

## SUPPLEMENTAL MATERIAL

### **Multimodal single-cell/nucleus RNA-sequencing data analysis uncovers molecular networks between disease-associated microglia and astrocytes with implications for drug repurposing in Alzheimer's disease**

Jielin Xu<sup>1,#</sup>, Pengyue Zhang<sup>2,#</sup>, Yin Huang<sup>1,#</sup>, Yadi Zhou<sup>1</sup>, Yuan Hou<sup>1</sup>, Lynn Bekris<sup>1,3</sup>, Justin Lathia<sup>3,4,5</sup>, Chien-Wei Chiang<sup>6</sup>, Lang Li<sup>6</sup>, Andrew A. Pieper<sup>7-12</sup>, James B. Leverenz<sup>13</sup>, Jeffrey Cummings<sup>14</sup>, Feixiong Cheng<sup>1,3,5,\*</sup>

<sup>1</sup>Genomic Medicine Institute, Lerner Research Institute, Cleveland Clinic, Cleveland, OH 44195, USA

<sup>2</sup>Department of Biostatistics, School of Medicine, Indiana University, Indianapolis, IN 46202, USA

<sup>3</sup>Department of Molecular Medicine, Cleveland Clinic Lerner College of Medicine, Case Western Reserve University, Cleveland, OH 44195, USA

<sup>4</sup>Case Comprehensive Cancer Center, Case Western Reserve University School of Medicine, Cleveland, Ohio 44106, USA

<sup>5</sup>Department of Cardiovascular & Metabolic Sciences, Lerner Research Institute, Cleveland Clinic, Cleveland, OH 44195, USA

<sup>6</sup>Department of Biomedical Informatics, College of Medicine, Ohio State University, Columbus, OH 43210, USA

<sup>7</sup>Harrington Discovery Institute, University Hospitals Cleveland Medical Center, Cleveland, OH 44106, USA

<sup>8</sup>Department of Psychiatry, Case Western Reserve University, Cleveland, OH 44106, USA

<sup>9</sup>Geriatric Psychiatry, GRECC, Louis Stokes Cleveland VA Medical Center; Cleveland, OH 44106, USA

<sup>10</sup>Institute for Transformative Molecular Medicine, School of Medicine, Case Western Reserve University, Cleveland 44106, OH, USA

<sup>11</sup>Weill Cornell Autism Research Program, Weill Cornell Medicine of Cornell University, NY, NY 10065, USA

<sup>12</sup>Department of Neuroscience, Case Western Reserve University, School of Medicine, Cleveland, OH 44106, USA

<sup>13</sup>Lou Ruvo Center for Brain Health, Neurological Institute, Cleveland Clinic, Cleveland, OH 44195, USA

<sup>14</sup>Chambers-Grundy Center for Transformative Neuroscience, Department of Brain Health, School of Integrated Health Sciences, University of Nevada Las Vegas, Las Vegas, NV 89154, USA.

#Co-first authors

Correspondence to Feixiong Cheng, Ph.D.

Lerner Research Institute, Cleveland Clinic

Tel: +1-216-444-7654; Fax: +1-216-636-0009

Email: [chengf@ccf.org](mailto:chengf@ccf.org)

# TABLES OF CONTENTS

## I. SUPPLEMENTAL TABLES

Supplemental Table S1. Summary of single-cell/nucleus RNA-sequencing datasets used in this study.....	Page 11
Supplemental Table S2. Numbers of nuclei in disease associated microglia (DAM) and homeostasis associated microglia (HAM) in WT and 5XFAD mice (GSE140511). .....	Page 12
Supplemental Table S3. Numbers of cells in disease associated microglia (DAM) and homeostasis associated microglia (HAM) in WT and 5XFAD mice (GSE98969).....	Page 13
Supplemental Table S4. Disease molecular network results for mouse snRNA-seq dataset GSE140511. (.xlsx)	
Supplemental Table S5. Disease molecular network results for mouse scRNA-seq dataset GSE98969. (.xlsx)	
Supplemental Table S6. Numbers of nuclei in disease associated astrocytes (DAAs) and non-DAAs in WT and 5XFAD mice (GSE143758).....	Page 14
Supplemental Table S7. Disease molecular network results for mouse snRNA-seq dataset GSE143758. (.xlsx)	
Supplemental Table S8. Disease molecular network results for human snRNA-seq dataset GSE138852. (.xlsx)	
Supplemental Table S9. Disease molecular network results for human snRNA-seq dataset GSE147528. (.xlsx)	
Supplemental Table S10. Numbers of nuclei in disease associated astrocytes (DAAs) and non-DAAs in human AD brains entorhinal cortex (EC) region with Braak stages 0, 2, and 6 (GSE147528).....	Page 15
Supplemental Table S11. Numbers of nuclei in disease associated astrocytes (DAAs) and non-DAAs in human AD brains superior frontal gyrus (SFG) region with Braak stages 0, 2, and 6 (GSE147528).....	Page 16
Supplemental Table S12. Numbers of nuclei in disease associated astrocytes (DAAs) and non-DAAs in human AD brains entorhinal cortex (EC) region with AD (Braak stage VI) and control (GSE138852).....	Page 17
Supplemental Table S13. Network proximity evaluation between different molecular networks (disease-associated microglia and astrocytes) or gene sets. (.xlsx)	

**Supplemental Table S14. Network Analysis of metabolites-enzyme associations triggering molecular networks between microglia and astrocyte. (.xlsx)**

**Supplemental Table S15. Gene set enrichment analysis (GSEA) results of network-based discovery of repurposable drug candidates for potential AD treatments. (.xlsx)**

**Supplemental Table S16. Phenotype definitions for Alzheimer’s disease, Type 2 diabetes, Hypertension and Coronary artery disease..... Pages 18-20**

## **II. SUPPLEMENTAL FIGURES**

**Supplemental Figure S1. Discovery of disease-associated microglia (DAM) specific molecular networks using a scRNA-seq dataset (GSE98969)..... Page 21**

**Supplemental Figure S2. Nucleus / cell distributions in different immune cell subtypes in both AD mouse model and human AD brain samples..... Page 22**

**Supplemental Figure S3. Differentially expressed genes (DEGs) and pathway enrichment analyses for disease associated microglia (DAM) in 2 AD mouse model datasets..... Page 23**

**Supplemental Figure S4. Differentially expressed genes and pathway enrichment analysis for disease associated astrocytes (DAAs) built from the AD transgenic mouse model (GSE143758) ..... Page 24**

**Supplemental Figure S5. Differentially expressed genes and pathway enrichment analyses for disease associated astrocytes (DAAs) built from human AD patient snRNA-seq data (GSE147528)..... Page 25**

**Supplemental Figure S6. Expression levels (UMAP plots) of human DAA marker genes..... Page 26**

**Supplemental Figure S7. Differentially expressed genes and pathway enrichment analysis for disease associated astrocytes (DAAs) built from 6 human AD brains and 6 healthy controls (GSE138852)..... Page 27**

**Supplemental Figure S8. Network visualization and pathway enrichment analysis for disease associated astrocyte (DAA) and disease associated microglia (DAM)..... Page 28**

**I. SUPPLEMENTAL REFERENCES..... Pages 29 - 32**

## Supplemental Methods and Materials

### Bioinformatics analysis of single cell/nucleus RNA-sequencing data

All bioinformatic analyses were performed with Seurat (Butler et al. 2018) (v3.1.5), scran (Lun et al. 2016) (v1.16.0), and scater (McCarthy et al. 2017) (v1.16.1) packages in R (R Core Team, 2020). Data were normalized using a scaling factor of 10,000 and all differential expression analyses are conducted by function *FindMarkers* in Seurat (Butler et al. 2018) R package with parameter *test.use* = 'MAST'. The detailed data analysis steps for each dataset (GSE98969, GSE140511, GSE143758, GSE147528 and GSE138852) are illustrated as below.

**GSE98969.** The data used are collected from whole brain cells of 6 months 5XFAD (n=16) and C57BL/6 (n=16) mice (Keren-Shaul et al. 2017). For quality control, cells with mitochondrial content >5% and UMIs < 500 were removed. Genes with mean expression smaller than 0.005 UMIs/cell were discarded for analysis. Data were normalized using a scaling factor of 10,000 and functions *FindIntegrationAnchors* and *IntegrateData* in Seurat (Butler et al. 2018) R package are used for batch effect correction. Principle component (PC) analysis was performed using the top 2000 most variable genes and clustering was performed using the top 40 PCs and resolution of 0.4. After identifying clusters for DAM and HAM, separately (gene markers, see **Fig. 1b** (Keren-Shaul et al. 2017)), DEGs are calculated between DAM and HAM by considering cells from 5XFAD mice.

**GSE140511.** The process for clustering different cell types are provided in the original literature (Zhou et al. 2020b). We used microglia nuclei and reproduced the clustering procedures to isolates DAM and HAM nuclei. Considering all microglia nuclei, PC analysis was performed using the top 3000 most variable genes and sub-clustering was performed using the top 10 PCs and resolution of 0.1. Again, after identifying clusters enriched in DAM and HAM

nuclei using gene markers (Keren-Shaul et al. 2017) in **Fig. 1b**, DEGs are compared between DAM and HAM by considering nuclei from 5XFAD mice.

**GSE143758.** The process for clustering different cell types are provided in the original literature (Habib et al. 2020). In this study, we used astrocyte nuclei and reproduced the clustering procedures to isolate DAA and non-DAA nuclei. Considering all astrocyte nuclei, PC analysis was performed using the top 2000 most variable genes and sub-clustering was performed using the top 10 PCs and resolution of 0.3. After identifying clusters enriched in DAA nuclei by comparing the expression pattern of marker genes (Habib et al. 2020) in **Fig. 1e** among sub-clusters. We computed DEGs between DAA and non-DAA by considering nuclei from 5XFAD mice as well.

**GSE147528.** We considered astrocyte nuclei and clustering analysis was first performed by *quickCluster* function and size factors were computed by *computeSumFactors* function with parameter *min.mean = 0.1* in scran R package. Then count matrix was normalized by the computed size factors and log-transformed by function *logNormCounts* in scater R package. Top 1000 highly variable genes were selected by functions *modelGeneVar* and *getTopHVGs* in scran R package. Functions *FindIntegrationAnchors* and *IntegrateData* in Seurat (Butler et al. 2018) R package were used for batch effect correction, and clustering was performed using the top 12 PCs and resolution of 0.2 for EC and top 10 PCs and resolution of 0.4 for SFG. After identifying clusters enriched in disease associated astrocytes (DAAs, gene markers (Leng et al. 2020) as listed in **Supplemental Fig. S5**), DEGs are compared between DAAs and non-DAAs for nuclei from both superior frontal gyrus and entorhinal cortex regions.

**GSE138852.** The process for clustering different cell types are provided in the original literature (Grubman et al. 2019). In this study, we used astrocyte nuclei and reproduced the clustering procedures to extract RNA-seq profiles from DAA and non-DAA nuclei. Considering all astrocyte nuclei, PC analysis was performed using the top 2000 most variable genes and

sub-clustering was performed using the top 10 PCs and resolution of 0.1. In order to identify clusters enriched in DAA nuclei, we compared the expression pattern of the selected marker genes (Leng et al. 2020) as listed in **Supplemental Fig. S5** among sub-clusters. DEGs are compared between DAAs and non-DAAs of the entorhinal cortex region.

