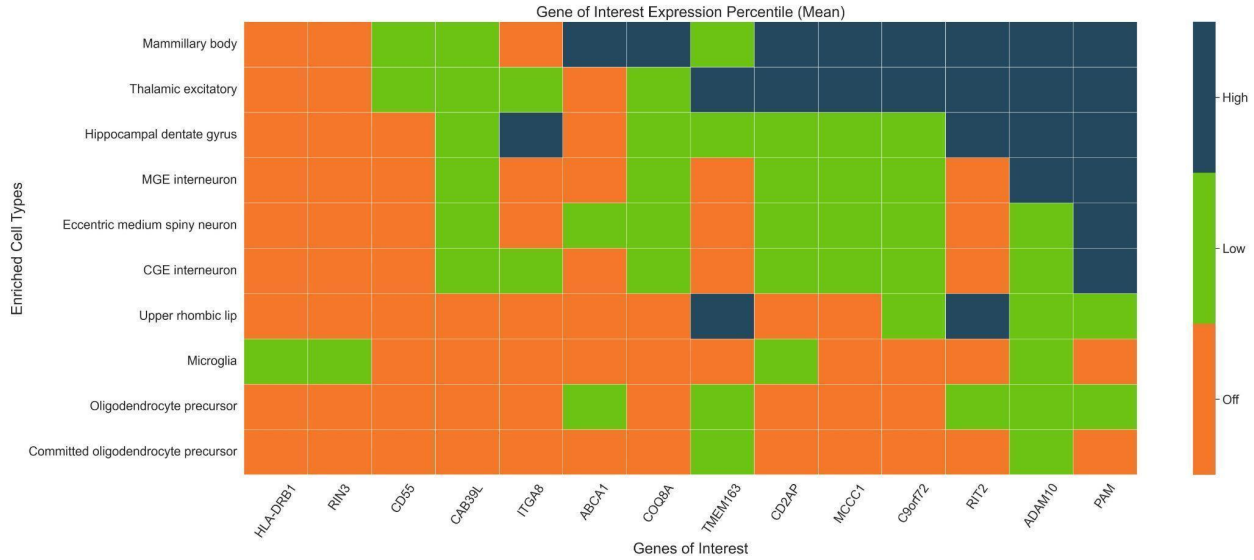

Figure 3: scRNA-seq expression distribution across colocalized genes. This graphical representation visualizes the scRNA-seq expression distribution across MAGMA nominated enriched supercluster cell types. Expression percentile rank (EPR) for each cell-gene combination was binned into *high expression* (dark blue, median expression within top 10% of all genes), *low expression* (green, middle 80%) or *off* (orange, bottom 10%).

Discussion

We identified enriched cell types in various brain regions for three (AD, ALS, and PD) out of six tested NDDs utilizing snRNA-seq data of the 461 superclusters identified by Siletti et al.[3] We did not detect significant cell type enrichments, at either the supercluster or class level, in FTLD, LBD, and PSP. We speculate that due to the smaller sample sizes of each GWAS we were unable to identify any significantly enriched cell types after applying the MAGMA-implemented Bonferroni correction ($p < 9.2\times10^{-4}$; **Supplementary Table S2, Supplementary Tables S8-10**).

All significantly enriched cell types had an associated positive beta for risk genes, and our identified significant cell type enrichments fall in line with current knowledge on cell types implicated in various NDD pathologies which we will discuss further. Broadly, our results highlight the importance of immune-related cell types in the pathology of varying NDDs. Additionally, our results add to the growing evidence that NDDs, such as AD, ALS, and PD, are highly related to autoimmune diseases or may even be autoimmune diseases of the brain.[17–20]

Monocytes were the only cell type significantly enriched across the three diseases AD, ALS, and PD. It is the most significantly enriched cell type in ALS and PD ($p_{ALS}=8.51\times10^{-4}$, $p_{PD} = 1.31\times10^{-6}$) and the third most enriched cell in AD ($p_{AD} = 9.25\times10^{-8}$). Monocytes are the precursor cells to dendritic cells (DCs) and macrophages which both play important roles in immune response, neuroinflammation, and neuroimmunological response.[21–23] Of the genes enriched in the monocyte cell type (previously described in our results), *KYNU* is a gene of note, being part of the tryptophan metabolic pathway which has been found to play a role in Aβ formation.[19]

