



The Caribbean-Hispanic Alzheimer's disease brain transcriptome reveals ancestry-specific disease mechanisms

Daniel Felsky^{a,b,c}, Ismael Santa-Maria^{d,e}, Mehmet Ilyas Cosacak^f, Leon French^{a,b}, Julie A. Schneider^{g,h}, David A. Bennett^{g,h}, Philip L. De Jager^{c,d}, Caghan Kizil^{d,f,i}, Giuseppe Tosto^{d,i,j,*}

^a Krembil Centre for Neuroinformatics, Centre for Addiction and Mental Health, Toronto, ON, 250 College St., M5T 1R8, Canada

^b Department of Psychiatry, University of Toronto, 27 King's College Circle, Toronto, Ontario M5S 1A1, Canada

^c Center for Translational and Computational Neuroimmunology, Department of Neurology, Columbia University Medical Center, 630 West 168th Street, New York, NY 10032, USA

^d The Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Columbia University Medical Center, 630 West 168th Street, New York, NY 10032, USA

^e Department of Pathology and Cell Biology, Columbia University Medical Center, 630 West 168th Street, New York, NY 10032, USA

^f German Center for Neurodegenerative Diseases (DZNE) Dresden, Helmholtz Association, Tatzberg 41, 01307 Dresden, Germany

^g Department of Neurology, Rush University Medical Center, 1653 West Congress Parkway, Chicago, IL 60612, USA

^h Rush Alzheimer's Disease Center, Rush University Medical Center, 1653 West Congress Parkway, Chicago, IL 60612, USA

ⁱ The Department of Neurology, Columbia University Medical Center, 630 West 168th Street, New York, NY 10032, USA

^j Gertrude H. Sergievsky Centre, Columbia University Medical Center, 630 West 168th St., New York, NY 10032, USA

ARTICLE INFO

Keywords:

Caribbean-Hispanic
Brain gene expression
Alzheimer's disease
Bulk tissue
Single cell

ABSTRACT

Identifying ancestry-specific molecular profiles of late-onset Alzheimer's Disease (LOAD) in brain tissue is crucial to understand novel mechanisms and develop effective interventions in non-European, high-risk populations. We performed gene differential expression (DE) and consensus network-based analyses in RNA-sequencing data of postmortem brain tissue from 39 Caribbean Hispanics (CH). To identify ancestry-concordant and -discordant expression profiles, we compared our results to those from two independent non-Hispanic White (NHW) samples ($n = 731$). In CH, we identified 2802 significant DE genes, including several LOAD known-loci. DE effects were highly concordant across ethnicities, with 373 genes transcriptome-wide significant in all three cohorts. Cross-ancestry meta-analysis found *NPNT* to be the top DE gene. We replicated over 82% of meta-analyses genome-wide signals in single-nucleus RNA-seq data (including *NPNT* and LOAD known-genes *SORL1*, *FBXL7*, *CLU*, *ABCA7*). Increasing representation in genetic studies will allow for deeper understanding of ancestry-specific mechanisms and improving precision treatment options in understudied groups.

1. Introduction

Genetic and environmental factors conferring risk for late-onset Alzheimer's disease (LOAD) are known to differ across ancestral and ethnic populations (Reitz and Mayeux, 2014), though non-Hispanic White (NHW) cohorts still dominate genomic studies (Sirugo et al., 2019), particularly for neurodegenerative diseases. This work has been foundational for building our understanding of the pathological mechanisms of late-onset Alzheimer's disease (LOAD), and points toward perturbations in a broad set of biological pathways most often including neuroimmune responses and inflammation, synaptic regulation, oxidative stress and glucose metabolism, and fatty acid metabolism (Calabrò

et al., 2020; Morgan et al., 2022). In fact, based on existing literature, the functional reach of LOAD-related pathways is nearly universal; a recent text-mining study of over 200,000 abstracts identified that 91% of Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways show LOAD associations within at least 5 published studies (Morgan et al., 2022). Direct bulk and single-cell transcriptomic analyses have also found consistent changes in excitatory and inhibitory neuronal, microglial, and astrocytic cellular populations in LOAD (Consens et al., 2022; Grubman et al., 2019; Mathys et al., 2019). However, with respect to identifying potentially targetable molecular risk mechanisms, the relative lack of data from populations other than NHW has the potential to exacerbate existing disparities in global health equity (Jooma et al.,

* Corresponding author at: Gertrude H. Sergievsky Centre, Columbia University Medical Center, 630 West 168th St., New York, NY 10032, USA.

E-mail address: gt2260@cumc.columbia.edu (G. Tosto).

<https://doi.org/10.1016/j.nbd.2022.105938>

Received 12 August 2022; Received in revised form 21 September 2022; Accepted 29 November 2022

Available online 30 November 2022

0969-9961/© 2022 Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

2019; Weinberger et al., 2020).

Recent work has shown important differences in the genes and molecular pathways involved in risk for and progression of LOAD among different populations, though the scarcity of tissue samples outside of NHW populations has led to a strong bias in the literature toward the least invasive modalities. For example, multi-ancestry genetic analyses how include tens of thousands of individuals from African, African American (AA), Asian, and Hispanic populations (Bellenguez et al., 2022; Kunkle et al., 2021; Lake et al., 2022). Fewer studies have evaluated gene expression profiles in minority groups, with some evidence of differences in cell-free, protein coding messenger RNA (cf-mRNA) abundance in blood between LOAD and controls in AA individuals (Reddy et al., 2022). Transcriptomic analyses of brain tissue in LOAD outside of NHW populations are rarer yet, though single-nucleus RNA sequencing (RNAseq) analysis of frontal cortex from six African American (AA) and NHW individuals found that expression of LOAD-implicated genes (i.e. *APOE* in $\epsilon 4$ carriers) and gene clusters (neuronal and astrocytic) are differentially expressed between populations (Celis et al., 2020). Additionally, multi-ancestry expression analyses have demonstrated that the inclusion of diverse ancestral populations in expression quantitative trait loci (eQTL) experiments reveals previously unknown mechanistic explanations for known LOAD risk variants (e.g. rs117618017 and expression of *APHB1*) (Zeng et al., 2022).

Specifically, Caribbean-Hispanics represent an important population of individuals for the study of late-life neurodegenerative disease, as they are uniquely genetically admixed (Moreno-Estrada et al., 2013) and have undergone the most rapid demographic aging within Latin America over the last ten years (Acosta et al., 2021). Recent RNA sequencing

analyses of blood in CH individuals (including LOAD and cognitively normal elderly) from Puerto Rico found that approximately 30% of gene expression regulatory variants were not shared between CH and either NHW or African American (AA) populations (Griswold et al., 2021). Despite this, there have been no postmortem brain gene expression studies of LOAD in Hispanic/Latinos populations to date, largely due to the scarcity of available tissue samples and the historical de-prioritization of molecular-genetic research in this group. Here we present the first RNA-sequencing analysis of cortical brain tissue from Caribbean-Hispanics (CH) with and without a pathological diagnosis of LOAD. We aimed to identify genes and pathways with ancestry-specific (i.e. those identified in CH *only*) and ancestry-independent (i.e. those replicating *across groups*) differential gene expression by comparing transcriptome-wide association differential expression (DE) analyses and co-expression network-based analyses performed in CH and two independent NHW cohorts, processed with identical pipelines.