### **Building Human Protein-protein interactome**

To build the comprehensive human interactome from the most contemporary data available, we assemble 18 commonly used PPI databases with experimental evidence and the in-house systematic human PPI we have previously utilized (Menche et al. 2015): (i) binary PPIs tested by high-throughput yeast-two-hybrid (Y2H) systems (Luck et al. 2020); (ii) kinase-substrate interactions by literature-derived low-throughput and high-throughput experiments from KinomeNetworkX (Cheng et al. 2014), Human Protein Resource Database (HPRD) (Peri 2004), PhosphoNetworks (Hu et al. 2014), PhosphositePlus (Hornbeck et al. 2015), DbPTM 3.0 and Phospho.ELM (Dinkel et al. 2011); (iii) signaling networks by literature-derived low-throughput experiments from the Signalink2.0 (Fazekas et al. 2013); (iv) binary PPIs from three-dimensional protein structures from Instruct (Meyer et al. 2013); (v) protein complexes data (~56,000 candidate interactions) identified by a robust affinity purification-mass spectrometry collected from BioPlex V2.0 (Huttlin et al. 2015); and (vi) carefully literature-curated PPIs identified by affinity purification followed by mass spectrometry from BioGRID (Chatr-aryamontri et al. 2015), PINA (Cowley et al. 2012), HPRD (Goel et al. 2012), MINT(Licata et al. 2012), IntAct (Orchard et al. 2014), and InnateDB (Breuer et al. 2013). In total, a human interactome including 351,444 PPIs connecting 17,706 unique proteins was used in this study.

### **Network analysis metabolite-enzyme associations**

We collected 155 AD-related metabolites from 12 studies (**Supplemental Table S14**) and the Human Metabolome Database (HMDB) (Wishart et al. 2018). All metabolites were identified in AD-related human samples, including brain tissues, cerebrospinal fluid, and blood. All of these results are free available in our website AlzGPS (<https://alzgps.lerner.ccf.org/>). We collected experimentally reported metabolite-enzyme associations from three commonly used data sources, including KEGG (Kanehisa et al. 2017), Recon3D (Brunk et al. 2018), and HMDB (Wishart et al. 2018), and assembled them with the human PPI network. The updated network contains 373,320 links connecting with 17,826 unique proteins (including metabolic enzymes) and 1,419 metabolites. Then we mapped 240 DAM and DAA disease module genes and the 155 AD-related metabolites to the network and computed the maximal subgraph: (1) we found 624 unique nodes which were the first or second order neighbors of 64 DAM and DAA immune genes; (2) we obtained 73 metabolites by considering the intersection of 624 unique nodes and 155 AD-related metabolites; and (3) a subnetwork connecting 236 enzymes and 30 metabolites was generated. Finally, we computed the network paths connecting the DAM and DAA gene products on the network as well as the betweenness centrality of each node.

### **Connectivity Map (CMap) database**

The CMap database used in this study contains 6,100 expression profiles relating 1,309 compounds (Lamb et al. 2006). A parameter  $\alpha$  defined below is used to leverage the extent of differential expression for a given set of genes.

$$\alpha = \frac{t - c}{(t + c)/2} \quad (\text{S1})$$

Here  $t$  is the scaled and thresholded average difference value for the drug treatment group and  $c$  is the thresholded average difference value for the control group. Therefore, a zero  $\alpha$  value indicates no expression change after drug treatment, and a positive  $\alpha$  value means elevated

expression level after drug treatment and vice versa. Drug gene signatures with  $\alpha > 0.67$  (0.67 equals the 2-folds increment) are considered as up-regulated drug-gene pairs, and  $\alpha < -0.67$  are denoted as down-regulated drug-gene pairs.

### Gene Set Enrichment Analysis (GSEA)

We utilized GSEA algorithm to predict drugs for each cell subtype. GSEA algorithms takes two inputs: CMap database and the extracted molecular network. Detailed descriptions of GSEA have been illustrated in our recent study (Zhou et al. 2020a). To be specific, the GSEA enrichment score (ES) is calculated as shown below.

$$ES = \begin{cases} ES_{up} - ES_{down} & \text{sgn}(ES_{up}) \neq \text{sgn}(ES_{down}) \\ 0 & \text{otherwise} \end{cases} \quad (S2)$$

Both  $ES_{up}$  and  $ES_{down}$  are computed for up- and down-regulated genes in input molecular network separately with the same scheme as shown below in a 2-step manner. We first compute intermediate parameters  $a$  and  $b$ :

$$a = \max_{1 \leq j \leq s} \left( \frac{j}{s} - \frac{V(j)}{r} \right) \quad b = \max_{1 \leq j \leq s} \left( \frac{V(j)}{r} - \frac{j-1}{s} \right) \quad (S3)$$

where  $j = 1, 2, \dots, s$  were the gene sets from molecular network sorted in ascending order by their rank in the gene profiles of the drug being evaluated. The rank of gene  $j$  is denoted by  $V(j)$ , where  $1 \leq V(j) \leq r$ , with  $r$  being the number of genes (12,849) from the drug profile. Then, the corresponding  $ES_{up}$  and  $ES_{down}$  equal:

$$ES_{up} = \begin{cases} a_{up} & \text{if } a_{up} > b_{up} \\ -b_{up} & \text{if } b_{up} > a_{up} \end{cases} \quad ES_{down} = \begin{cases} a_{down} & \text{if } a_{down} > b_{down} \\ -b_{down} & \text{if } b_{down} > a_{down} \end{cases} \quad (S4)$$

In the above equations,  $a_{up/down}$  and  $b_{up/down}$  are computed with respect to up- and down-regulated genes in molecular network, separately. The GSEA ES represents drug potential capability to reverse the expression of the input molecular network. Permutation tests repeated 100 times using randomly gene lists consisting of the same numbers of up- and down-regulated

genes as the input molecular network were performed to leverage the significance of the computed ES value. The corresponding false positive rate (FDR) values for drug-disease scores are computed with 'qvalue' (Dabney et al. 2013) R package. Therefore, drugs with large positive ES value and significant FDR ( $q < 0.05$ ) were selected.

### **Pharmacoepidemiologic validation**

**Study cohorts.** We used the MarketScan Medicare Claims database from 2012 to 2017 for the pharmacoepidemiologic analysis. This dataset included individual-level procedure codes, diagnosis codes, and pharmacy claim data for 7.23 million patients. Pharmacy prescriptions of fluticasone and mometasone were identified by using RxNorm and National Drug Code (NDC).

**Outcome measurement.** For an individual exposed to the aforementioned drugs, a drug episode was defined as from drug initiation to drug discontinuation. Specifically, drug initiation was defined as the first day of drug supply (i.e., 1st prescription date). Drug discontinuation is defined as the last day of drug supply (i.e., last prescription date + days of supply) and without drug supply for the next 60 days. The fluticasone cohort included the first fluticasone episode for each individual, as well as the mometasone cohort. Further, we excluded observations that started within 180-day of insurance enrollment. For the extracted cohorts, demographic variables including age, gender and geographical location were collected. Additionally, diagnoses of hypertension (HT), coronary artery disease (CAD), and type 2 diabetes (T2D) (the International Classification of Disease [ICD], **Supplemental Table S16**) codes before drug initiation were collected, to address potential confounding biases. Last, a control cohort was selected from patients who were not exposed to fluticasone. Specifically, non-exposures were matched to the exposures (ratio 4:1) by initiation time of fluticasone, enrollment history, gender and comorbidities (T2D, CAD and HT). The outcome defined by ICD codes was time from drug initiation to diagnosis of AD. For the fluticasone and mometasone cohorts, observations without

diagnosis of AD were censored at the end of drug episodes. For the control cohort, the corresponding fluticasone episode starting date was used as the starting time. Observations without diagnosis of AD were censored at the corresponding fluticasone episode's end date.

**Propensity score estimation.** We define NE = north east, NC = north central, S = south, W = west, T2D = type 2 diabetes, HT = hypertension and CAD = coronary artery disease. The propensity score of taking fluticasone vs. a comparator drug was estimated by the following logistic regression model:

$$\begin{aligned} \text{logit}[\text{Pr}(\text{Drug} = \text{fluticasone})] = & \beta_0 + \beta_1 \text{Age} + \beta_2 \text{Gender} + \beta_3 1(\text{Location} = \text{NE}) + \\ & \beta_4 1(\text{Location} = \text{NC}) + \beta_5 1(\text{Location} = \text{S}) + \beta_6 \text{T2D} + \beta_7 \text{HT} + \beta_8 \text{CAD}. \end{aligned} \quad (\text{S5})$$

**Statistical analysis.** The survival curves for time to AD were estimated using a Kaplan-Meier estimator approach. We used the large number of covariates generated throughout the process to address clinical scenarios evaluated in each study. Additionally, propensity score stratified survival analyses were conducted to investigate the risk of AD between fluticasone users and non-fluticasone users, as well as fluticasone users and mometasone users. Specifically, for each comparison, the propensity score of taking fluticasone was estimated by using a logistic regression model, in which the covariates included age, gender, geographical location, T2D diagnosis and HT diagnosis. Further, propensity score stratified Cox-proportional hazards models were used to conduct statistical inference for the hazard ratios (HR) of developing AD between cohorts.

**Supplemental Table S1:** Summary of single-cell/nucleus RNA-sequencing datasets used in this study.

	GSE98989	GSE140511	GSE143758	GSE147528	GSE138852
Organism	Mus musculus	Mus musculus	Mus musculus	Homo sapiens	Homo sapiens
Brain region	whole brain	whole brain	hippocampus	entorhinal cortex + superior frontal gyrus	entorhinal cortex
Focused Cell Type	microglia	microglia	astrocyte	astrocyte	astrocyte
Batch Effect Correction	Yes (with Seurat Integration tool)	NA (do not provide sample-batch information in data)	Yes (Combat function from 'sva' (Leek et al. 2012) package in R)	Yes (with Seurat Integration tool)	NA (do not provide sample-batch information in data)
Cell / Nucleus Num (Focused Cell Type)	8,277	4,389	7,748	5,599 (EC), 8,348 (SFG)	2,119

**Supplemental Table S2:** Numbers of nuclei in disease associated microglia (DAM) and homeostasis associated microglia (HAM) in WT and 5XFAD mice (GSE140511).

	5XFAD1 (GSM4173510)	5XFAD2 (GSM4173511)	5XFAD3 (GSM4173512)	WT1 (GSM4173504)	WT2 (GSM4173505)	WT3 (GSM4173506)
Nucleus Number						
HAM	272	96	298	312	205	213
DAM	214	151	421	33	12	14

Unpaired *t*-test 1:

Null hypothesis: means of DAM nucleus numbers in mouse AD and WT samples are equal.

Alternative hypothesis: mean of DAM nucleus numbers in mouse AD samples is GREATER than that in mouse WT samples.

Result:  $t = 2.96$ ,  $p\text{-value} = 0.048$

Unpaired *t*-test 2:

Null hypothesis: means of HAM nucleus numbers in mouse AD and WT samples are equal.

Alternative hypothesis: means of HAM nucleus numbers in mouse AD and WT samples are NOT equal.

