

## SUPPLEMENTARY METHODS

### Study Participants

Study participants were recruited as part of ongoing studies of Alzheimer's disease (AD) and related neurodegenerative diseases at the University of California, San Francisco (UCSF) Memory and Aging Center (MAC). All participants or their surrogates provided written informed consent prior to study participation, and all aspects of the studies described here were approved by the UCSF institutional review board (IRB). Sixteen individuals ( $n = 8$  early-onset Alzheimer's disease [EOAD] cases and  $n = 8$  cognitively normal controls) participated in the single-cell RNA sequencing (scRNA-seq) study. Nineteen individuals ( $n = 9$  EOAD cases and  $n = 10$  cognitively normal controls) participated in the droplet digital (dd)PCR study. To ensure replicability across methodologies, a subset of 4 EOAD cases was included in both studies. EOAD cases and cognitively normal controls were sex-matched, and equal numbers of female and male participants were included in the scRNA-seq study. The EOAD group was significantly older than the control group in the scRNA-seq study ( $P = 0.0004$ ), while ages did not significantly differ between cases and controls in the ddPCR study ( $P = 0.37$ ), as assessed using unpaired  $t$ -tests. Two control participants included in the studies described herein are siblings; however, these two controls were included separately in either the scRNA-seq study or the ddPCR study. No EOAD cases studied herein are related to any other included EOAD cases or controls. Demographic information for the scRNA-seq and ddPCR study participants is included in Table 1.

### Cell Isolation

Human peripheral blood mononuclear cells (PBMCs) were obtained from study participants at the UCSF MAC. Blood samples were collected in yellow-cap ACD Vacutainer tubes (BD Biosciences) and processed within five hours of collection, as described previously [1]. PBMCs were isolated by Ficoll density gradient sedimentation using Lymphosep separation medium (MP Biomedicals), washed with  $\text{Ca}^{2+}$ - and  $\text{Mg}^{2+}$ -free PBS (ThermoFisher), and treated with red blood cell lysis buffer (BioLegend). PBMCs were then washed again with PBS and diluted to a density of  $1.5 \times 10^6$  cells/ml in freezing medium composed of 10% DMSO in FBS and immediately frozen at  $-80^\circ\text{C}$ . Samples were transferred to liquid nitrogen for long-term storage after two weeks of storage at  $-80^\circ\text{C}$ . All PBMC samples used in our primary analyses were cryopreserved.

## Single-Cell RNA-seq

PBMCs were thawed and prepared for scRNA-seq using the Chromium Single Cell 3' v3 kit according to the manufacturer's instructions (10x Genomics). Samples were processed in two separate batches of eight samples each, with four EOAD cases and four controls included in each batch. To minimize the potential for batch effects, each batch contained equal numbers of samples from female and male participants. After sample thawing, counting, and dilution, PBMCs underwent standard 10x processing, 3' gene expression library construction steps, and next-generation sequencing at the UCSF Genomics CoLab and Institute for Human Genetics.

## Sequencing Data Processing

For each of the two batches, single-cell 3' libraries generated from eight samples were pooled and sequenced on one lane of a NovaSeq S4 flow cell. Raw sequencing reads were aligned to GRCh38-2020-A, and feature-barcode matrices were generated using Cell Ranger version 7.1.0 with intronic reads excluded.

## Quality Control

We obtained a total of  $7.4 \times 10^9$  reads and detected ~260,000 cells across the two independent 10x and sequencing batches, yielding a moderate sequencing depth [2,3] of ~30,000 mean reads/cell. We detected ~5,300 median UMI counts/cell and ~1,600 median genes/cell (supplementary Table S1). There were no significant differences in the number of cells captured per sample, the number of reads per sample, or the mean read depth per sample when comparing the EOAD group to the control group. Subsequent quality-control (QC) and downstream analysis steps were performed using Seurat v4.3.0.1 [4,5]. QC filtering was applied to individual-sample feature-barcode matrices and consisted of the following steps: (i) genes detected in < 10 cells were removed; (ii) cells with  $\leq 500$  detected genes were removed; (iii) cells with  $\leq 500$  counts and those with  $\geq 20,000$  counts were removed; (iv) cells with mitochondrial mapping percentages  $\geq 10$  were removed; (v) doublets were identified and removed using DoubletFinder v2.0.3 [6,7] using the recommended parameter settings. After stringent QC filtering, ~182,000 cells remained for downstream analysis (supplementary Table S2).