2. Results

Fig. 1 provides a schematic of our overall study design and Table 1 reports demographic and RNA sequencing depth measures for each cohort included in our study. For the CH cohort, global admixture analysis (Tosto et al., 2015b) (see Supplementary file, Fig. S1) revealed that six individuals did not show all three ancestral components (European, African and Native-American), resulting in 39 genetically-confirmed three-way admixed subjects ($n_{LOAD} = 23$, $n_{non-LOAD} = 16$); five carried a *PSEN1* mutation, the G206A substitution - a known mutation associated with familial LOAD in Caribbean Hispanic populations

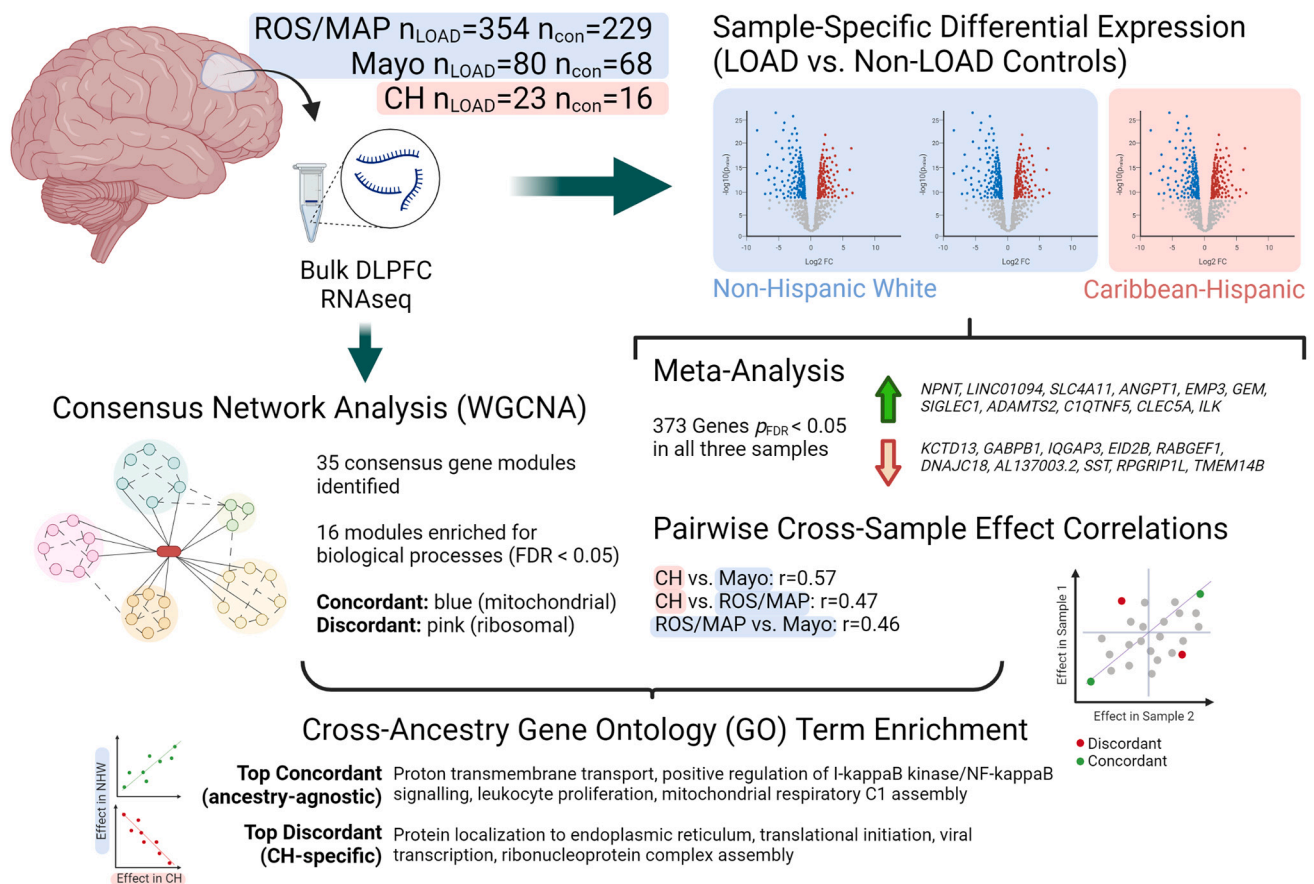


Fig. 1. Schematic showing study design and top results. CH = Caribbean Hispanic; con = non-LOAD control; DLPFC = dorsolateral prefrontal cortex; FDR = false discovery rate; LOAD = late-onset Alzheimer's disease; logFC = log fold-change; ROS/MAP = Religious Orders Study / Memory and Aging Project; WGCNA = weighted gene correlation network analysis. Created with BioRender.com.

Table 1
Sample Demographics for Caribbean-Hispanic and Non-Hispanic White RNAseq Samples.

Sample	Non-LOAD	LOAD	P-value
CH	(n = 16)	(n = 23)	
sex (M/F)	8 M, 8F	7M, 16F	0.32
age	66.9 (11.8)	79 (10.9)	0.0026
RIN	5.8 (1.5)	4.2 (1.5)	0.0024
<i>PSEN1</i> carriers (+/-)	16-	5+/18-	0.07
Library size, millions	16.2 (4.1)	15.6 (3.8)	0.65
ROS/MAP (NHW)	(n = 229)	(n = 354)	
sex (M/F)	88 M, 141F	113 M, 241F	0.11
age	87 (7.2)	90.1 (5.9)	7.7×10^{-8}
RIN	7.2 (1.1)	7 (0.9)	0.015
PMI	7.1 (4.1)	7.6 (5.2)	0.18
Library size, millions	30.1 (13.5)	26.5 (8.8)	4.9×10^{-4}
Mayo (NHW)	(n = 68)	(n = 80)	
sex (M/F)	33 M, 35F	31 M, 49F	0.25
age	82.9 (8.4)	82.6 (7.7)	0.84
RIN	7.6 (1)	8.6 (0.6)	1.4×10^{-9}
PMI	6.5 (6.5)	6.7 (4.3)	0.84
Library size, millions	40.9 (11.1)	45.4 (10)	0.013

Note: *p-values correspond to those from two-sided, two-sample *t*-tests for continuous measures (age, RIN, library size, and PMI) and from two-sided Fisher's exact tests for dichotomous measures (sex and *PSEN1* G206A mutation status). Summary measures for continuous outcomes are means (standard deviations in brackets). RIN = RNA integrity number; LOAD = late-onset Alzheimer's disease; M = male; F = female; PMI = postmortem interval; CH = Caribbean-Hispanic (all genetically-confirmed, three-way admixed subjects); NHW = non-Hispanic white; ROS/MAP = Religious Orders Study / Memory and Aging Project; PMI = postmortem interval.

(Athán et al., 2001).

2.1. Caribbean-Hispanics LOAD transcriptome-wide differential expression analysis

We first sought to identify LOAD-differentially expressed genes in CH brain. Robust regression modelling of LOAD status on gene transcript abundance ($\log_2(\text{CPM})$) was adjusted for demographic and technical covariates, including a term for *PSEN1* mutation carrier status. To further ensure that *PSEN1* status would not bias our results, we performed principal components analyses on the CH expression dataset with and without the removal of co-variate effects, confirming that mutation carriers did not cluster separately from the rest of the sample (Fig. S2). Differential expression analysis found a total of 2802 differentially-expressed genes at an FDR significant threshold of $q < 0.05$ (Table S1). Among the most significantly differentially expressed genes were many with known roles in LOAD, including *CD33*, *TREM2*, *CR1*, *CD2AP*, *HLA-DRB1*, *MS4A4A*, *FBXL7*, among additional genes identified by the most recent LOAD GWAS meta-analysis (Bellenguez et al., 2022). To functionally characterize differentially expressed genes, we analyzed those significantly up- ($n = 1335$) and down-regulated ($n = 1467$) separately, using the FUMA-GWAS tool. Up-regulated genes were significantly enriched for the "BLALOCK_ALZHEIMERS_DISEASE_UP" gene set ($p_{\text{FDR}} = 9.4 \times 10^{-29}$) as well as 2578 additional gene sets primarily related to immune cell activation, vascular function, cancers, and components of the extracellular matrix (full results are reported in Table S2). In contrast, the LOAD down-regulated gene set was enriched for the "BLALOCK_ALZHEIMERS_DISEASE_DN" gene set ($p_{\text{FDR}} = 6.2 \times 10^{-21}$) and showed fewer additional significant enrichments (152), mostly corresponding to neuronal signaling components, consistent with synaptic loss observed in LOAD.

2.2. Non-Hispanic differential expression analyses and cross-ancestry gene expression comparison

In the Religious Orders Study and Memory and Aging Project (ROS/MAP) and Mayo Clinic non-Hispanic samples, differential expression analyses found 2066/17,665 and 10,878/19,380 significant genes, respectively (Fig. 2D-E; summary statistics in Tables S3 and S4). Validation of these results against published differential expression analyses in these cohorts (Canchi et al., 2019; Logsdon et al., 2019; Mostafavi et al., 2018) demonstrated substantial coherence of transcriptome-wide effects (Supplementary file; Figs. S3 and S4). Comparing between populations, 373 genes reached genome-wide significance in all three differential expression analyses ($n_{\text{universe}} = 15,834$, expected overlap = 184; overlap hypergeometric $p = 6.6 \times 10^{-43}$); 364 (98%) with concordant directions of effect (Fig. 2F). Among these, many were of known significance to LOAD, including *TREM2*, *ICAI*, *VASP*, *MTMR3*, *GFAP*, and *GRIK1* (cross-sample effects of top up- and down-regulated genes shown in Fig. 2G). At uncorrected $p < 0.05$, 1441 genes were significant in all three analyses, 1360 (94%) with concordant directions of effect. Cross-ancestry meta-analysis found *NPNT* to be the top DE gene ($p = 1.04 \times 10^{-19}$). Among the LOAD-known genes, several were significantly DE, including *FBXL7*, *TREM2*, *IQCK*, *RIN3*, *CD33*, *INPP5D*, *PTK2B*, *CLU*, *AD2AP*, *MEF2C* and *WVVOX* (full list in Table S5), with 43 LOAD-known genes at least nominally DE.