Result:  $t = -0.296$ ,  $p\text{-value} = 0.786$

**Supplemental Table S3:** Numbers of cells in disease associated microglia (DAM) and homeostasis associated microglia (HAM) in wild-type (WT) and 5XFAD mice (GSE98969).

Cell Number						
	AD_P1-1 (GSM2629341)	AD_P1-2 (GSM2629406)	AD_P1-3 (GSM2629412)	WT_P1-1 (GSM2629343)	WT_P1-2 (GSM2629418)	WT_P1-3 (GSM2629424)
HAM	152	173	166	219	208	207
DAM	13	19	36	1	0	0
	AD_P2-1 (GSM2629344)	AD_P2-2 (GSM2629407)	AD_P2-3 (GSM2629413)	WT_P2-1 (GSM2629342)	WT_P2-2 (GSM2629419)	WT_P2-3 (GSM2629425)
HAM	177	193	195	210	268	274
DAM	13	38	27	3	3	3
	AD_P3-1 (GSM2629347)	AD_P3-2 (GSM2629408)	AD_P3-3 (GSM2629414)	WT_P3-1 (GSM2629345)	WT_P3-2 (GSM2629420)	WT_P3-3 (GSM2629426)
HAM	181	209	185	215	233	262
DAM	17	19	25	4	1	4
	AD_P4-1 (GSM2629348)	AD_P4-2 (GSM2629409)	AD_P4-3 (GSM2629415)	WT_P4-1 (GSM2629346)	WT_P4-2 (GSM2629421)	WT_P4-3 (GSM2629427)
HAM	184	225	201	213	237	267
DAM	25	19	29	2	1	1
		AD_P5-2 (GSM2629410)	AD_P5-3 (GSM2629416)		WT_P5-2 (GSM2629422)	WT_P5-3 (GSM2629428)
HAM		174	155		239	206
DAM		22	29		1	0
		AD_P6-2 (GSM2629411)	AD_P6-3 (GSM2629417)		WT_P6-2 (GSM2629423)	WT_P6-3 (GSM2629429)
HAM		147	139		219	240
DAM		26	28		2	2

Unpaired *t*-test 1:

Null hypothesis: means of DAM cell numbers in mouse AD and WT samples are equal.

Alternative hypothesis: mean of DAM cell numbers in mouse AD samples is GREATER than that in mouse WT samples.

Result:  $t = 12.1$ ,  $p\text{-value} = 9.11 \times 10^{-10}$

Unpaired *t*-test 2:

Null hypothesis: means of HAM cell numbers in mouse AD and WT samples are equal.

Alternative hypothesis: mean of HAM cell numbers in mouse AD samples is LESS than that in mouse WT samples.

Result:  $t = -6.42$ ,  $p\text{-value} = 2.16 \times 10^{-17}$

**Supplemental Table S6:** Numbers of nuclei in disease associated astrocytes (DAAs) and non-DAAs in WT and 5XFAD mice (GSE143758). The sample IDs mentioned in below matches with those from Supplementary Table S1 in the original literatures.

Nucleus Number						
	Batch1_AD1 (HipR-AD- G3-2w)	Batch1_AD2 (HipR-AD- G1-4w)	Batch1_WT1(HipR- WT-G3-2w)	Batch1_WT2 (HipR-WT- G1-4w)		
non-DAA	1,104	1,143	1,320	1,275		
DAA	329	261	25	20		
	Batch3_AD1 (Untreated- Hip-S2-R)	Batch3_AD2 (Untreated- Hip-S2-L)	Batch3_AD3 (Untreated-Hip-S1- L)	Batch3_WT1 (Wt-Hip-S2- R)	Batch3_WT2 (Wt-Hip-S2-L)	Batch3_WT3 (Wt-Hip-S1-L)
non-DAA	122	129	291	450	336	551
DAA	82	68	215	8	3	16

Paired *t*-test 1:

Null hypothesis: means of DAA nucleus numbers in mouse AD and WT samples are equal.

Alternative hypothesis: mean of DAA nucleus numbers in mouse AD samples is GREATER than that in mouse WT samples.

Result:  $t = 3.77$ ,  $p\text{-value} = 0.0098$

Paired *t*-test 2:

Null hypothesis: means of non-DAA nucleus numbers in mouse AD and WT samples are equal.

Alternative hypothesis: mean of non-DAA nucleus numbers in mouse AD samples is LESS than that in mouse WT samples.

Result:  $t = -7.08$ ,  $p\text{-value} = 0.001$

**Supplemental Table S10:** Numbers of nuclei in disease associated astrocytes (DAAs) and non-DAAs in human AD brains entorhinal cortex (EC) region with Braak stages 0, 2, and 6 (GSE147528).

Nucleus Number				
Braak Stage 0				
	EC1 (GSM4432646, Batch C)	EC2 (GSM4432645, Batch C)	EC3 (GSM4432647, Batch C)	
non-DAA	523	271	77	
DAA	16	62	14	
Braak Stage 2				
	EC4 (GSM4432648, Batch C)	EC5 (GSM4432651, Batch B)	EC6 (GSM4432649, Batch D)	EC7 (GSM4432650, Batch D)
non-DAA	312	341	111	523
DAA	7	1	1	137
Braak Stage 6				
	EC8 (GSM4432653, Batch B)	EC9 (GSM4432652, Batch B)	EC10 (GSM4432654, Batch D)	
non-DAA	670	1545	707	
DAA	245	23	13	

One-way ANOVA test 1:

Null hypothesis: means of DAA nucleus numbers are same across different Braak stages (0,2,6) human brain EC samples.

Alternative hypothesis: at least one pairs (Braak stages) of means (DAA nucleus numbers) are different from each other in human brain EC samples.

Result:  $F = 0.53$ ,  $p\text{-value} = 0.608$

One-way ANOVA test 2:

Null hypothesis: means of non-DAA nucleus numbers are same across different Braak stages (0,2,6) human brain EC samples.

Alternative hypothesis: at least one pairs (Braak stages) of means (non-DAA nucleus numbers) are different from each other in human brain EC samples.

Result:  $F = 4.83$ ,  $p\text{-value} = 0.048$

**Supplemental Table S11:** Numbers of nuclei in disease associated astrocytes (DAAs) and non-DAAs in human AD brains superior frontal gyrus (SFG) region with Braak stages 0, 2, and 6 (GSE147528).

Nucleus Number				
Braak Stage 0				
	SFG1 (GSM4432636, Batch C)	SFG2 (GSM4432635, Batch C)	SFG3 (GSM4432637, Batch C)	
non-DAA	585	633	445	
DAA	13	117	47	
Braak Stage 2				
	SFG4 (GSM4432638, Batch C)	SFG5 (GSM4432641, Batch A)	SFG6 (GSM4432639, Batch D)	SFG7 (GSM4432640, Batch D)
non-DAA	921	552	784	357
DAA	56	105	56	253
Braak Stage 6				
	SFG8 (GSM4432643, Batch B)	SFG9 (GSM4432642, Batch A)	SFG10 (GSM4432644, Batch D)	
non-DAA	1411	552	940	
DAA	107	236	178	

One-way ANOVA test 1:

Null hypothesis: means of DAA nucleus numbers are same across different Braak stages (0,2,6) human brain SFG samples.

Alternative hypothesis: at least one pairs (Braak stages) of means (DAA nucleus numbers) are different from each other in human brain SFG samples.

Result:  $F = 1.72$ ,  $p\text{-value} = 0.246$

One-way ANOVA test 2:

Null hypothesis: means of non-DAA nucleus numbers are same across different Braak stages (0,2,6) human brain SFG samples.

Alternative hypothesis: at least one pairs (Braak stages) of means (non-DAA nucleus numbers) are different from each other in human brain SFG samples.

Result:  $F = 1.73$ ,  $p\text{-value} = 0.246$

**Supplemental Table S12:** Numbers of nuclei in disease associated astrocytes (DAAs) and non-DAAs in human AD brains entorhinal cortex (EC) region with AD (Braak stage VI) and control (GSE138852).

Nucleus Number						
	AD1-AD2 (GSM4120429)	AD3-AD4 (GSM4120424)	AD5-AD6 (GSM4120423)	Ct1-Ct2 (GSM4120427)	Ct3-Ct4 (GSM4120426)	Ct5-Ct6 (GSM4120425)
non-DAA	1	15	144	149	706	623
DAA	0	229	2	28	155	67

Unpaired *t*-test 1:

Null hypothesis: means of DAA nucleus numbers in human AD and control brain samples are equal.

Alternative hypothesis: mean of DAA nucleus numbers in human AD brain samples is GREATER than that in human control brain samples.

Result:  $t = -0.07$ ,  $p\text{-value} = 0.527$

Unpaired *t*-test 2:

Null hypothesis: means of non-DAA nucleus numbers in human AD and control brain samples are equal.

Alternative hypothesis: mean of non-DAA nucleus numbers in human AD brain samples is LESS than that in human control brain samples.

Result:  $t = -2.45$ ,  $p\text{-value} = 0.059$

**Supplemental Table S16:** Phenotype definitions for Alzheimer's disease, Type 2 diabetes, Hypertension and Coronary artery disease.

---

**Alzheimer's disease** (Wei et al. 2016; Wilkinson et al. 2018)

3310 (Alzheimer's disease), F00 (Dementia in Alzheimer disease), F000A (Dementia in Alzheimer disease with early onset), F001A (Dementia in Alzheimer disease with late onset), F002A (Dementia in Alzheimer disease, atypical or mixed type), F009A (Dementia in Alzheimer disease, unspecified), G30 (Alzheimer's disease), G300 (Alzheimer disease with early onset), G301 (Alzheimer disease with late onset), G308 (Other Alzheimer disease), G309 (Alzheimer disease, unspecified)