## Clustering

After QC, we performed the following additional processing steps: (i) we applied `sctransform` v2 regularization [8,9] at the individual-sample level, including mitochondrial mapping percentage as a covariate [8,10], to minimize variability due to differences in sequencing depth between samples; (ii) the 16 individual samples were integrated with `FindIntegrationAnchors` and `IntegrateData`, specifying 'sctransform' as the normalization method and canonical correlation analysis (CCA) as the reduction. Subsequently, PCA was performed followed by uniform manifold approximation and projection (UMAP) reduction using the first 30 PCs; clustering was performed using a resolution parameter of 0.5. This resulted in the generation of 19 clusters that were annotated via multimodal reference mapping to a large, well-characterized human PBMC dataset [5]. After omitting two clusters (one dominated by contaminating platelets and the other with mixed myeloid and lymphoid markers), 17 primary clusters remained for analysis of relative abundance. Sub-clustering was performed by subsetting all T cells (as identified by multimodal reference mapping) followed by PCA, UMAP reduction using the first 30 PCs, and clustering using a resolution of 0.5. To integrate additional cognitively normal control samples from a previously published study [1] into the dataset, we performed the same analyses outlined above using a resolution of 0.4.

## Cluster Proportionality

Cluster proportions (expressed as % of total PBMCs) were determined for individual samples by dividing the number of cells in a given cluster by the total number of cells in all clusters (after QC filtering) for each individual. To determine cluster abundance relative to all T cells or all CD4 T cells, a similar calculation was performed, using as the denominator all cells annotated as T cells or CD4 T cells by multimodal reference mapping. Differences in cluster proportionality were initially assessed visually for all clusters according to disease status. Only the interferon (IFN) signaling-associated gene (ISAG)<sup>hi</sup> T-cell cluster (cluster 15) showed a clear difference in proportionality between cases and controls. Statistical significance for cluster 15 proportionality, as well as ISAG<sup>hi</sup> T-cell subcluster 11 proportionality, was determined by linear modeling, covarying for age and sex. To account for multiple testing across the 17 primary clusters, we performed false discovery rate [FDR] correction with a  $P_{FDR}$  threshold of 0.1, striking a balance to avoid both type I and type II error. To test for significant differences in ISAG<sup>hi</sup> T-cell relative abundance in the dataset containing seven additional cognitively normal controls, the cluster

proportions were  $\log_2$ -transformed prior to analysis to minimize the effect of the single outlier control sample.

### Differential Expression Analysis

Differential expression (DE) analysis was performed using limma [11–14] on individual clusters and  $\text{sc}t\text{ransform}$   $v_2$ -normalized data, with disease status as the contrast and age, sex and batch as covariates (as in [1]). The publicly available CSF dataset (GSE134577) was analyzed essentially in the same manner, covarying for age and sex. To account for multiple testing, a FDR-corrected  $p$ -value ( $P_{FDR}$ )  $< 0.05$  was considered statistically significant, and only genes with estimated absolute  $\log_2$  fold-changes (LFC)  $> 0.1$  were considered differentially expressed.

### CD4 T cell isolation

CD4 T cells were isolated from PBMCs using the magnetic-activated cell sorting (MACS) human CD4 T cell isolation kit (Miltenyi). The purity of the isolated CD4 T cells was assessed using a droplet digital (dd)PCR assay that measured the enrichment of the *CD4/CD8A* transcript ratio relative to the starting PBMC population. This assay indicated an  $\sim 4,800$ -fold enrichment of the *CD4/CD8A* ratio in the isolated CD4 T cells relative to starting PBMCs, indicating successful isolation of CD4 T cells.

### RNA Extraction

RNA was extracted from CD4 T cells using the RNeasy Micro Kit (Qiagen) and isolated RNA was quantified and its quality was assessed using the RNA 6000 Pico Bioanalyzer kit (Agilent). CD4 T-cell RNA samples had RNA integrity number (RIN) values ranging from 8.5–10, indicating high-quality RNA (supplementary Table S3) [15].

### Droplet Digital (dd) PCR Cohort

For ddPCR experiments, we isolated CD4 T cell RNA from 19 participants. In the first ddPCR batch, we selected 5 independent controls and 4 EOAD cases with a clear expansion of ISAG<sup>hi</sup> T cells, as assessed by scRNA-seq. These cases were selected to determine the suitability of ddPCR for assessing the ISAG<sup>hi</sup> T-cell phenotype. The second ddPCR batch consisted of an additional 5 independent cases and 5 independent controls.