Microglia are another significantly enriched cell type identified in our analyses, though only significantly enriched in AD. Previous literature highlights the role that microglia play in neuronal loss and neuroinflammation as well as their involvement in immune response.[24–26] Microglia are known to function similarly to DCs and macrophages, derived from monocytes.[25] In AD literature, microglia is implicated as an affected cell type associated with neuroinflammation.[25,27] Our results highlight microglia as the second most enriched cell type in AD after Bonferroni correction ($p_{AD} = 4.19\times10^{-8}$). Microglia were found to be nominally significant in LBD and PD ($p_{LBD(Linear)} = 0.0119$, $p_{PD(Linear)} = 9.25\times10^{-3}$, $p_{PD(Linear)} = 1.96\times10^{-03}$, **Supplementary Tables 2, 3, 6**). Interestingly, while not shown as enriched in our ALS analyses, microglia have been shown to play an important role in the pathogenesis of ALS in part due to their activation being neurotoxic to motor neurons.[25,27–30] It is possible that due to the limited sample size in the used snRNA-seq data, the analyses were unable to detect any significant cell type enrichment in microglia for ALS.

Common risk factors across neurodegenerative diseases have always been of interest to the basic science and therapeutic industries. The association at the *SLC26A1* locus is of particular interest across multiple NDDs, as it is strongly associated with *IDUA* protein level in QTL studies, with deficiencies in this protein causing severe lysosomal storage disorders.[31,32] The *HLA* region is a complex locus from a structural genetic standpoint, but also in terms of general risk of neuroinflammation so not surprisingly resulting in a number of fine-mapped associations across this locus for multiple genes.[5,33] *GRN* is a positive control for this colocalization analysis effort as previous efforts show increased genetic risk in the region coincides with decreased expression of the gene in PD, AD and ALS.[34]

Fine-mapping efforts localizing signals to single genes within risk loci help to identify novel therapeutic targets with known biological plausibility. *ABCA1* is known to be associated with Tangier disease, which is a rare autosomal recessive disorder with low plasma levels of high-density lipoprotein (HDL) causing peripheral neuropathy.[35] *ADAM10* is implicated in the formation of amyloid plaques in the brain and the processing of APP.[35,36] CD55 is known to interact with viruses and cause neuroinflammation potentially leading to increased AD risk.[37] *COQ8A* mutations have been shown to cause coenzyme Q10 deficiency leading to autosomal recessive ataxia, cerebellar atrophy, and progressive movement disorders.[38] *C9orf72*, known to be linked to ALS, acts as a positive control here and displays a unique ALS/FTD colocalization. *ADAM10* and *C9orf72* were also nominated as potential therapeutic drug targets for neurodegenerative disease via Mendelian randomization.[16]

Limitations to this study generally relate to availability of data in this context. First and foremost, there is a limited amount of multi-ancestry or non-European data available for the GWAS and single-cell or QTL resources used here. This potentially introduces bias into therapeutic development and precision medicine applications. Secondly, low sample sizes for single nucleus analyses (in terms of the number of humans involved) reduces our ability to generate eQTL databases as compared to coarse methods of bulk RNA sequencing at scale.

Here we provided insights that could potentially aid in therapeutic development for NDDs. On the macro-level, we have identified cell-type level enrichments associated with disease risk in multiple neurodegenerative diseases allowing biologists and drug developers to better focus their mechanistic and therapeutic research. On the micro-level, we have used eQTL colocalization methods to narrow down the

large tracts of associated loci in GWAS to potentially functional variants, metaphorically going from a neighborhood to building level resolution on a map.

Data and Code availability

Code used to generate and process our data can be accessed at our Github Repository. All data for this project are publicly available via the original publications accessed by our team. Summaries of enrichment and colocalization data are also available to browse in our community target discovery and due diligence resource omicSynth web application.[16]

References

1. Dugger BN, Dickson DW. Pathology of Neurodegenerative Diseases. *Cold Spring Harb Perspect Biol*. 2017;9(7):a028035. doi:10.1101/cshperspect.a028035

2. Cuevas-Diaz Duran R, Gonzalez-Orozco JC, Velasco I, Wu JQ. Single-cell and single-nuclei RNA sequencing as powerful tools to decipher cellular heterogeneity and dysregulation in neurodegenerative diseases. *Front Cell Dev Biol*. 2022;10:884748. doi:10.3389/fcell.2022.884748

3. Siletti K, Hodge R, Mossi Albiach A, et al. Transcriptomic diversity of cell types across the adult human brain. Published online 2022. doi:10.1101/2022.10.12.511898