**Type 2 diabetes** (Kho et al. 2012; Wei et al. 2012)

25000 (Diabetes mellitus without mention of complication, type II or unspecified type, not stated as uncontrolled), 25050 (Diabetes with ophthalmic manifestations, type II or unspecified type, not stated as uncontrolled), 25002 (Diabetes mellitus without mention of complication, type II or unspecified type, uncontrolled), 25052 (Diabetes with ophthalmic manifestations, type II or unspecified type, uncontrolled), 25010 (Diabetes with ketoacidosis, type II or unspecified type, not stated as uncontrolled), 25060 (Diabetes with neurological manifestations, type II or unspecified type, not stated as uncontrolled), 25012 (Diabetes with ketoacidosis, type II or unspecified type, uncontrolled), 25062 (Diabetes with neurological manifestations, type II or unspecified type, uncontrolled), 25020 (Diabetes with hyperosmolarity, type II or unspecified type, not stated as uncontrolled), 25070 (Diabetes with peripheral circulatory disorders, type II or unspecified type, not stated as uncontrolled), 25022 (Diabetes with hyperosmolarity, type II or unspecified type, uncontrolled), 25072 (Diabetes with peripheral circulatory disorders, type II or unspecified type, uncontrolled), 25030 (Diabetes with other coma, type II or unspecified type, not stated as uncontrolled), 25080 (Diabetes with other specified manifestations, type II or unspecified type, not stated as uncontrolled), 25032 (Diabetes with other coma, type II or unspecified type, uncontrolled), 25082 (Diabetes with other specified manifestations, type II or unspecified type, uncontrolled), 25040 (Diabetes with renal manifestations, type II or unspecified type, not stated as uncontrolled), 25090 (Diabetes with unspecified complication, type II or unspecified type, not stated as uncontrolled), 25042 (Diabetes with renal manifestations, type II or unspecified type, uncontrolled), 25092 (Diabetes with unspecified complication, type II or unspecified type, uncontrolled), E089 (Diabetes mellitus due to underlying condition without complications), E1100 (Type 2 diabetes mellitus with hyperosmolarity without nonketotic hyperglycemic-hyperosmolar coma), E1101 (Type 2 diabetes mellitus with hyperosmolarity with coma), E1110 (Type 2 diabetes mellitus with ketoacidosis without coma), E1121 (Type 2 diabetes mellitus with diabetic nephropathy), E1122 (Type 2 diabetes mellitus with diabetic chronic kidney

disease), E1129 (Type 2 diabetes mellitus with other diabetic kidney complication), E11311 (Type 2 diabetes mellitus with unspecified diabetic retinopathy with macular edema), E11319 (Type 2 diabetes mellitus with unspecified diabetic retinopathy without macular edema), E113219 (Type 2 diabetes mellitus with mild nonproliferative diabetic retinopathy with macular edema, unspecified eye), E113291 (Type 2 diabetes mellitus with mild nonproliferative diabetic retinopathy without macular edema, right eye), E113292 (Type 2 diabetes mellitus with mild nonproliferative diabetic retinopathy without macular edema, left eye), E113293 (Type 2 diabetes mellitus with mild nonproliferative diabetic retinopathy without macular edema, bilateral), E113299 (Type 2 diabetes mellitus with mild nonproliferative diabetic retinopathy without macular edema, unspecified eye), E113319 (Type 2 diabetes mellitus with moderate nonproliferative diabetic retinopathy with macular edema, unspecified eye), E113391 (Type 2 diabetes mellitus with moderate nonproliferative diabetic retinopathy without macular edema, right eye), E113392 (Type 2 diabetes mellitus with moderate nonproliferative diabetic retinopathy without macular edema, left eye), E113393 (Type 2 diabetes mellitus with moderate nonproliferative diabetic retinopathy without macular edema, bilateral), E113399 (Type 2 diabetes mellitus with moderate nonproliferative diabetic retinopathy without macular edema, unspecified eye), E113419 (Type 2 diabetes mellitus with severe nonproliferative diabetic retinopathy with macular edema, unspecified eye), E113491 (Type 2 diabetes mellitus with severe nonproliferative diabetic retinopathy with macular edema, right eye), E113492 (Type 2 diabetes mellitus with severe nonproliferative diabetic retinopathy with macular edema, left eye), E113493 (Type 2 diabetes mellitus with severe nonproliferative diabetic retinopathy without macular edema, bilateral), E113499 (Type 2 diabetes mellitus with severe nonproliferative diabetic retinopathy without macular edema, unspecified eye), E113519 (Type 2 diabetes mellitus with proliferative diabetic retinopathy with macular edema, unspecified eye), E113591 (Type 2 diabetes mellitus with proliferative diabetic retinopathy without macular edema, right eye), E113592 (Type 2 diabetes mellitus with proliferative diabetic retinopathy without macular edema, left eye), E113593 (Type 2 diabetes mellitus with proliferative diabetic retinopathy without macular edema, bilateral), E113599 (Type 2 diabetes mellitus with proliferative diabetic retinopathy without macular edema, unspecified eye), E1136 (Type 2 diabetes mellitus with diabetic cataract), E1139 (Type 2 diabetes mellitus with other diabetic ophthalmic complication), E1140 (Type 2 diabetes mellitus with diabetic neuropathy, unspecified), E1142 (Type 2 diabetes mellitus with diabetic polyneuropathy), E1143 (Type 2 diabetes mellitus with diabetic autonomic (poly)neuropathy), E1144 (Type 2 diabetes mellitus with diabetic amyotrophy), E1149 (Type 2 diabetes mellitus with other diabetic neurological complication), E1151 (Type 2 diabetes mellitus with diabetic peripheral angiopathy without gangrene), E1152 (Type 2 diabetes mellitus with diabetic peripheral angiopathy with gangrene), E1159 (Type 2 diabetes

mellitus with other circulatory complications), E11610 (Type 2 diabetes mellitus with diabetic neuropathic arthropathy), E11618 (Type 2 diabetes mellitus with other diabetic arthropathy), E11620 (Type 2 diabetes mellitus with diabetic dermatitis), E11621 (Type 2 diabetes mellitus with foot ulcer), E11628 (Type 2 diabetes mellitus with other skin complications), E11630 (Type 2 diabetes mellitus with periodontal disease), E11641 (Type 2 diabetes mellitus with hypoglycemia with coma), E11649 (Type 2 diabetes mellitus with hypoglycemia without coma), E1165 (Type 2 diabetes mellitus with hyperglycemia), E1169 (Type 2 diabetes mellitus with other specified complication), E118 (Type 2 diabetes mellitus with unspecified complications), E119 (Type 2 diabetes mellitus without complications),

E1300 (Other specified diabetes mellitus with hyperosmolarity without nonketotic hyperglycemic-hyperosmolar coma), E1310 (Other specified diabetes mellitus with ketoacidosis without coma), E1311 (Other specified diabetes mellitus with ketoacidosis with coma), E1322 (Other specified diabetes mellitus with diabetic chronic kidney disease), E1329 (Other specified diabetes mellitus with other diabetic kidney complication), E1339 (Other specified diabetes mellitus with other diabetic ophthalmic complication), E1349 (Other specified diabetes mellitus with other diabetic neurological complication), E1351 (Other specified diabetes mellitus with diabetic peripheral angiopathy without gangrene), E13628 (Other specified diabetes mellitus with other skin complications), E1365 (Other specified diabetes mellitus with hyperglycemia), E1369 (Other specified diabetes mellitus with other specified complication), E138 (Other specified diabetes mellitus with unspecified complications), E139 (Other specified diabetes mellitus without complications)

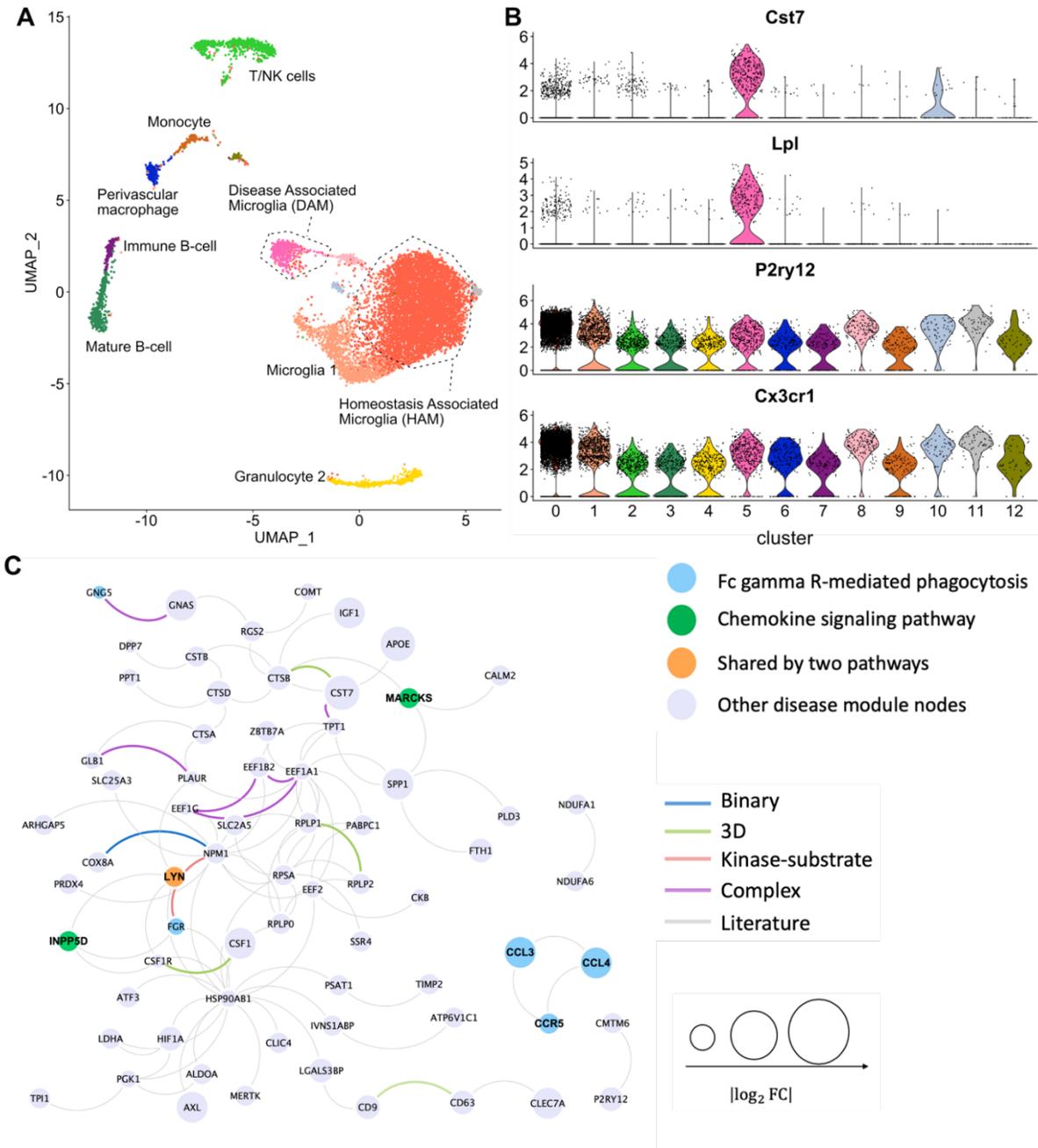
**Hypertensions** (Federman et al. 2005; Banerjee et al. 2012)

4010 (Malignant essential hypertension), 4011 (Benign essential hypertension), 4019 (Unspecified essential hypertension), I10 (Essential (Primary) Hypertension), I169 (Sequelae of cerebrovascular disease)