We then aimed to identify genes with ancestry-specific effects by comparing differential expression statistics between 1) CH vs. ROS/MAP, 2) CH vs. Mayo, and 3) ROS/MAP vs. Mayo, among the set of genes analyzed in all three studies ($n_{\text{genes}} = 15,834$). We observed a greater proportion of replicated signals between CH and Mayo (1937 in common out of 2588 in CH = 75% replication; expected overlap = 1511, overlap $p = 7.8 \times 10^{-81}$) than between CH and ROS/MAP (450/2588 = 17% replication; expected overlap = 315, overlap $p = 1.3 \times 10^{-17}$). Consistently, pairwise Pearson correlations of the three study's t-statistics revealed moderate positive relationships (CH vs. Mayo $r = 0.57$, C.I._{99%} = [0.56,0.59]; CH vs. ROS/MAP $r = 0.47$, C.I._{99%} = [0.46,0.49]; ROS/MAP vs. Mayo $r = 0.46$, C.I._{99%} = [0.45,0.48]) (Fig. S5). There were 118 individual genes with significant effects in CH ($p_{\text{FDR}} < 0.05$) and an opposite direction of effect in both Mayo and ROS/MAP (top 10 over- and under-expressed shown in Table S6).

2.3. Identification of ancestry-discordant and -concordant LOAD-related biological processes

To harness the full sets of ranked summary statistics for differential expression in all three analyses, we calculated the pairwise product of t-statistics ("tprod") for each gene between each pair of analyses. This metric provides a transcriptome-wide ranked measure of similarity in expression magnitude and direction between samples, and was used as input for rank-based gene set enrichment tests. Using this approach, we identified 20 unique, semantically non-redundant biological processes with significant ancestry-discordant or -concordant LOAD-related effects in at least one cross-sample comparison (Fig. 3A). CH vs. Mayo (the comparison with the greatest number of discordant ($n = 5$) and concordant ($n = 11$) gene sets): discordant processes included those related to transcription and translation, whereas concordant processes included those related to cellular metabolism, immune response, synaptic transmission, and biomineralization. Notably, the "response to beta-amyloid" ($p_{\text{FDR}} = 0.013$) process was also concordantly down-regulated in both datasets, though it was not assigned a dispensability score of 0 by REVIGO (disp = 0.07). Similar themes emerged for both the CH vs. ROS/MAP and ROS/MAP vs. Mayo comparisons, where ribosomal genes were again found to be enriched for discordant effects and neuronal genes for concordant effect, though concordant signals between ROS/MAP and Mayo also included exocrine system development (Fig. 3A,B). Importantly, several of the identified GO groups across sample comparisons had strong overlap in gene identity (S7 Figure),

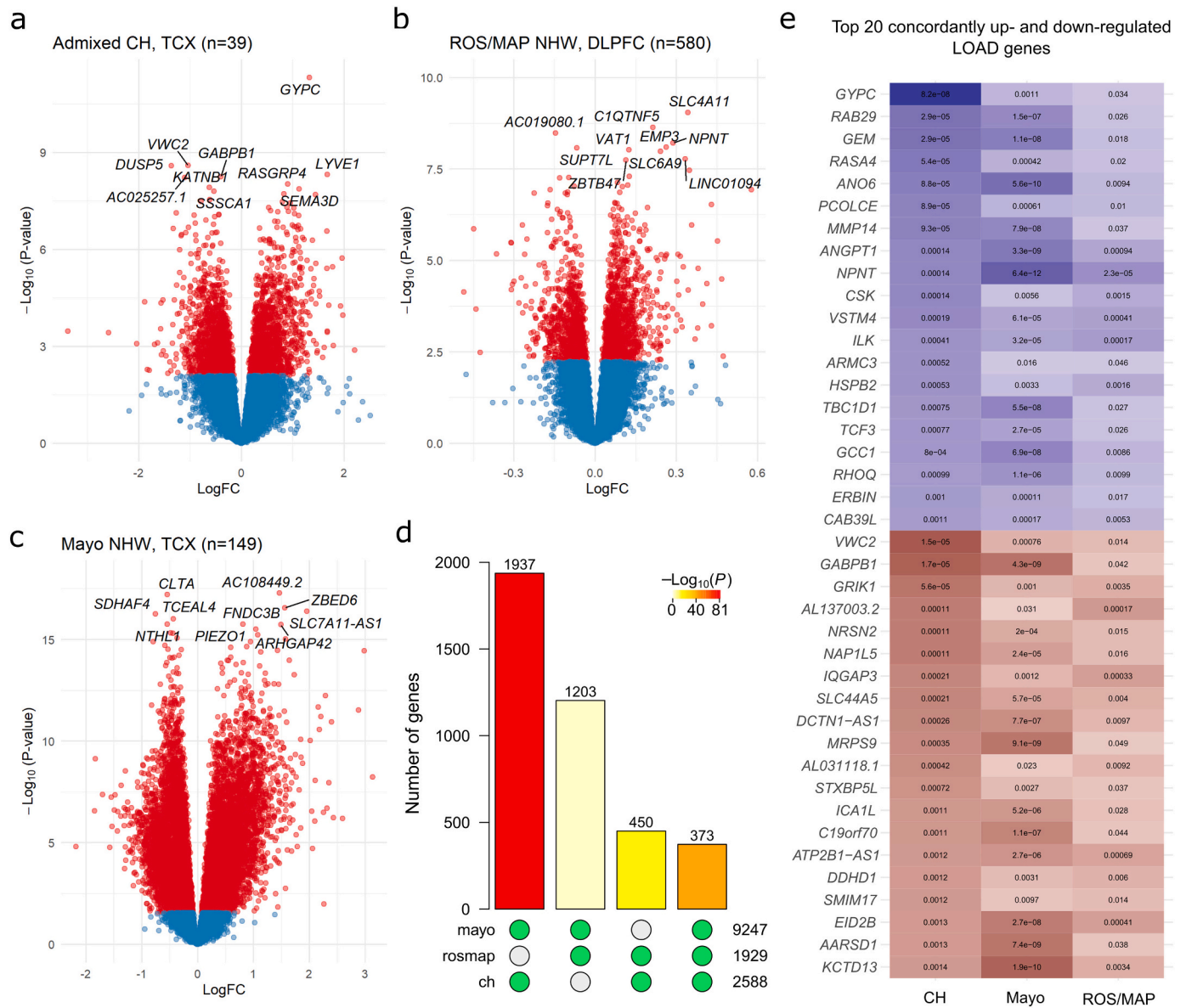


Fig. 2. Results of single-gene transcriptome-wide associations with LOAD in Caribbean-Hispanics and non-Hispanic Whites. Gene-level volcano plots for association of genes with LOAD status in A) genetically-confirmed CH subjects ($n = 39$), and two independent samples of non-Hispanic whites: B) ROS/MAP and C) the Mayo RNAseq cohort. The top ten significantly differentially expressed genes in each analysis are labeled. The Y-axes indicate two-sided $-\log_{10}(p\text{-values})$ for robust regression testing differential expression. The X-axes indicate \log_2 fold-change in expression. D) Barplot showing the number of transcriptome-wide FDR-significant ($q < 0.05$) differentially expressed genes in common between each pair of samples and across all three datasets. Green dots below each bar indicate which sets of significant genes are intersecting. Bar color indicates the p -value for hypergeometric testing of overlap probability (scale shown). E) The top 20 up- and top 20 down-regulated LOAD genes with significant, concordant effects across all three samples, ranked by significance in the CH sample. The color is proportional to moderated t -statistics (blue for positive effect, or higher expression in LOAD, and red for negative effect, or lower expression in LOAD), with FDR-adjusted two-sided p -values labeled inside of tiles. DLPFC = dorsolateral prefrontal cortex source tissue; LogFC = \log_2 fold-change in expression; NHW = non-Hispanic White; TCX = temporal cortex source tissue. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

with a core set of 78 genes highly overrepresented in protein localization, viral transcription, and translational initiation categories driving discordant effects.