4. Van Rheenen W, Van Der Spek RAA, Bakker MK, et al. Common and rare variant association analyses in amyotrophic lateral sclerosis identify 15 risk loci with distinct genetic architectures and neuron-specific biology. *Nat Genet*. 2021;53(12):1636-1648. doi:10.1038/s41588-021-00973-1

5. Bellenguez C, Küçükali F, Jansen IE, et al. New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat Genet*. 2022;54(4):412-436. doi:10.1038/s41588-022-01024-z

6. Pottier C, Ren Y, Perkerson RB, et al. Genome-wide analyses as part of the international FTLD-TDP whole-genome sequencing consortium reveals novel disease risk factors and increases support for immune dysfunction in FTLD. *Acta Neuropathol*. 2019;137(6):879-899. doi:10.1007/s00401-019-01962-9

7. Chia R, Sabir MS, Bandres-Ciga S, et al. Genome sequencing analysis identifies new loci associated with Lewy body dementia and provides insights into its genetic architecture. *Nat Genet*. 2021;53(3):294-303. doi:10.1038/s41588-021-00785-3

8. Nalls MA, Blauwendraat C, Vallerga CL, et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet Neurol*. 2019;18(12):1091-1102. doi:10.1016/s1474-4422(19)30320-5

9. Höglinger GU, Melhem NM, Dickson DW, et al. Identification of common variants influencing risk of the tauopathy progressive supranuclear palsy. *Nat Genet*. 2011;43(7):699-705. doi:10.1038/ng.859

10. de Klein N, Tsai EA, Vochteloo M, et al. Brain expression quantitative trait locus and network analyses reveal downstream effects and putative drivers for brain-related diseases. *Nat Genet*. 2023;55(3):377-388. doi:10.1038/s41588-023-01300-6

11. Skene NG, Bryois J, Bakken TE, et al. Genetic identification of brain cell types underlying schizophrenia. *Nat Genet*. 2018;50(6):825-833. doi:10.1038/s41588-018-0129-5

12. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol*. 2015;11(4):e1004219. doi:10.1371/journal.pcbi.1004219

13. Murphy AE, Schilder BM, Skene NG. MungeSumstats: a Bioconductor package for the standardization and quality control of many GWAS summary statistics. *Bioinformatics*. 2021;37(23):4593-4596. doi:10.1093/bioinformatics/btab665

14. Skene NG, Grant SGN. Identification of Vulnerable Cell Types in Major Brain Disorders Using Single Cell Transcriptomes and Expression Weighted Cell Type Enrichment. *Front Neurosci*. 2016;10:16. doi:10.3389/fnins.2016.00016

15. Giambartolomei C, Vukcevic D, Schadt EE, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet*. 2014;10(5):e1004383. doi:10.1371/journal.pgen.1004383

16. Alvarado CX, Makarious MB, Vitale D, et al. omicSynth: an Open Multi-omic Community Resource for Identifying Druggable Targets across Neurodegenerative Diseases. *medRxiv*. Published online April 12, 2023:2023.04.06.23288266. doi:10.1101/2023.04.06.23288266

17. Lyon MS, Wosiski-Kuhn M, Gillespie R, Caress J, Milligan C. Inflammation, Immunity, and amyotrophic lateral sclerosis: I. Etiology and pathology. *Muscle Nerve*. 2019;59(1):10-22. doi:10.1002/mus.26289

18. Bonam SR, Muller S. Parkinson's disease is an autoimmune disease: A reappraisal. *Autoimmun Rev*. 2020;19(12):102684. doi:10.1016/j.autrev.2020.102684

19. Meier‐Stephenson FS, Meier‐Stephenson VC, Carter MD, et al. Alzheimer's disease as an autoimmune disorder of innate immunity endogenously modulated by tryptophan metabolites. *Alzheimers Dement*. 2022;8(1). doi:10.1002/trc2.12283

20. Li M, Wan J, Xu Z, Tang B. The association between Parkinson's disease and autoimmune diseases: A systematic review and meta-analysis. *Front Immunol*. 2023;14:1103053. doi:10.3389/fimmu.2023.1103053

21. Feng Y, Li L, Sun XH. Monocytes and Alzheimer's disease. *Neurosci Bull*. 2011;27(2):115-122. doi:10.1007/s12264-011-1205-3

22. Wlodarczyk A, Løbner M, Cédile O, Owens T. Comparison of microglia and infiltrating CD11c+ cells as antigen presenting cells for T cell proliferation and cytokine response. *J Neuroinflammation*. 2014;11(1):57. doi:10.1186/1742-2094-11-57