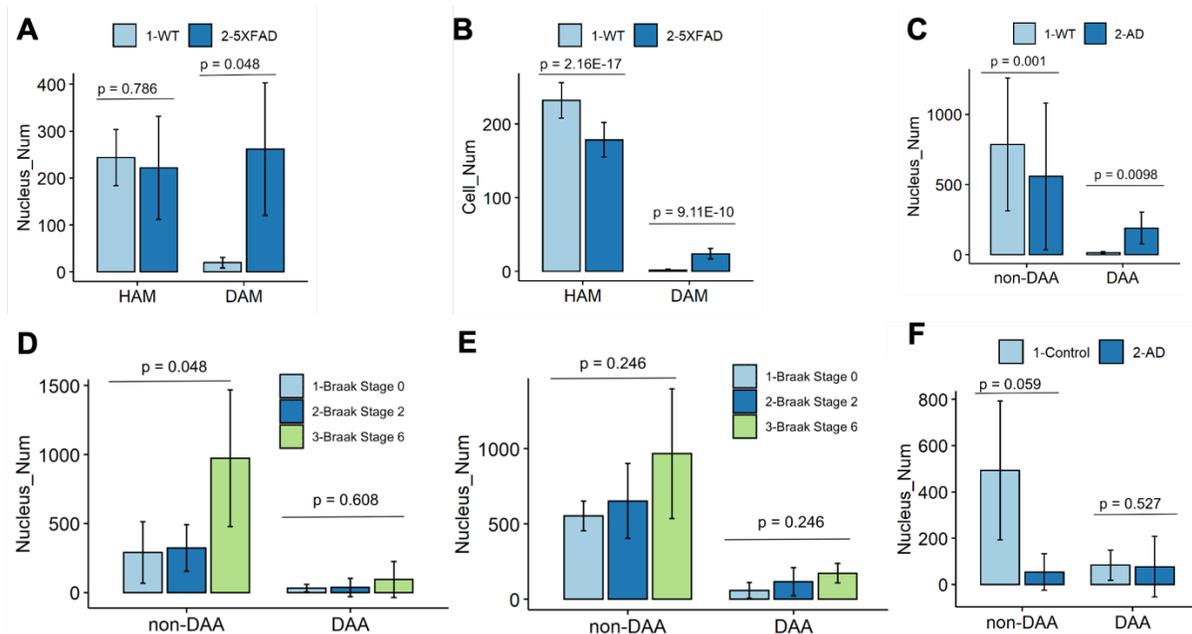
**Coronary artery disease** (Federman et al. 2005; Abul-Husn et al. 2016; Smilowitz et al. 2016)

410 (Acute myocardial infarction), 411 (Other acute and subacute forms of ischemic heart disease), 412 (Old myocardial infarction), 4131 (Prinzmetal angina), 414 (Other forms of chronic ischemic heart disease), I20 (Angina pectoris), I21 (ST elevation (STEMI) and non-ST elevation (NSTEMI) myocardial infarction), I22 (Subsequent ST elevation (STEMI) and non-ST elevation (NSTEMI) myocardial infarction), I23 (Certain current complications following ST elevation (STEMI) and non-ST elevation (NSTEMI) myocardial infarction (within the 28 day period)), I24 (Other acute ischemic heart diseases), I25 (Chronic ischemic heart disease)

---

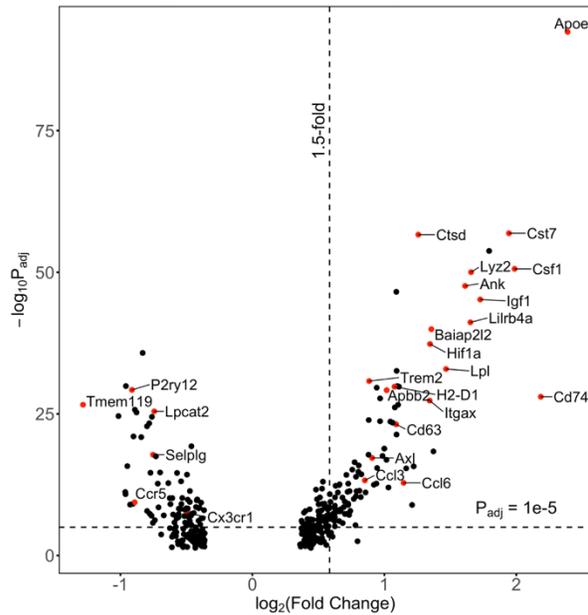


**Supplemental Figure S1:** Discovery of disease-associated microglia (DAM) specific molecular networks using a scRNA-seq dataset (GSE98969). (A) UMAP plot of clustering 10,836 CD45+ cells into 13 subgroups: red dash-circled cluster denoting the homeostasis associated microglia (HAM) and pink dash-circled cluster denoting the DAM. (B) Expression levels (stacked violin plots) of representative marker genes (up-regulation in DAM: *Cst7* and *Lpl* and down-regulation in DAM: *P2ry12* and *Cx3cr1*) in different clusters; (C) Extracted cell subtype DAM specific molecular network includes 69 nodes (proteins) and 97 edges (protein-protein interactions [PPIs]). Node sizes are proportional to their corresponding  $|\log_2 FC|$ . Node color is coded by known immune pathways from the KEGG database. Edge color is coded by experimental evidences of PPIs. Key immune modulators related to AD are highlighted by **bold** text.

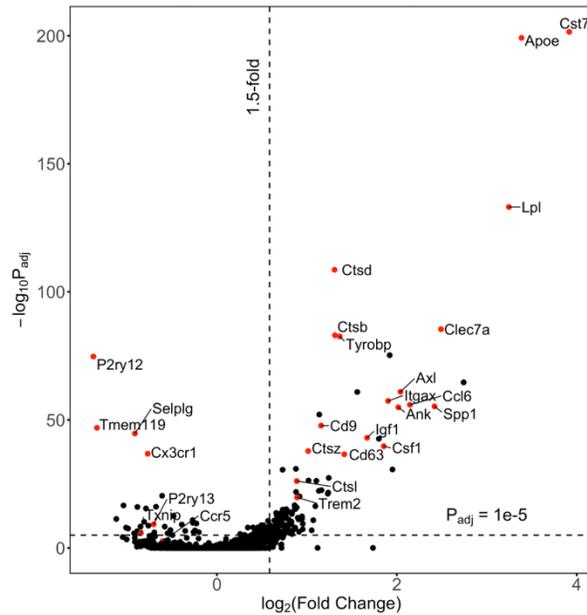


**Supplemental Figure S2:** Nucleus / cell distributions in different immune cell subtypes in both AD mouse model and human AD brain samples. Nucleus / cell abundance for homeostasis associated microglia (HAM) and disease associated microglia (DAM) clusters in both wild-type (WT) and 5XFAD mouse models (A) from snRNA-seq dataset – GSE140511 and (B) from scRNA-seq dataset – GSE98969. (C) Bar plot of nucleus abundance in both disease-associated astrocyte (DAA) and non-DAA clusters considering both WT and 5XFAD mice (GSE143758). (D-E) Bar plot of nucleus abundance in both DAA and non-DAA clusters considering human AD brains (GSE147528) with Braak stages 0, 2, and 6 with respect to 2 brain regions – (D) entorhinal cortex (EC) and superior frontal gyrus (SFG). (F) Bar plot of nucleus abundance in both disease-associated astrocyte (DAA) and non-DAA clusters considering both human AD brains and healthy controls (GSE138852). Detailed results are presented in **Supplemental Tables S2,S3,S6,S10-S12**.

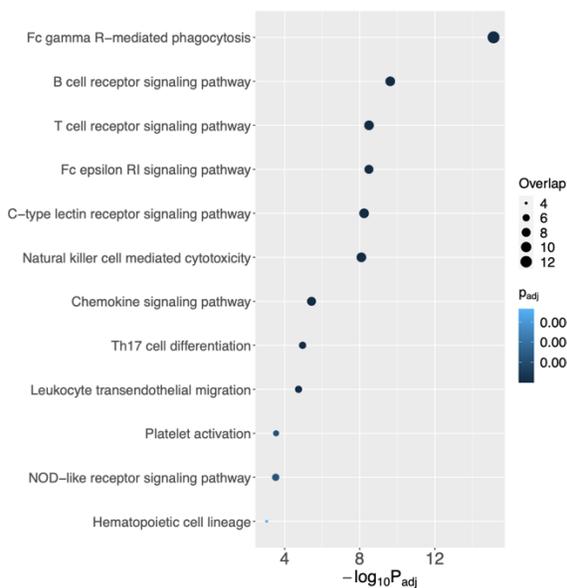
**A** Differentially Expressed Genes Analyses between HAM and DAM, Genotype = '5XFAD', sn-RNA seq: GSE140511



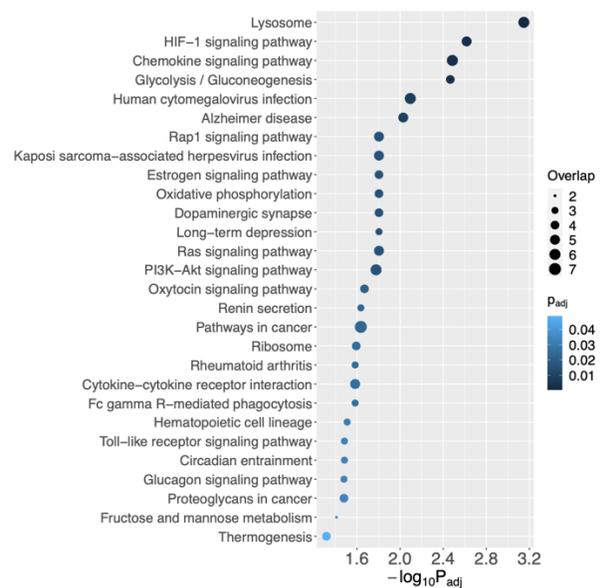
**B** Differentially Expressed Genes Analyses between HAM and DAM, Genotype = '5XFAD', sc-RNA seq: GSE98969



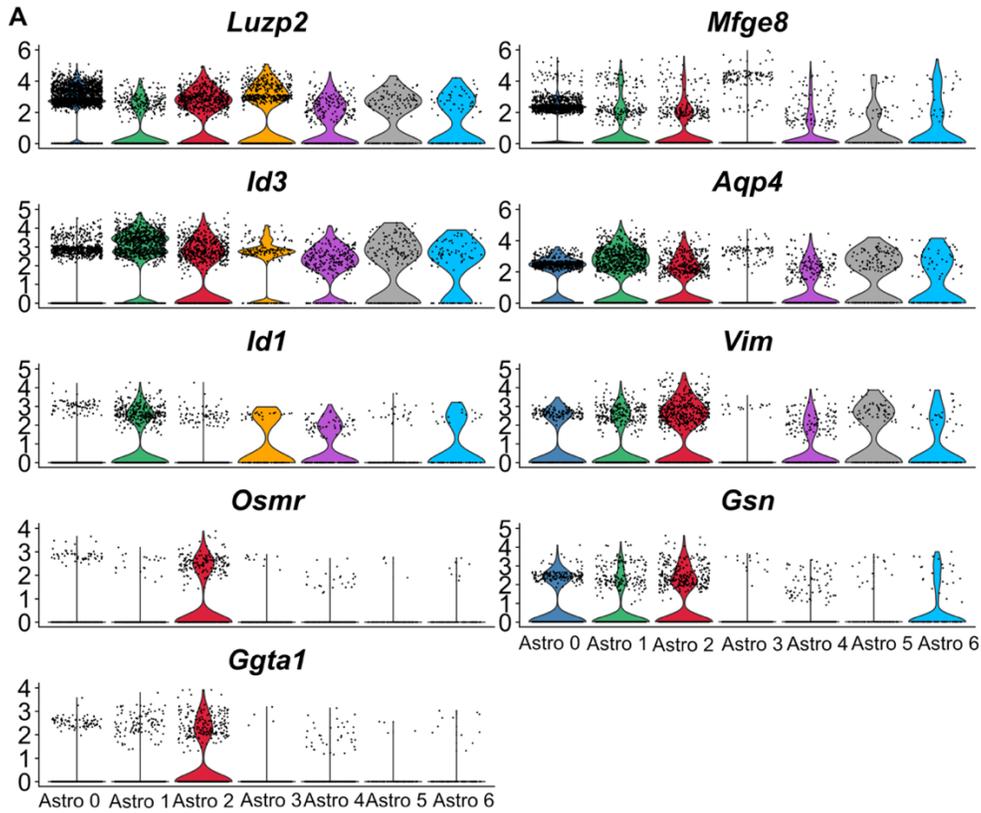
**C** KEGG Pathway Enrichment Analyses, mouse, sn-RNA seq: GSE140511, Immune pathways only