2.4. Weighted Gene Co-Expression Network Analysis (WGCNA)

Moving beyond the individual gene level to gene networks, we performed consensus co-expression network analyses on all three datasets to understand the degree of shared and distinct network architecture among CH and NHW transcriptomes. A total of 35 discrete consensus gene modules were identified across CH and NHW samples, ranging from 51 to 2715 genes in size (Fig. 4A; Table S7), with 16 modules

significantly enriched for at least one biological process (Table S8). Association tests of module eigengenes with LOAD in each dataset revealed largely conserved network-level effects, where CH vs. ROS/MAP effects were most cohesive ($r = 0.81$, C.I._{95%} = [0.65,0.90]), followed by Mayo vs. CH ($r = 0.68$, C.I._{95%} = [0.45,0.83]) and Mayo vs. ROS/MAP ($r = 0.56$, C.I._{95%} = [0.28,0.75]). Several modules with concordant and discordant LOAD associations were identified (Fig. 4B-E; full results in Table S9), among which the blue concordant module was enriched for mitochondrial genes. Further, the discordant pink module (significantly over-expressed in CH, but under-expressed in Mayo) largely represented ribosomal genes (“SRP-dependent cotranslational protein targeting to membrane” $p_{\text{FDR}} = 7.0 \times 10^{-80}$).

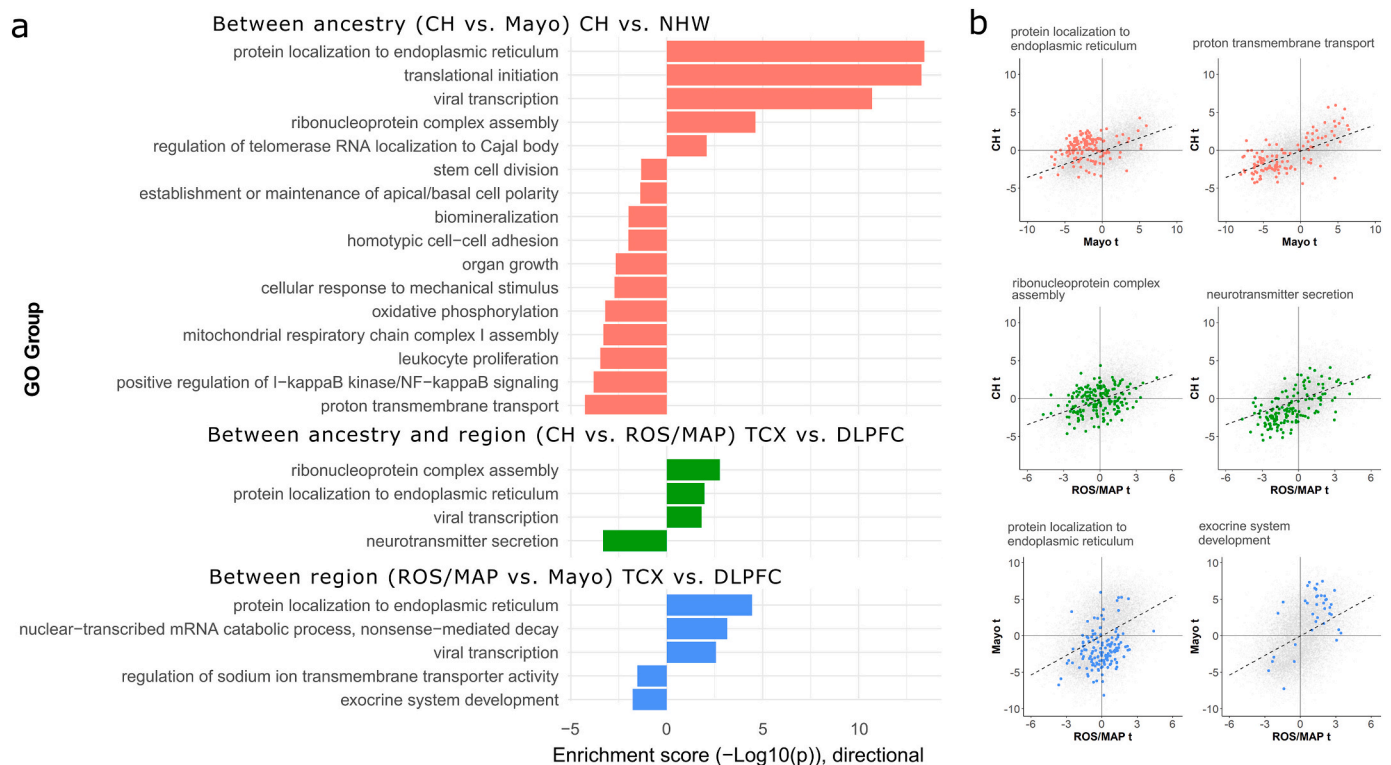


Fig. 3. Enrichment analyses for genes with concordant and discordant LOAD effects between ancestry and brain region tissue source. A) Barplot summarizing AUC-based GO enrichment analysis using tprod ranks for all FDR-significant GO terms with a REVIGO dispensability score of 0. The x-axes indicate the enrichment $-\log_{10}$ (two-sided p -values), with values above 0 indicating enrichment toward higher ranks (greater discordance of between-sample effect) and values below 0 indicating enrichment toward lower ranks (greater concordance of between-sample effect). B) Scatterplots for the top discordant (left column) and concordant (right column) groups from panel A for each sample contrast, showing the LOAD differential expression t -values for genes belonging to each enriched gene set within the context of all genes. Genes belonging to labeled GO sets in panel B are colored to match the barplot in panel A. CH = Caribbean-Hispanic; FDR = false discovery rate; GO = gene ontology; DLPFC = dorsolateral prefrontal cortex source tissue; ROS/MAP = Religious Orders Study and Memory and Aging Project; TCX = temporal cortex source tissue.

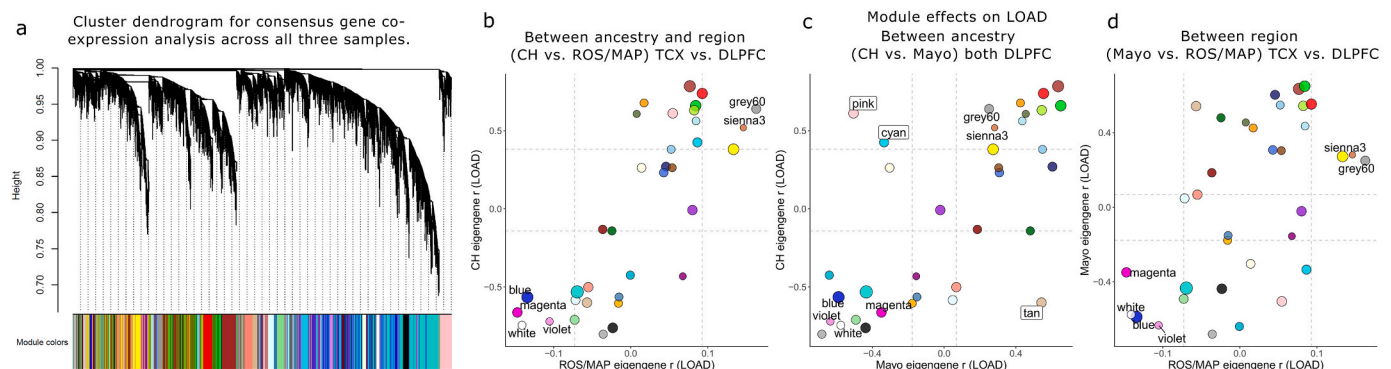


Fig. 4. Consensus gene co-expression network module analysis with LOAD. A) Clustered dendrogram showing the common gene set hierarchical structure and consensus gene module definitions. Panels B–D show consensus module eigengene associations with LOAD (Pearson r) for all modules compared between pairwise sample combinations. Modules with concordant effects ($p_{FDR} < 0.05$ with same direction of effect) in all three samples, or significantly discordant ($p_{FDR} < 0.05$ in at least two samples but with opposite direction of effect), are labeled; six modules were significantly associated with LOAD in all three samples; greenyellow, brown, and royalblue modules were consistently downregulated, while the red, purple, and lightyellow modules were upregulated. In contrast, four modules showed significant but directionally discordant effects on LOAD when comparing CH to Mayo; salmon and lightcyan modules were significantly upregulated in CH but downregulated in Mayo, whereas black and cyan modules were downregulated in CH but upregulated in Mayo. Colors correspond to module definitions. CH = Caribbean-Hispanic; DLPFC = dorsolateral prefrontal cortex; ROS/MAP = Religious Orders Study and Memory and Aging Project; TCX = temporal cortex. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2.5. Comparison of bulk RNAseq to single cell transcriptomics

To localize our previous findings to specific brain cell types, we analyzed high-depth single-nucleus RNAseq data (snRNAseq) from entorhinal cortex (EC) and superior frontal gyrus (SFG) (Leng et al., 2021).