23. Mammana S, Fagone P, Cavalli E, et al. The Role of Macrophages in Neuroinflammatory and Neurodegenerative Pathways of Alzheimer's Disease, Amyotrophic Lateral Sclerosis, and Multiple Sclerosis: Pathogenetic Cellular Effectors and Potential Therapeutic Targets. *Int J Mol Sci*. 2018;19(3):831. doi:10.3390/ijms19030831

24. Neher JJ, Neniskyte U, Zhao JW, Bal-Price A, Tolkovsky AM, Brown GC. Inhibition of microglial phagocytosis is sufficient to prevent inflammatory neuronal death. *J Immunol*. 2011;186(8):4973-4983. doi:10.4049/jimmunol.1003600

25. Gallizioli M, Miro-Mur F, Otxoa-de-Amezaga A, et al. Dendritic Cells and Microglia Have Non-redundant Functions in the Inflamed Brain with Protective Effects of Type 1 cDCs. *Cell Rep*. 2020;33(3):108291. doi:10.1016/j.celrep.2020.108291

26. Kwon HS, Koh SH. Neuroinflammation in neurodegenerative disorders: the roles of microglia and astrocytes. *Transl Neurodegener*. 2020;9(1). doi:10.1186/s40035-020-00221-2

27. Monterey MD, Wei H, Wu X, Wu JQ. The Many Faces of Astrocytes in Alzheimer's Disease. *Front Neurol*. 2021;12:619626. doi:10.3389/fneur.2021.619626

28. Zhao W, Beers DR, Appel SH. Immune-mediated Mechanisms in the Pathoprogression of Amyotrophic Lateral Sclerosis. *J Neuroimmune Pharmacol*. 2013;8(4):888-899. doi:10.1007/s11481-013-9489-x

29. Quek H, Cuní-López C, Stewart R, et al. ALS monocyte-derived microglia-like cells reveal cytoplasmic TDP-43 accumulation, DNA damage, and cell-specific impairment of phagocytosis associated with disease progression. *J Neuroinflammation*. 2022;19(1). doi:10.1186/s12974-022-02421-1

30. Humphrey J, Venkatesh S, Hasan R, et al. Integrative transcriptomic analysis of the amyotrophic lateral sclerosis spinal cord implicates glial activation and suggests new risk genes. *Nat Neurosci*. 2023;26(1):150-162. doi:10.1038/s41593-022-01205-3

31. Yogalingam G, Guo XH, Muller VJ, et al. Identification and molecular characterization of α-L-iduronidase mutations present in mucopolysaccharidosis type I patients undergoing enzyme replacement therapy. *Hum Mutat*. 2004;24(3):199-207. doi:10.1002/humu.20081

32. Ruffieux H, Carayol J, Popescu R, et al. A fully joint Bayesian quantitative trait locus mapping of human protein abundance in plasma. *PLoS Comput Biol*. 2020;16(6):e1007882. doi:10.1371/journal.pcbi.1007882

33. Yu E, Ambati A, Andersen MS, et al. Fine mapping of the HLA locus in Parkinson's disease in Europeans. *npj Parkinson's Disease*. 2021;7(1). doi:10.1038/s41531-021-00231-5

34. Nalls MA, Blauwendraat C, Sargent L, et al. Evidence for GRN connecting multiple neurodegenerative diseases. *Brain Commun*. 2021;3(2):fcab095. doi:10.1093/braincomms/fcab095

35. Brunham LR, Kang MH, Van Karnebeek C, et al. Clinical, Biochemical, and Molecular Characterization of Novel Mutations in ABCA1 in Families with Tangier Disease. In: Springer Berlin Heidelberg; 2014:51-62. doi:10.1007/8904_2014_348

36. Manzine PR, Ettcheto M, Cano A, et al. ADAM10 in Alzheimer's disease: Pharmacological modulation by natural compounds and its role as a peripheral marker. *Biomed Pharmacother*. 2019;113:108661. doi:10.1016/j.biopha.2019.108661

37. Shinjyo N, Kagaya W, Pekna M. Interaction Between the Complement System and Infectious Agents - A Potential Mechanistic Link to Neurodegeneration and Dementia. *Front Cell Neurosci*. 2021;15:710390. doi:10.3389/fncel.2021.710390

38. Zhang L, Ashizawa T, Peng D. Primary coenzyme Q10 deficiency due to *COQ8A* gene mutations. *Molecular Genetics & Genomic Medicine*. 2020;8(10):e1420. doi:10.1002/mgg3.1420