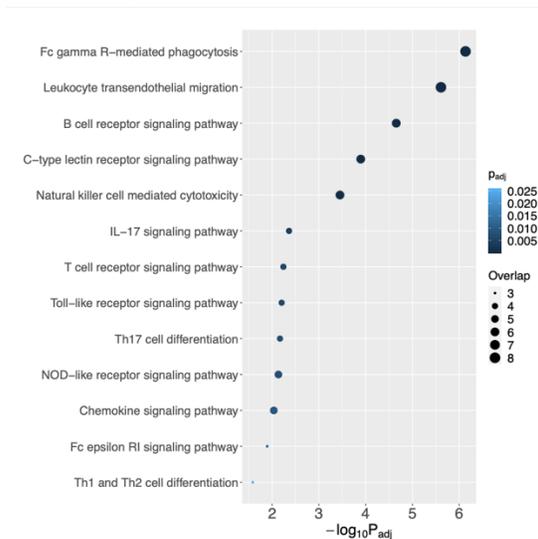
**D** KEGG Pathway Enrichment Analyses, mouse, sc-RNA seq: GSE98969



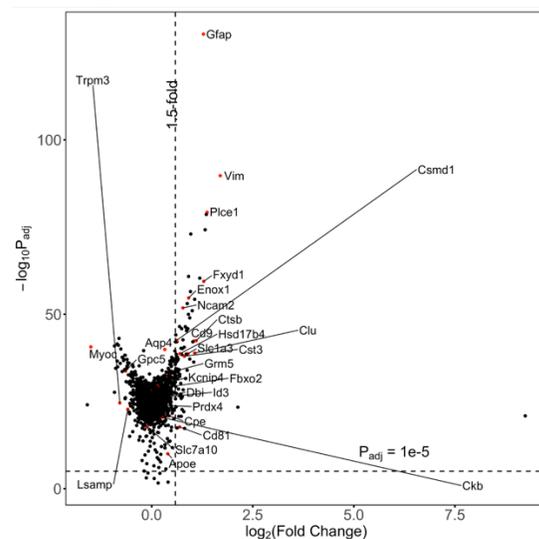
**Supplemental Figure S3:** Differentially expressed genes (DEGs) and pathway enrichment analyses for disease associated microglia (DAM) in 2 AD mouse model datasets. Differential expressed gene analyses (volcano plot) were compared between DAM and homeostasis associated microglia (HAM) in (A) GSE140511 (**Supplemental Table S4**) and (B) GSE98969 (**Supplemental Table S5**). (C) Pathway enrichment analysis (**Supplemental Table S4**) for GSE140511. (D) Pathway enrichment analysis (**Supplemental Table S5**) for GSE98969.



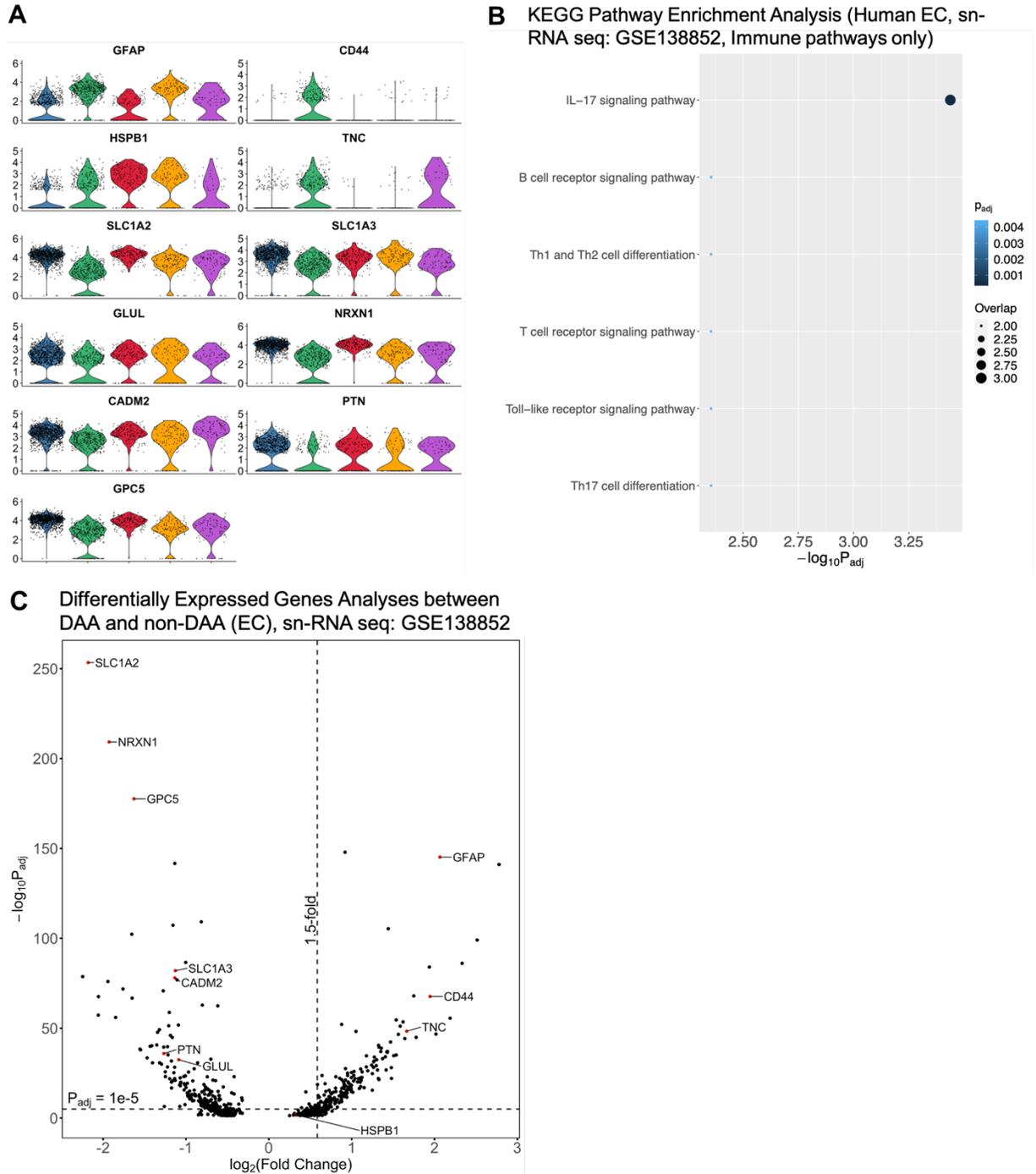
**B** KEGG Pathway Enrichment Analyses, mouse hippocampus, sn-RNA seq: GSE143758, Immune pathways only



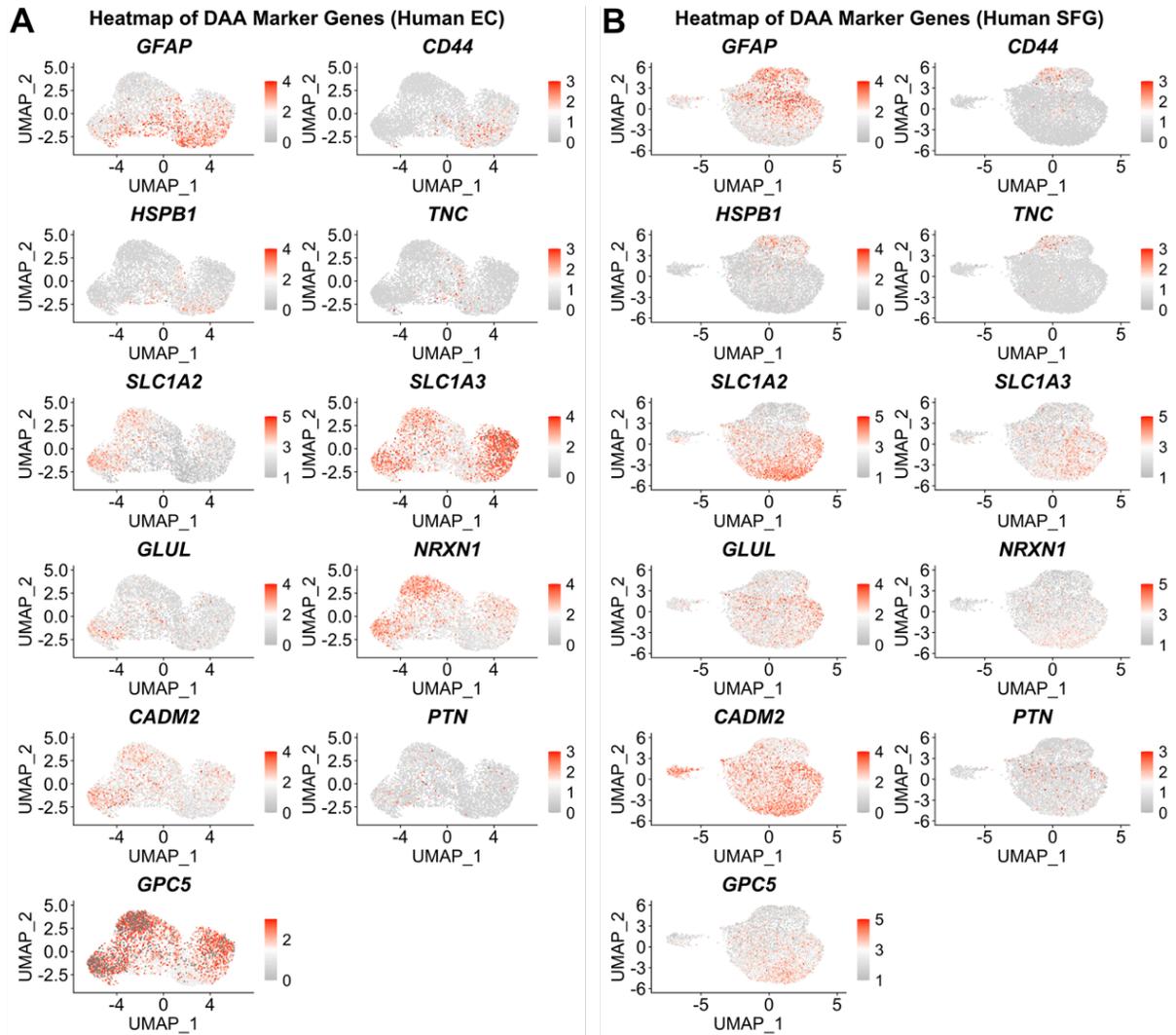
**C** Differently Expressed Genes Analyses between DAA and non-DAA, Genotype = '5XFAD', sn-RNA seq: GSE143758



**Supplemental Figure S4:** Differentially expressed genes and pathway enrichment analysis for disease associated astrocytes (DAAs) built from the AD transgenic mouse model (GSE143758). (A) Stacked violin plot displaying the expression patterns of 9 representative genes across different astrocyte sub-clusters. (B) Pathway enrichment analysis presented by 13 enriched KEGG immune system pathways (**Supplemental Table S7**). (C) Differential expressed gene analysis (volcano plot) between DAAs and non-DAAs in 5XFAD mice.