Following clustering, we determined seven major cell types: astroglia, microglia, inhibitory neurons, excitatory neurons, oligodendrocytes, oligodendrocyte progenitors and endothelia (Fig. 5A, B). We then performed differential expression analysis between AD cases and controls (Braak stage 6 versus Braak stage 0, full results for EC per cell type in

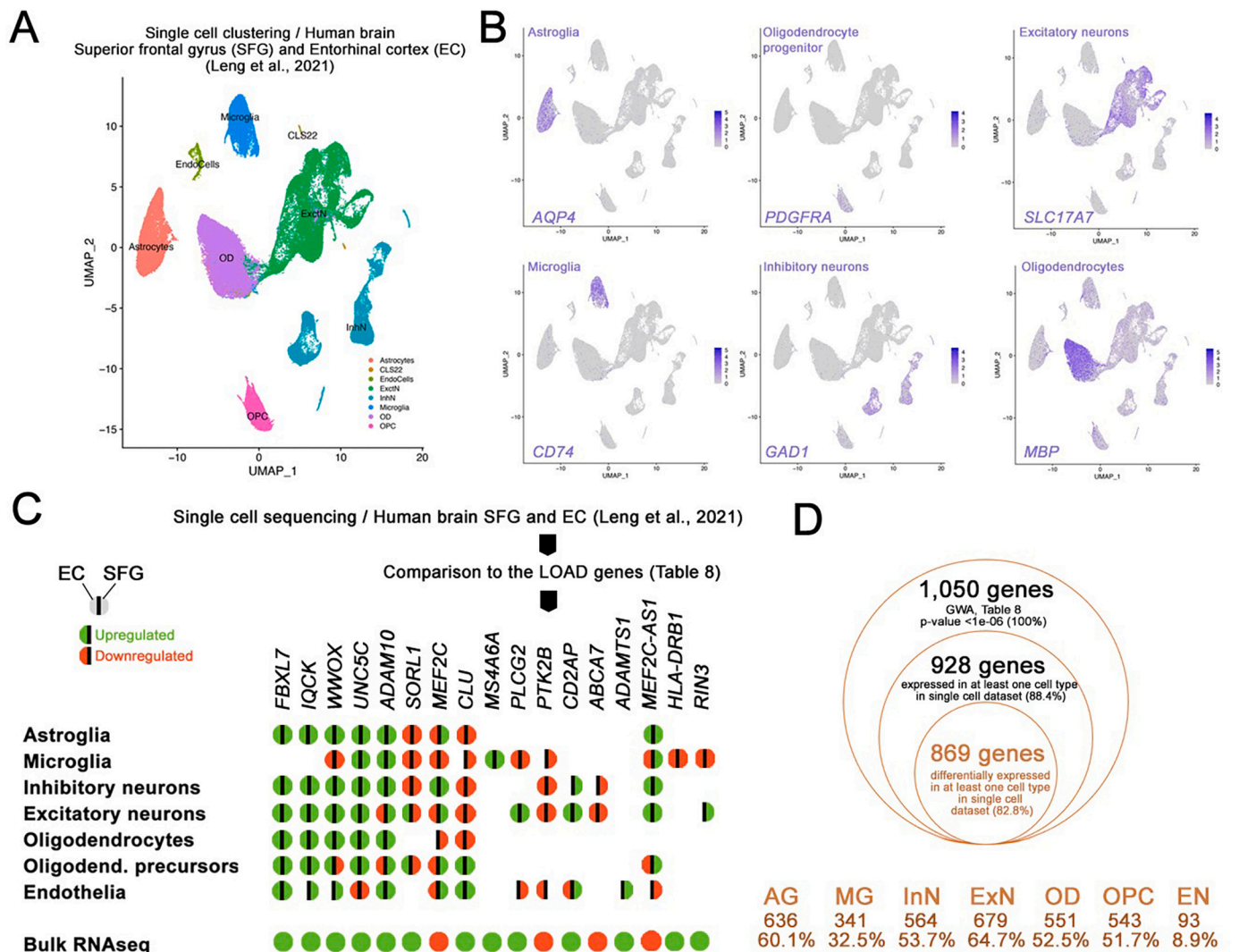


Fig. 5. Secondary independent validation of the bulk RNA sequencing with single cell transcriptomics from human brains. (A) Clustering of single cell dataset in Leng et al., 2021 (human brains, EC and SFG, AD cases versus controls; Braak stage 6 versus 0). 7 major cell types were identified as indicated. (B) tSNE plots for major cell type markers (only one out of many were shown) *AQP4* for astroglia, *PDGFRA* for oligodendrocyte progenitors, *CD74* for microglia, *GAD1* for inhibitory neurons, *MBP* for oligodendrocytes, *SLC17A7* for excitatory neurons. (C) Genes in Table 8 were run against the single cell datasets and differentially expressed genes in various cell types in entorhinal cortex (EC) and superior frontal gyrus (SFG) of human brains with and without 17 CE genes are shown for their differential gene expression in major cell types. Green: upregulation, red: downregulation, left hemisphere: EC, right hemisphere: SFG. No color indicates no significant differential expression. (D) Comparison of GWA genes (Table 8) in single cell dataset (human EC, Braak stage 6 vs 0). 869 genes (82.6% of the initially selected GWA genes) were validated for their differential expression in human EC in at least 1 cell type. Cell types and the number/percentage of the initially selected GWA genes are shown for astroglia (AG), microglia (MG), inhibitory neurons (InN), excitatory neurons (ExN), oligodendrocytes (OD), oligodendrocyte progenitors (OPC), and endothelia (EN). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table S10) within each cell type cluster separately and cross-referenced these results with our bulk brain-based findings. Regarding the cross-ancestry meta-analysis, 1050 genes were gene-wide significant ($p < 1 \times 10^{-6}$, Table S8): 938 of these genes were expressed in at least 1 cell type in the scRNA-seq, while 829 genes were differentially expressed (DEGs) in at least one cell type in the single cell (Fig. 5D). Excitatory neurons (ExN) showed 679 DEGs (64.7% of the GWA genes in Table S8), followed by astroglia (AG, 636 DEGs, 60.1%), inhibitory neurons (InN, 564 DEGs, 53.7%), oligodendrocytes (OD, 551 DEGs, 52.5%), oligodendrocyte progenitors (OPC, 543 DEGs, 51.7%), microglia (MG, 341 DEGs, 32.5%), and endothelia (EN, 93 DEGs, 8.9%) (Fig. 5D). These genes included the top hit *NPNT* (AG, InN, OPC, ExN). In addition, 69 out of 75 known LOAD-associated genes were also differentially expressed in at least one cell type in our single-nucleus analysis (Fig. 5C), including *SORL1*, *ABCA7*, *PLCG2*, *CLU*, *FBXL7*, *UNC5C*. For instance, *ABCA7* was found DE in neurons only while *FBXL7* and *IQCK*

were differentially expressed in all cell types but microglia. Finally, *GYPC* (top DE gene in CH transcriptome) was also found upregulated in astroglia ($p = 0.012$), microglia ($p = 0.009$), OPC ($p = 3.9 \times 10^{-6}$), excitatory neurons ($p = 2.2 \times 10^{-8}$).

3. Discussion

To our knowledge, this is the first investigation of gene expression from postmortem LOAD brain in CH individuals. Understanding ancestry-specific molecular profiles of LOAD brain tissue is a first step toward developing interventions effective in non-European Caucasian high-risk populations. We report over 2000 genes differentially expressed in LOAD brain vs. controls within the CH cohort, many with known relevance to LOAD pathophysiology. For example, the top up-regulated gene was *GYPC*, or glycophorin C, which is found upregulated in human LOAD microglia (Srinivasan et al., 2020) (notably, this

gene is also nominally significant in both ROS/MAP and Mayo; $p = 0.003$, $p = 0.0003$ respectively). In our single-cell analyses, we consistently found this gene upregulated in astroglia, microglia, oligodendrocyte progenitors and excitatory neurons (Table S9). Up-regulated genes in CH were found significantly enriched for several GWAS catalog traits, including “late onset Alzheimer’s disease” ($p_{FDR} = 0.03$) and “intraocular pressure” ($p_{FDR} = 5.6 \times 10^{-3}$). The latter is a condition associated with open-angle glaucoma, which shows higher prevalence in Hispanics than NHW (Grosvenor and Hennis, 2011; Quigley et al., 2001) as well as overlaps with genetic risk for LOAD across multiple populations (Gharahkhani et al., 2020).