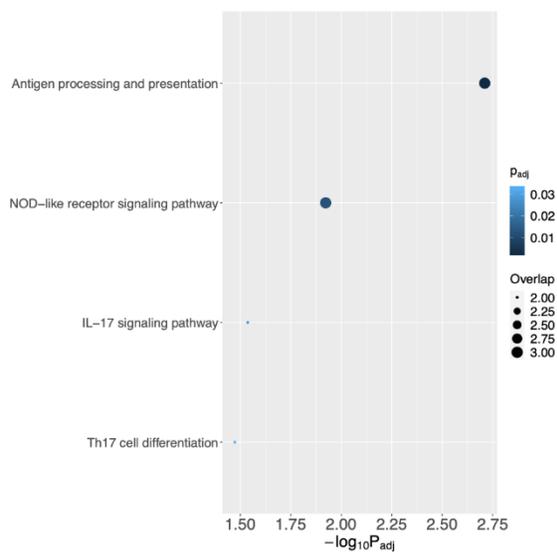


**Supplemental Figure S5:** Differentially expressed genes and pathway enrichment analysis for disease associated astrocytes (DAAs) built from 6 human AD brains and 6 healthy controls (GSE138852). (A) Stacked violin plot displaying the expression patterns of DAA marker genes across different astrocyte sub-clusters. (B) Pathway enrichment analysis presented by 6 enriched KEGG immune system pathways (**Supplemental Table S8**). (C) Differential expressed gene analysis (volcano plot) between DAAs and non-disease associated astrocytes (non-DAAs) among human brain samples.

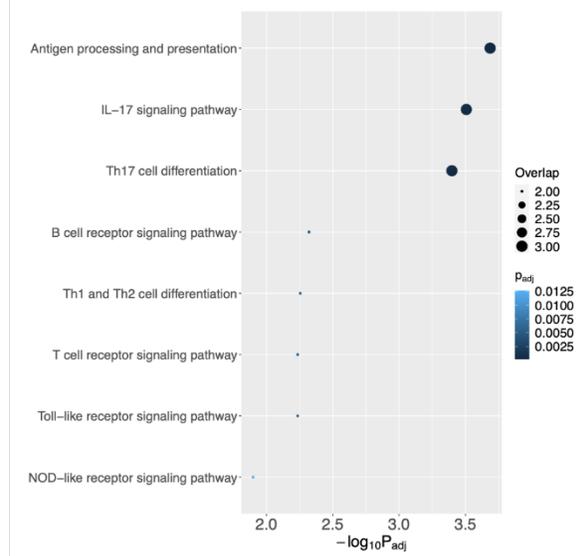


**Supplemental Figure S6:** Expression levels (UMAP plots) of human DAA marker genes (up-regulation in DAA: *GFAP*, *CD44*, *HSPB1* and *TNC*, and down-regulation in DAA: *SLC1A2*, *SLC1A3*, *GLUL*, *NRXN1*, *CADM2*, *PTN* and *GPC5*) in all astrocyte sub-clusters with respect to AD human brains. (A) brain region: entorhinal cortex (EC) and (B) brain region: superior frontal gyrus (SFG). Data source: GSE147528.

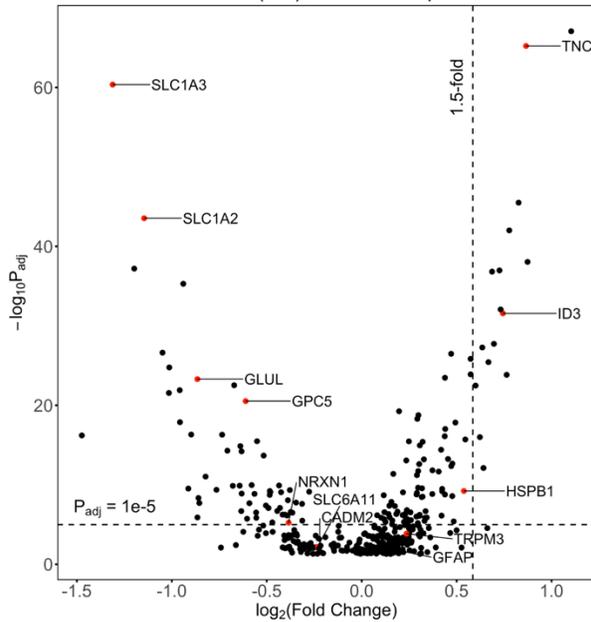
**A** KEGG Pathway Enrichment Analysis (Human EC, sn-RNA seq: GSE147528, Immune pathways only)



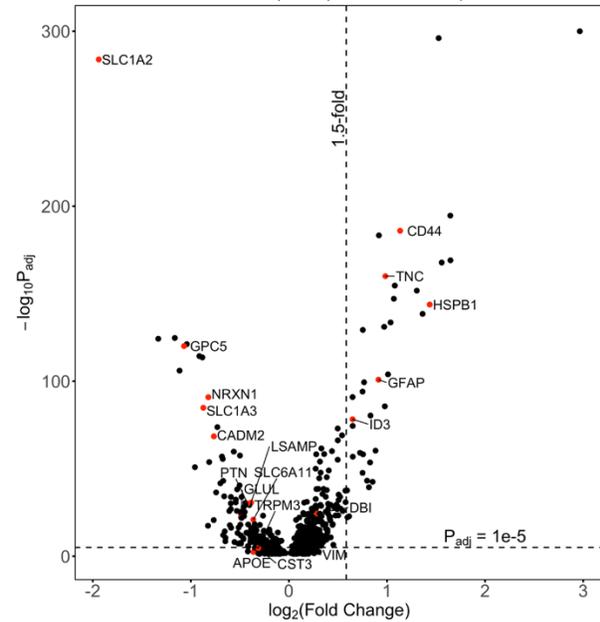
**B** KEGG Pathway Enrichment Analysis (Human EC, sn-RNA seq: GSE147528, Immune pathways only)



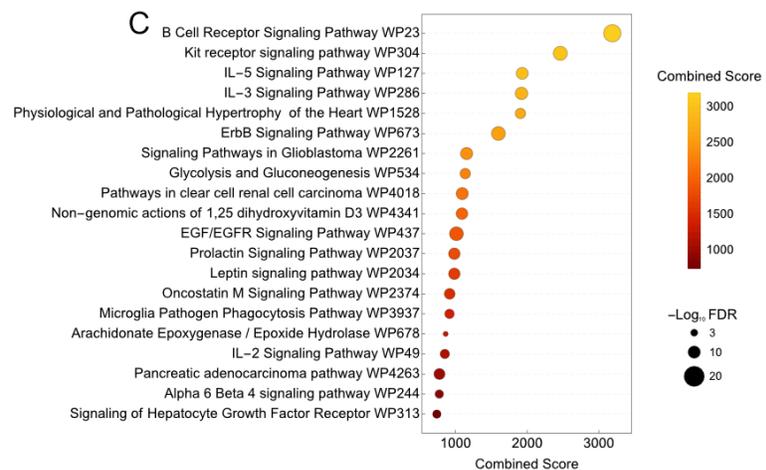
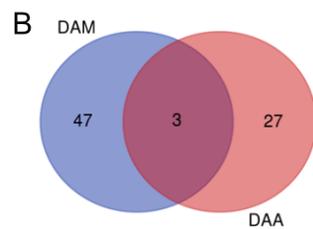
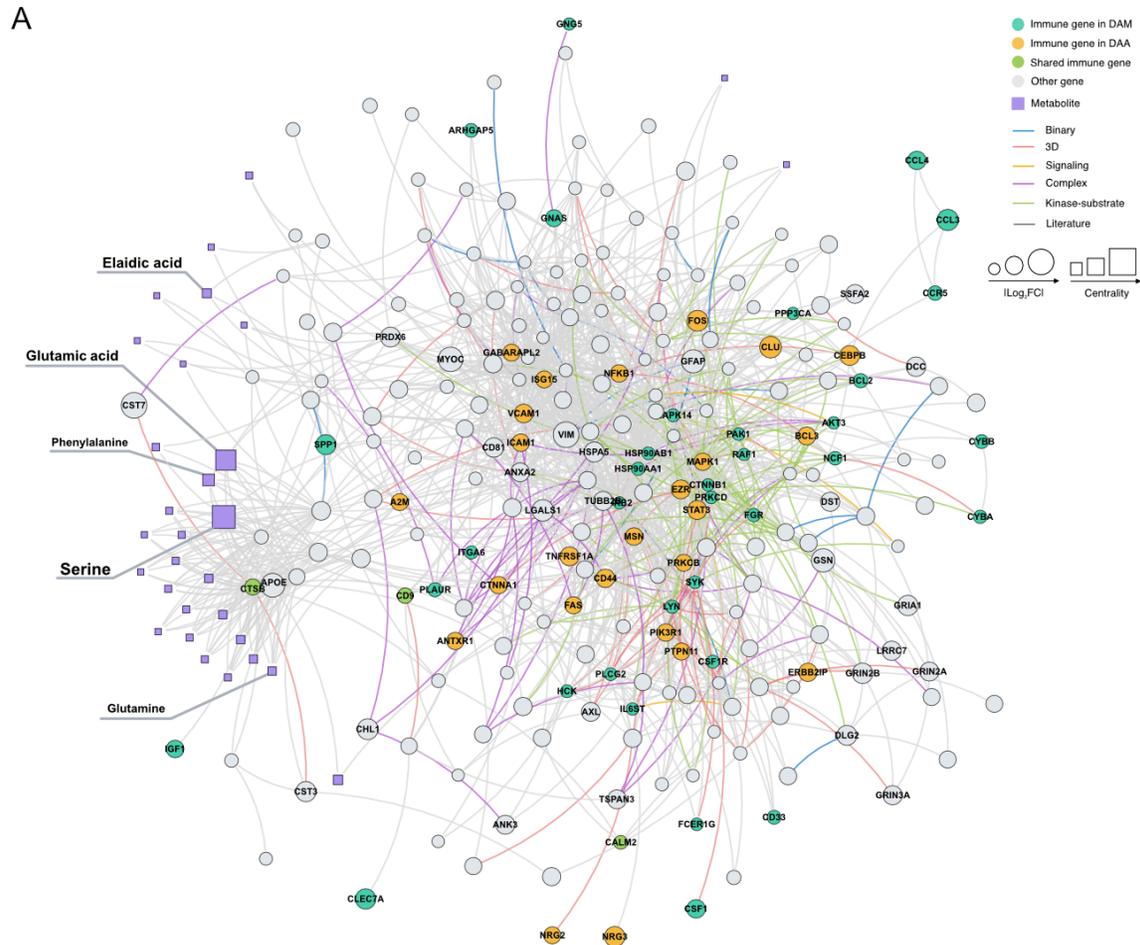
**C** Differentially Expressed Genes Analyses between DAA and non-DAA (EC), sn-RNA seq: GSE147528



**D** Differentially Expressed Genes Analyses between DAA and non-DAA (SFG), sn-RNA seq: GSE147528



**Supplemental Figure S7:** Differentially expressed genes and pathway enrichment analyses for disease associated astrocytes (DAAs) built from human AD patient snRNA-seq data (GSE147528). (A-B) Pathway enrichment analyses (**Supplemental Table S9**) for molecular networks built from (A) entorhinal cortex (EC) and (B) superior frontal gyrus (SFG). Differentially expressed gene analyses (volcano plot) between DAAs and non-disease associated astrocytes (non-DAAs) in patients' 2 brain regions: (C) entorhinal cortex (D) superior frontal gyrus.



**Supplemental Figure S8:** Network visualization and pathway enrichment analysis for disease associated astrocyte (DAA) and disease associated microglia (DAM). (A) A module illustrating the network-based relationship between DAA and DAM immune genes associated with AD-related metabolites. (B) Venn diagram of enzymes from DAA and DAM. (C) Pathway enrichment of 77 enzymes in DAA and DAM.

## Supplemental References

- Abul-Husn NS, Manickam K, Jones LK, Wright EA, Hartzel DN, Gonzaga-Jauregui C, O'Dushlaine C, Leader JB, Lester Kirchner H, Lindbuchler DM, et al. 2016. Genetic identification of familial hypercholesterolemia within a single U.S. health care system. *Science* **354**. doi:10.1126/science.aaf7000
- Banerjee D, Chung S, Wong EC, Wang EJ, Stafford RS, Palaniappan LP. 2012. Underdiagnosis of hypertension using electronic health records. *Am J Hypertens* **25**: 97–102. doi:10.1038/ajh.2011.179
- Breuer K, Foroushani AK, Laird MR, Chen C, Sribnaia A, Lo R, Winsor GL, Hancock REW, Brinkman FSL, Lynn DJ. 2013. InnateDB: systems biology of innate immunity and beyond—recent updates and continuing curation. *Nucleic Acids Res* **41**: D1228–D1233. doi:10.1093/nar/gks1147
- Brunk E, Sahoo S, Zielinski DC, Altunkaya A, Dräger A, Mih N, Gatto F, Nilsson A, Preciat Gonzalez GA, Aurich MK, et al. 2018. Recon3D enables a three-dimensional view of gene variation in human metabolism. *Nat Biotechnol* **36**: 272–281. doi:10.1038/nbt.4072
- Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. 2018. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **36**: 411–420. doi:10.1038/nbt.4096
- Chatr-aryamontri A, Breitkreutz B-J, Oughtred R, Boucher L, Heinicke S, Chen D, Stark C, Breitkreutz A, Kolas N, O'Donnell L, et al. 2015. The BioGRID interaction database: 2015 update. *Nucleic Acids Res* **43**: D470–D478. doi:10.1093/nar/gku1204
- Cheng F, Jia P, Wang Q, Zhao Z. 2014. Quantitative network mapping of the human kinome interactome reveals new clues for rational kinase inhibitor discovery and individualized cancer therapy. *Oncotarget* **5**: 3697–3710. doi:10.18632/oncotarget.1984
- Cowley MJ, Pinese M, Kassahn KS, Waddell N, Pearson JV, Grimmond SM, Biankin AV, Hautaniemi S, Wu J. 2012. PINA v2.0: mining interactome modules. *Nucleic Acids Res* **40**: D862–D865. doi:10.1093/nar/gkr967
- Dabney A, Storey JD, R AFG. 2013. Title Q-value estimation for false discovery rate control Version 1.32.0. <https://rdr.io/bioc/qvalue/f/inst/doc/qvalue.pdf>
- Dinkel H, Chica C, Via A, Gould CM, Jensen LJ, Gibson TJ, Diella F. 2011. Phospho.ELM: a database of phosphorylation sites—update 2011. *Nucleic Acids Res* **39**: D261–D267. doi:10.1093/nar/gkq1104
- Fazekas D, Koltai M, Türei D, Módos D, Pálffy M, Dúl Z, Zsákai L, Szalay-Bekő M, Lenti K, Farkas IJ, et al. 2013. Signalink 2 – a signaling pathway resource with multi-layered regulatory networks. *BMC Syst Biol* **7**: 7. doi:10.1186/1752-0509-7-7
- Federman DG, Krishnamurthy R, Kancir S, Goulet J, Justice A. 2005. Relationship between provider type and the attainment of treatment goals in primary care. *Am J Manag Care* **11**: 561–566.

- Goel R, Harsha HC, Pandey A, Keshava Prasad TS. 2012. Human Protein Reference Database and Human Proteinpedia as Resources for Phosphoproteome Analysis. *Mol Biosyst* **8**: 453–463. doi:10.1039/c1mb05340j
- Grubman A, Chew G, Ouyang JF, Sun G, Choo XY, McLean C, Simmons RK, Buckberry S, Vargas-Landin DB, Poppe D, et al. 2019. A single-cell atlas of entorhinal cortex from individuals with Alzheimer's disease reveals cell-type-specific gene expression regulation. *Nat Neurosci* **22**: 2087–2097. doi:10.1038/s41593-019-0539-4
- Habib N, McCabe C, Medina S, Varshavsky M, Kitsberg D, Dvir-Szternfeld R, Green G, Dionne D, Nguyen L, Marshall JL, et al. 2020. Disease-associated astrocytes in Alzheimer's disease and aging. *Nat Neurosci* **23**: 701–706. doi:10.1038/s41593-020-0624-8
- Hornbeck PV, Zhang B, Murray B, Kornhauser JM, Latham V, Skrzypek E. 2015. PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res* **43**: D512–D520. doi:10.1093/nar/gku1267
- Hu J, Rho H-S, Newman RH, Zhang J, Zhu H, Qian J. 2014. PhosphoNetworks: a database for human phosphorylation networks. *Bioinformatics* **30**: 141–142. doi:10.1093/bioinformatics/btt627
- Huttlin EL, Ting L, Bruckner RJ, Gebreab F, Gygi MP, Szpyt J, Tam S, Zarraga G, Colby G, Baltier K, et al. 2015. The BioPlex Network: A Systematic Exploration of the Human Interactome. *Cell* **162**: 425–440. doi:10.1016/j.cell.2015.06.043
- Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. 2017. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* **45**: D353–D361. doi:10.1093/nar/gkw1092
- Keren-Shaul H, Spinrad A, Weiner A, Matcovitch-Natan O, Dvir-Szternfeld R, Ulland TK, David E, Baruch K, Lara-Astaiso D, Toth B, et al. 2017. A Unique Microglia Type Associated with Restricting Development of Alzheimer's Disease. *Cell* **169**: 1276-1290.e17. doi:10.1016/j.cell.2017.05.018
- Kho AN, Hayes MG, Rasmussen-Torvik L, Pacheco JA, Thompson WK, Armstrong LL, Denny JC, Peissig PL, Miller AW, Wei W-Q, et al. 2012. Use of diverse electronic medical record systems to identify genetic risk for type 2 diabetes within a genome-wide association study. *J Am Med Inform Assoc* **19**: 212–218. doi:10.1136/amiainjnl-2011-000439
- Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, Lerner J, Brunet J-P, Subramanian A, Ross KN, et al. 2006. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* **313**: 1929–1935. doi:10.1126/science.1132939
- Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. 2012. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**: 882–883. doi:10.1093/bioinformatics/bts034

- Leng K, Li E, Eser R, Piergies A, Sit R, Tan M, Neff N, Li SH, Rodriguez RD, Suemoto CK, et al. 2020. Molecular characterization of selectively vulnerable neurons in Alzheimer's Disease. *Nat Neurosci*, **24**: 276-287. doi:10.1038/s41593-020-00764-7
- Licata L, Briganti L, Peluso D, Perfetto L, Iannuccelli M, Galeota E, Sacco F, Palma A, Nardoza AP, Santonico E, et al. 2012. MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res* **40**: D857–D861. doi:10.1093/nar/gkr930
- Luck K, Kim D-K, Lambourne L, Spirohn K, Begg BE, Bian W, Brignall R, Cafarelli T, Campos-Laborie FJ, Charlotheaux B, et al. 2020. A reference map of the human binary protein interactome. *Nature* **580**: 402–408. doi:10.1038/s41586-020-2188-x
- Lun ATL, McCarthy DJ, Marioni JC. 2016. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Res* **5**: 2122. doi:10.12688/f1000research.9501.2
- McCarthy DJ, Campbell KR, Lun ATL, Wills QF. 2017. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* **33**: 1179–1186. doi:10.1093/bioinformatics/btw777
- Meyer MJ, Das J, Wang X, Yu H. 2013. INstruct: a database of high-quality 3D structurally resolved protein interactome networks. *Bioinformatics* **29**: 1577–1579. doi:10.1093/bioinformatics/btt181
- Orchard S, Ammari M, Aranda B, Breuza L, Briganti L, Broackes-Carter F, Campbell NH, Chavali G, Chen C, del-Toro N, et al. 2014. The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. *Nucl Acids Res* **42**: D358–D363. doi:10.1093/nar/gkt1115
- Peri S. 2004. Human protein reference database as a discovery resource for proteomics. *Nucleic Acids Res* **32**: 497D – 501. doi:10.1093/nar/gkh070
- Smilowitz NR, Oberweis BS, Nukala S, Rosenberg A, Zhao S, Xu J, Stuchin S, Iorio R, Errico T, Radford MJ, et al. 2016. Association Between Anemia, Bleeding, and Transfusion with Long-term Mortality Following Noncardiac Surgery. *Am J Med* **129**: 315-323.e2. doi:10.1016/j.amjmed.2015.10.012
- Wei W-Q, Leibson CL, Ransom JE, Kho AN, Caraballo PJ, Chai HS, Yawn BP, Pacheco JA, Chute CG. 2012. Impact of data fragmentation across healthcare centers on the accuracy of a high-throughput clinical phenotyping algorithm for specifying subjects with type 2 diabetes mellitus. *J Am Med Inform Assoc* **19**: 219–224. doi:10.1136/amiajnl-2011-000597
- Wei W-Q, Teixeira PL, Mo H, Cronin RM, Warner JL, Denny JC. 2016. Combining billing codes, clinical notes, and medications from electronic health records provides superior phenotyping performance. *J Am Med Inform Assoc* **23**: e20-27. doi:10.1093/jamia/ocv130
- Wilkinson T, Ly A, Schnier C, Rannikmäe K, Bush K, Brayne C, Quinn TJ, Sudlow CLM. 2018. Identifying dementia cases with routinely collected health data: A systematic review. *Alzheimers Dement* **14**: 1038–1051. doi:10.1016/j.jalz.2018.02.016

- Wishart DS, Feunang YD, Marcu A, Guo AC, Liang K, Vázquez-Fresno R, Sajed T, Johnson D, Li C, Karu N, et al. 2018. HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res* **46**: D608–D617. doi:10.1093/nar/gkx1089
- Zhou Y, Hou Y, Shen J, Huang Y, Martin W, Cheng F. 2020a. Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell Discov* **6**: 14. doi:10.1038/s41421-020-0153-3
- Zhou Y, Song WM, Andhey PS, Swain A, Levy T, Miller KR, Poliani PL, Cominelli M, Grover S, Gilfillan S, et al. 2020b. Human and mouse single-nucleus transcriptomics reveal TREM2-dependent and TREM2-independent cellular responses in Alzheimer's disease. *Nat Med* **26**: 131–142. doi:10.1038/s41591-019-0695-9