Our results show substantial overlap in LOAD-related genes between CH and NHW - particularly when comparing same brain region (i.e. in temporal cortex; CH vs. Mayo). This higher concordance suggests that differences in LOAD-related gene expression patterns are greater between sampled brain regions than between ancestries. This degree of inter-regional concordance mirrors recent findings of differential gene expression across multiple brain regions in LOAD (Patel et al., 2019) and Schizophrenia (Collado-Torres et al., 2019). Several genes were consistently differentially expressed across cohorts: among the most significant, many were LOAD known-loci. Notably, *BACE2*, the lesser known homolog (Wang et al., 2019, p. 2) of the LOAD-associated β -secretase enzyme *BACE1*, was the eighth most strongly overexpressed gene in our study ($p_{FDR} = 4 \times 10^{-5}$). Most importantly, we found the known CH GWAS-implicated LOAD risk gene, *FBXL7* (Tosto et al., 2015a), overexpressed in LOAD at FDR-corrected $p_{FDR} < 0.05$ in all three samples. At the pathways level, those related to immune cell proliferation, cell killing, and neurotransmitter secretion were found overlapping.

In contrast, among the strongest ancestry-specific genes was *CSPG4* (see Table S6), a chondroitin sulfate proteoglycan (also known as neuron-glia antigen 2) which is expressed in oligodendrocyte precursor cells and a subpopulation of NG2+ astrocytes, has been strongly linked to LOAD pathology in humans and mice (Nielsen et al., 2013; Vanzulli et al., 2020), and harbors rare genetic variants in Central American/Hispanic families linked to schizophrenia (de Vrij et al., 2019). Core genes that were increased in LOAD in CH, but decreased in NHW, suggested ancestry-driven differences in the activation of translational machinery in LOAD. Ancestry-specific enrichment analyses highlighted roles of ribosomal genes and those involved in protein synthesis and trafficking, which are emerging as robust and important markers of LOAD pathogenesis (Garcia-Esparcia et al., 2017; Ghosh et al., 2020; Grothe et al., 2018; Hernández-Ortega et al., 2016; Patel et al., 2020; Rasmussen et al., 2015; Shigemizu et al., 2020). We speculate that (SRP)-dependent protein targeting may be disproportionately perturbed in CH patients, given the strong population-specific enrichment in both the single gene and network co-expression analyses. Intriguingly, SRP-dependent genes are strongly dysregulated in periodontitis-affected periodontal tissue (Lundmark et al., 2018). The main driver of periodontitis pathogenesis, *P. gingivalis*, is known to secrete gingipains which have been linked to LOAD pathogenesis in mice (Ilievski et al., 2018) and humans (Dominy et al., 2019). This connection is re-enforced by the significant enrichment of the GWAS catalog periodontitis gene set in our CH LOAD-upregulated genes. Notably, nephronectin, coded by the *NPNT* gene (the top DE gene in our cross-ancestral meta-analysis) has been shown to have a critical role in human dental pulp stem cells (Tang and Saito, 2017) and is a top ranking gene in large AD proteomic studies (Bai et al., 2020).

Among our top results, the “protein localization to endoplasmic reticulum” GO group was unanimously enriched in discordantly LOAD-related genes, whether comparing between ancestry, between sampled brain regions, or both ancestry and region (Fig. 3A). Notably, this enrichment was most significant in the between ancestry comparison (same brain region; CH vs. Mayo), highlighting a hierarchy of effects for protein trafficking and related biological processes (e.g. “translational initiation” and “viral transcription; see Fig. S7 for overlap) between the

three cohorts. In CH, these genes are largely up-regulated, whereas in Mayo (NHW), these genes are downregulated. While there is moderate trend toward downregulation in ROS/MAP (also NHW), the overall signal driving the strongest enrichment appears to be driven by the unique increases in protein trafficking genes in the CH cohort. While this result is difficult to contextualize given the lack of comparable CH samples, previous work has demonstrated that the effects of LOAD biomarkers (e.g. CSF tau, CSF IL-9, MRI-based functional connectivity) may be strongly attenuated or even reversed in direction between NHW and African American populations (Babulal et al., 2019; Howell et al., 2017; Wharton et al., 2019). Pilot analyses ($n = 9$ AA; $n = 11$ NHW) from postmortem brain has also showed over 180 proteins with significant race x LOAD diagnosis interactions, pointing again to ancestry discordant molecular signatures of LOAD and motivating future multi-ancestry ‘omic investigations (Stepler et al., 2020).

Bulk RNA sequencing is a powerful method for determining tissue-level differences in gene abundance. However, is incapable of determining cell type-specific changes in a definitive manner. In our single-nucleus analyses, we validated over 82% of the genes identified by our cross-ancestry meta-analysis, including the top hit, *NPNT*. This gene was found upregulated in astroglia, oligodendrocyte progenitors and neurons. Several genes with known significant to LOAD showed high consistency between our bulk and scRNAseq data. For example, *ABCA7* was found downregulated in neurons only and *MS4A6A* was found differentially expressed in microglia. *UNC5C* was found upregulated in all major cell types, with the exception of endothelial cells, where it was found to be downregulated. *CLU* (found to be upregulated in bulk analyses) was upregulated in endothelial cells and oligodendrocyte precursors, but downregulated in other cell types. This suggests that these genes may have different roles in cell-specific physiology. In contrast, *FBXL7* - which has a genome-wide significant risk variant for LOAD in CH individuals (Tosto et al., 2015a) - was found upregulated in nearly all cell types. Our results demonstrate that comparison between bulk and single cell sequencing data can provide useful information on cell types that differentially express genes of interest. This approach can refine biological hypotheses on the functional relevance of these genes to LOAD pathophysiology.

Key limitations in this study include different methods of tissue ascertainment and study design that ultimately impact the results; ROS and MAP are community-based, prospective cohort studies, whereas the CH and Mayo samples are case-control designs selected for diagnosis. Additionally, CH samples were not scanned for rare LOAD mutations using whole genome sequencing, though *PSEN1* carriers were identified by imputation or brain-bank record (see Supplementary file). CH specifically are a relatively small brains sample collected over several decades; the difficulties of ascertaining brains in this specific population have mired the availability of large sample sizes of tissue collected over shorter periods (i.e. higher quality). Hispanics tend to not participate in either organ donation in general, or brain donation more specifically, to the same extent as non-Hispanic Whites (NHW) (Frates and Garcia Bohrer, 2002). This resulted in samples with relatively low RNA quality (median RIN = 4.5). Nevertheless, by performing ribosomal RNA depletion prior to sequencing, we ensured that the impact of low RNA quality was mitigated. In fact, ribosomal RNA depletion has been shown to perform very well (Kumar et al., 2017; Schuierer et al., 2017) at amounts far below recommendation and over a wide range of intact and degraded material. Despite these challenges, our analyses proved reliable by showing a moderate concordance between results from all three samples at the genome-wide scale. We were also able to identify many well-known LOAD-associated loci from previous GWAS and sequencing studies such as *FBXL7* (previously identified by a CH-specific GWAS from our group (Tosto et al., 2015a)). Finally, we note that differences in age at death between LOAD and non-LOAD participants in the ROS/MAP and CH cohorts were observed (Table 1), meaning that correction for age at death in linear modelling may have partially accounted for a small portion of LOAD-specific differences between groups.

In sum, we performed RNA-sequencing on postmortem brain from a small sample of CH elderly and two large independent NHW cohorts, ultimately identifying numerous candidate genes and biological processes that are consistently dysregulated in LOAD across ancestry or show ancestry-specific dysregulation. Among those genes showing the strongest ancestry-specific effects, we found significant enrichment for ribosomal pathways. Further work in large admixed cohorts will permit a deeper understanding of ancestry-specific mechanisms that may be used to predict risk, onset of pathology, and potentially provide precision treatment options.

4. Materials and methods

4.1. Caribbean Hispanic cohort

We processed postmortem temporal cortex (TCX) tissue from 45 unrelated self-reported CH individuals ascertained from the New York Brain Bank (NYBB) at Columbia University (New York, NY, USA). This research clinic referral-based brain bank consists of over 600 brains from autopsies performed between 1989 and 2016. Cases were selected if age > 50 years old, and neuropathologically defined LOAD or control without any neuropathological diagnosis. The protocol was approved by the institutional review boards of the New York State Psychiatric Institute and Columbia University. Details of the tissue preparation and neuropathological assessment are reported in the Supplementary file and extensively described elsewhere (Vonsattel et al., 1995, 2008). RNA-sequencing methods, expression quantification, and initial quality control are reported in the Supplementary file. Whole-genotyping and admixture estimation has been extensively described in a recent publication (Sariya et al., 2021) from our group and also reported in the Supplementary file.

4.2. Non-Hispanic White (NHW) cohorts

We included in this study two NHW cohorts: 1) The Religious Orders Study and Memory and Aging Project (ROS/MAP) and 2) the Mayo Clinic RNAseq cohort. The sample characteristics of the ROS/MAP cohort subset studied here have been published in detail elsewhere (Bennett et al., 2018). Briefly, RNA sequencing data was analyzed from postmortem dorsolateral prefrontal cortex (DLPFC) of 595 subjects (Bennett et al., 2018). After quality control (QC; see Supplementary file), 583 subjects remained ($n_{AD} = 354$, $n_{non-AD} = 229$). The Mayo RNAseq cohort has also been described in detail (Allen et al., 2018; Allen et al., 2016). After QC (see Supplementary file), 148 subjects remained ($n_{AD} = 80$, $n_{non-AD} = 68$). Protocols were approved by the Mayo Clinic Institutional Review Board and all subjects or next of kin provided informed consent. Neuropathological evaluation details and RNAseq methods are reported in the Supplementary file.

4.3. Ethics approval and consent to participate

For the Caribbean-Hispanic Alzheimer's Disease Research Centre (ADRC) New York Brain Bank dataset, the protocol was approved by the institutional review boards of the New York State Psychiatric Institute and Columbia University. For the Religious Orders Study and Memory and Aging Project, all study participants provided informed consent and both studies were approved by the Rush University Institutional Review Board. For the Mayo dataset, protocols were approved by the Mayo Clinic Institutional Review Board and all subjects or next of kin provided informed consent.

4.4. RNAseq processing and LOAD Transcriptome-wide differential expression analysis

Gene counts for all three cohorts were post-processed identically. Counts were read into R (v3.6.3) for processing with edgeR (Robinson

et al., 2010) and limma/voom (Law et al., 2014; Ritchie et al., 2015). TMM normalization values (using edgeR calcNormFactors) and mean-variance derived observational-level weights were then calculated for linear modelling using limma/voom. For differential expression analyses, the limma package 'lmFit' function was used to model $\log_2(\text{expected counts})$ as a linear function of pathological LOAD status, sequencing batch (brain bank source and flowcell included for Mayo), age, sex, RIN, postmortem interval (not available for the CH sample), percent of usable bases (not included for CH, as it was highly co-linear with RIN, and additional parameters were detrimental to model performance given small sample size), percent duplicated reads, median 3' bias, and percent of mapped ribosomal bases. In CH models that included *PSEN1* mutation carriers, a parameter was added to model its effect. Significance of the contrast between LOAD and non-LOAD status (LOAD vs. control for Mayo) was performed using empirical Bayes moderation (eBayes function).

For ROS/MAP, a stratified sensitivity analysis was performed to ensure that differences in the age range between ROS/MAP and CH samples did not impact our comparisons (Supplementary file). Further validation analyses were performed to ensure consistency of our differential expression analyses of ROS/MAP and Mayo cohorts with previously published work (Supplementary file).

4.5. Differential expression cross-ancestry meta-analysis

We meta-analyzed results from the three transcriptome-wide differential expression analyses employing a fixed-effect model and using logFC and its standard error (SE) as inputs, consistently with previous studies of transcriptomics in LOAD (Patel et al., 2019). The SE was calculated from the logFC values for each gene, separately in each cohort, and used for standard meta-analysis using the R package "rmeta". LogFC and se ensured that direction of effect was taken into consideration when performing the meta-analysis. Multiple testing correction was applied using the false discovery rate (FDR) method.

4.6. Cross-ancestry comparison of differential expression

To compare differential expression results from our CH and NHW sample analyses (i.e. ROS/MAP and Mayo), we identified genes with 1) *ancestry-independent effects* - i.e. those with replicated associations with LOAD across CH and NHW samples, and 2) *ancestry-specific effects* - i.e. those significant in CH and neither NHW samples. To incorporate directionality of effect, magnitude of effect, and effect standard error into our assessment of effect concordance and discordance across samples, we ranked genes by the product of their t-statistics ("tprod") for association with AD, assigning the highest ranks to genes with the largest discrepancy in direction and effect size between ancestries (i.e. CH vs. Mayo and CH vs. ROS/MAP, separately). Thus, higher ranks indicated sample-specific ("ancestry-discordant") effects, whereas lower ranks indicated sample-nonspecific ("ancestry-concordant") effects. To test the impact of gene inclusion criteria, effect statistic, and effect direction on inter-sample correlations, we performed a series of sensitivity analyses (Supplementary file and S6 Figure).

4.7. Functional enrichment analyses

To functionally characterize AD-associated genes in our CH DE analysis we used the Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA-GWAS) GENE2FUNC online utility (<https://fuma.ctglab.nl/gene2func/>). We performed hypergeometric tests for gene set enrichment separately for genes significantly (transcriptome-wide FDR-adjusted p -value < 0.05) up- and down-regulated in LOAD. For rank-based enrichment analyses of cross-sample concordant and discordant gene lists, an area under the curve (AUC) rank-based method was used, as implemented in the tmod R package (Weiner, 2020). Resulting significant GO groups ($p_{FDR} < 0.05$) were consolidated

to minimize semantic redundancy using the REVIGO tool (Supek et al., 2011). For the presentation of results, we chose to select only GO terms with a REVIGO dispensability score of 0; this decision was made to provide the most succinct summary of an otherwise large number of enrichment results while retaining all major relevant pathways. Annotations were limited to those including between 10 and 200 tested genes. GO annotations were extracted from the org.Hs.eb.db and GO.db (version 3.10.0) R packages (Carlson, 2016a, 2016b). For gene co-expression modules, hypergeometric testing was used to evaluate functional enrichment, using the background universe of genes passing QC for consensus module analysis. For all three samples, genes assigned to each consensus module were tested for GO term enrichment including categories of biological processes, cellular components, and molecular functions (Ashburner et al., 2000). FDR correction was used ($q < 0.05$) to determine significant enrichment.

4.8. Consensus Weighted Gene Co-Expression Network Analysis (WGCNA)

Consensus gene co-expression modules were defined across all three samples using Weighted Gene Co-Expression Analysis (WGCNA) (Zhang and Horvath, 2005) applied to batch- and covariate-corrected partial gene expression residuals (partial with respect to LOAD diagnosis). Detailed description of the method and non-default parameters used in the Supplementary file.

4.9. Comparison of bulk tissue results with single-nucleus RNA sequencing data from human brain

single-nucleus RNA sequencing data ($n = 10$) for human superior frontal gyrus (SFG) and entorhinal cortex (EC) samples were downloaded from the Gene Expression Omnibus repository under the following GEO ID; GSE147528 (Leng et al., 2021). After extensive QC, we used Seurat V3.1.5 (Hao et al., 2021) for further analyses as described in the Supplementary file. In total, 26 cell clusters (numbered from 0 to 25) were identified. We then generated heat maps and cell types were inferred based on the characteristic gene expression patterns. Major cell types were identified using the following markers (based on Leng et al., 2021): *GFAP*, *SLC1A2* and *AQP4* (astroglia); *MBP* and *MOG* (oligodendrocytes or OD); *PDGFRA* and *SOX10* (oligodendrocyte precursors or OPC); *CD74* and *CX3CR1* (microglia); *SLC17A7* and *CAMK2A* (excitatory Neurons, or ExctN); *GAD1* and *GAD2* (inhibitory Neurons or OD); *CLND5* and *FLT1* (endothelial cells or EndoCells). Details are provided in the Supplementary file. To identify the differentially expressed genes (DEGs), we compared cases (samples with Braak stage 6) vs. controls (Braak Stage 0) within each main cell type, across Entorhinal Cortex (EC) and superior frontal gyrus (SFG), using the FindMarker function in Seurat. To identify which cell types exhibited overlapping DEGs with our findings from bulk RNAseq, we selected experiment-wide significant (p -value $< 1.0 \times 10^{-6}$) from our cross-ancestry meta-analysis, and searched for differentially expressed genes ($p < 0.05$) in the single-nucleus experiment. We also perform the same investigation for LOAD known-genes.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.nbd.2022.105938>.

Funding

This work was supported by grants from the National Institutes of Health: R56AG069118, R56AG066889, R21AG054832, R56AG059756, R01AG056531, P50AG008702, P30AG10161, R01AG15819, R01AG17917, U01AG61356, R01AG015473, R56AG063908, P30AG066462, R01NS095922 and P50AG0008702. Funding support for DF was provided by The Koerner Family Foundation New Scientist Program and The Krembil Family Foundation. Taub Institute Schaefer Research Scholars Award (CK) and German Center for

Neurodegenerative Diseases (DZNE) (MIC, CK) provided support.

Author contributions

DF was responsible for data curation from new and published sources, data quality control and analyses, manuscript writing and editing. SS contributed to data preprocessing, interpretation, and manuscript editing. LF contributed to several aspects of the enrichment analysis, study design, and drafting of the manuscript. ISM was responsible for preparation of CH samples for RNA sequencing, quality control, and manuscript editing. MIC and CK was responsible for comparisons to single-nucleus datasets and contributed to manuscript editing. JAS and DAB were responsible for ROS/MAP postmortem data collection, PLDJ was responsible for ROS/MAP RNA sequencing data generation, and all three authors contributed to elements of study design and manuscript editing. GT and RM were responsible for CH postmortem data ascertainment. GT was responsible for study design, analytics design, and drafting of the manuscript. All authors read and approved the final manuscript.

Data and materials availability

The datasets supporting the conclusions of this article (ROS/MAP and Mayo) are available via approved access at the Synapse AMP-AD Knowledge Portal (<https://adknowledgeportal.synapse.org/>, doi: <https://doi.org/10.7303/syn2580853>). All RNA sequencing data for the Caribbean-Hispanic sample will be made available via GEO repository following publication. All analyses were performed using open-source software. No custom algorithms or software were used that are central to the research or not yet described in published literature. Open-sourced code for running all analyses and generating figures presented in the manuscript can be found on github at https://www.github.com/dfelsky/CH_LOAD_TWAS.

CRedit authorship contribution statement

Daniel Felsky: Conceptualization, Methodology, Data curation, Formal analysis, Software, Data curation, Writing – original draft, Writing – review & editing. **Ismael Santa-Maria:** Data curation, Writing – review & editing. **Mehmet Ilyas Cosacak:** Investigation, Formal analysis, Validation, Writing – review & editing. **Leon French:** Conceptualization, Formal analysis, Software. **Julie A. Schneider:** Investigation. **David A. Bennett:** Funding acquisition, Investigation. **Philip L. De Jager:** Funding acquisition, Investigation. **Caghan Kizil:** Investigation, Formal analysis, Validation, Writing – review & editing. **Giuseppe Tosto:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Writing – review & editing.

Declaration of Competing Interest

Leon French owns shares in Cortexyme Inc., a company that is developing a gingipain inhibitor to treat Alzheimer's Disease. The other authors declare no conflict of interest. Funders did not play any role in the design, analysis, or writing of this study.

Data availability

Some data used is currently available within online repositories, and some will be made available online at time of publication (as described in article).

References

- Acosta, D., Llibre-Guerra, J.J., Jiménez-Velázquez, I.Z., Llibre-Rodríguez, J.J., 2021. Dementia research in the Caribbean Hispanic Islands: present findings and future trends. *Front. Public Health* 8.

- Shigemizu, D., Mori, T., Akiyama, S., Higaki, S., Watanabe, H., Sakurai, T., Niida, S., Ozaki, K., 2020. Identification of potential blood biomarkers for early diagnosis of Alzheimer's disease through RNA sequencing analysis. *Alzheimers Res. Ther.* 12, 87. <https://doi.org/10.1186/s13195-020-00654-x>.
- Sirugo, G., Williams, S.M., Tishkoff, S.A., 2019. The missing diversity in human genetic studies. *Cell* 177, 26–31. <https://doi.org/10.1016/j.cell.2019.02.048>.
- Srinivasan, K., Friedman, B.A., Etxeberria, A., Huntley, M.A., van der Brug, M.P., Foreman, O., Paw, J.S., Modrusan, Z., Beach, T.G., Serrano, G.E., Hansen, D.V., 2020. Alzheimer's patient microglia exhibit enhanced aging and unique transcriptional activation. *Cell Rep.* 31, 107843 <https://doi.org/10.1016/j.celrep.2020.107843>.
- Stepler, K.E., Mahoney, E.R., Kofler, J., Hohman, T.J., Lopez, O.L., Robinson, R.A.S., 2020. Inclusion of African American/black adults in a pilot brain proteomics study of Alzheimer's disease. *Neurobiol. Dis.* 146, 105129 <https://doi.org/10.1016/j.nbd.2020.105129>.
- Supek, F., Bošnjak, M., Škunca, N., Šmuc, T., 2011. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* 6, e21800. <https://doi.org/10.1371/journal.pone.0021800>.
- Tang, J., Saito, T., 2017. The role of Nephronectin on proliferation and differentiation in human dental pulp stem cells. *Stem Cells Int.* 2017, e2546261 <https://doi.org/10.1155/2017/2546261>.
- Tosto, G., Fu, H., Vardarajan, B.N., Lee, J.H., Cheng, R., Reyes-Dumeyer, D., Lantigua, R., Medrano, M., Jimenez-Velazquez, I.Z., Elkind, M.S.V., Wright, C.B., Sacco, R.L., Pericak-Vance, M., Farrer, L., Rogaeva, E., St George-Hyslop, P., Reitz, C., Mayeux, R., 2015a. F-box/LRR-repeat protein 7 is genetically associated with Alzheimer's disease. *Ann. Clin. Transl. Neurol.* 2, 810–820. <https://doi.org/10.1002/acn3.223>.
- Tosto, G., Lee, J.H., Vardarajan, B.N., Cheng, R., Reyes-Dumeyer, D., Barral, S., Reitz, C., Mayeux, R., 2015b. O3-05-04: admixture analysis of Alzheimer's disease in caribbean hispanics. *Alzheimers Dement.* 11, P229. <https://doi.org/10.1016/j.jalz.2015.07.264>.
- Vanzulli, I., Papanikolaou, M., De-La-Rocha, I.C., Pieropan, F., Rivera, A.D., Gomez-Nicola, D., Verkhratsky, A., Rodríguez, J.J., Butt, A.M., 2020. Disruption of oligodendrocyte progenitor cells is an early sign of pathology in the triple transgenic mouse model of Alzheimer's disease. *Neurobiol. Aging* 94, 130–139. <https://doi.org/10.1016/j.neurobiolaging.2020.05.016>.
- Vonsattel, J.P., Aizawa, H., Ge, P., DiFiglia, M., McKee, A.C., MacDonald, M., Gusella, J.F., Landwehrmeyer, G.B., Bird, E.D., Richardson, E.P., 1995. An improved approach to prepare human brains for research. *J. Neuropathol. Exp. Neurol.* 54, 42–56. <https://doi.org/10.1097/00005072-199501000-00006>.
- Vonsattel, J.P.G., Del Amaya, M.P., Keller, C.E., 2008. Twenty-first century brain banking. Processing brains for research: the Columbia University methods. *Acta Neuropathol. (Berl.)* 115, 509–532. <https://doi.org/10.1007/s00401-007-0311-9>.
- Wang, Z., Xu, Q., Cai, F., Liu, X., Wu, Y., Song, W., 2019. BACE2, a conditional β -secretase, contributes to Alzheimer's disease pathogenesis. *JCI Insight* 4, 123431. <https://doi.org/10.1172/jci.insight.123431>.
- Weinberger, D.R., Dzirasa, K., Crumpton-Young, L.L., 2020. Missing in action: African ancestry brain research. *Neuron* 107, 407–411. <https://doi.org/10.1016/j.neuron.2020.07.008>.
- Weiner, J., 2020. tmod: Feature Set Enrichment Analysis for Metabolomics and Transcriptomics.
- Wharton, W., Kollhoff, A.L., Gangishetti, U., Verble, D.D., Upadhy, S., Zetterberg, H., Kumar, V., Watts, K.D., Kippels, A.J., Gearing, M., Howell, J.C., Parker, M.W., Hu, W.T., 2019. Interleukin 9 alterations linked to alzheimer disease in african americans. *Ann. Neurol.* 86, 407–418. <https://doi.org/10.1002/ana.25543>.
- Zeng, B., Bendl, J., Kosoy, R., Fullard, J.F., Hoffman, G.E., Roussos, P., 2022. Multi-ancestry eQTL meta-analysis of human brain identifies candidate causal variants for brain-related traits. *Nat. Genet.* 54, 161–169. <https://doi.org/10.1038/s41588-021-00987-9>.
- Zhang, B., Horvath, S., 2005. A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* 4, Article17. <https://doi.org/10.2202/1544-6115.1128>.