## ddPCR Procedure and Analysis

For primary analyses, two ng of total RNA was used for single-tube reverse transcription (RT) and ddPCR using the One-Step RT-ddPCR Advanced kit (Bio-Rad). For validation of the CD4 T-cell isolation procedure, 30 ng of RNA was used to ensure that the low *CD8A* signal could be measured over background in the isolated CD4 T cells. Droplets were generated and subsequently analyzed using the QX100 system (Bio-Rad) at the UCSF Center for Advanced Technology (CAT). Reactions were prepared and run essentially according to the manufacturer's instructions. For steps in which a temperature range was specified, we used the following parameters: RT was performed at 50°C, annealing/extension occurred at 55°C, and samples were held at 12°C in the C1000 thermocycler (Bio-Rad) prior to analysis on the droplet reader. To confirm specificity, we ran the following control reactions: wells lacking RNA but containing all other components and wells lacking reverse transcriptase but containing all other components. PrimePCR ddPCR Gene Expression primer-probe mixes coupled to FAM or HEX (Bio-Rad) were used to amplify specific genes. For the analysis of ddPCR data, we performed linear modeling to assess whether disease status predicted differences in gene expression while covarying for age and sex. Log<sub>2</sub>-transformed absolute concentration data for *MX1* and *IFI6* (normalized to *CD4* as the reference gene) were used for analyses that assume normality, while non-transformed data are displayed in the plots. Differences were considered significant at  $p \leq 0.05$ .

## Secondary Analysis of Publicly Available CSF Leukocyte Dataset

Publicly available human CSF leukocyte scRNA-seq data were downloaded from GEO (accession GSE134577) [16]. The original cohort consists of samples from 18 individuals ( $n = 9$  healthy older controls,  $n = 5$  participants with mild cognitive impairment (MCI), and  $n = 4$  participants with AD). For additional information, see Gate et al., 2020 [16] and supplementary Table S6. The downloaded counts data was processed and analyzed as described above for our PBMC scRNA-seq data. We used the marker genes *MX1*, *IFI6*, *CD3E*, and *CD4* to identify a CSF leukocyte cluster corresponding to ISAG<sup>hi</sup> T cells. For our analyses, in order to focus on the cases with late onset, we excluded three cases that had ages at collection below 65 years. DE analysis was performed as described above for the EOAD PBMC dataset, except that cases were defined as samples with diagnoses of either MCI or AD. Genes significantly upregulated in MCI/AD with  $P_{FDR} < 0.05$  and LFC  $> 0.1$  from any CSF CD4 T-cell cluster (expressing *CD3E* and *CD4*; clusters 0, 1, 5, 6, 7, 8, 9, 12, and 15; see Figure 3) were pooled and submitted for

gene ontology (GO) analysis via g:Profiler [17]. We queried the GO biological process (BP) [18] and reactome databases [19] and selected for display significantly enriched terms related to IFN and antiviral responses. For the functional enrichment analysis of individual CSF cell clusters, we submitted upregulated DEGs from each cluster that had at least 10 upregulated genes; all such clusters, except cluster 11, showed significant enrichment for the reactome terms 'interferon signaling' and 'interferon alpha/beta signaling'.

### Secondary Analysis of Mouseac Dataset

To further evaluate the expression of ISAG<sup>hi</sup> T-cell marker genes in a mouse model of AD-like amyloidosis, we examined genes implicated by our prior analyses in an AD model compared to wild type (WT) mice. In particular, we queried the most significantly overlapping set of ISAG<sup>hi</sup> marker genes between our ISAG<sup>hi</sup> T-cell cluster and the CSF antiviral CD4 T-cell cluster from Heming et al., 2021 [20]. Data for these analyses was obtained from the Mouseac project, which has been described in detail elsewhere [21,22]. Briefly, the Mouseac project is a study of murine brain tissue at varying ages (i.e., 8, 16, 32, and 72 weeks) in models of neurodegeneration. Samples were collected from cortex, hippocampus, and cerebellum using the TASTPM model of AD (TAS10 × TPM AD mouse models; APP<sup>swe</sup> × PS1.M146V) and WT mice [23]. Of note, both heterozygous and homozygous TASTPM mice were included in our analyses. Gene expression was measured using Illumina Ref8 v2 microarray with processing completed by the Mouseac project as previously described [21]. Briefly, expression data was log<sub>2</sub> normalized with quantile normalization. Individual probes were excluded if the detection p-value was > 0.05 in > 50% of samples. A total of 16 ISAG<sup>hi</sup> marker genes were available for analysis in the downloaded dataset. ANOVA was used to analyze gene expression in comparing across age groups, gene status, and brain region. Following this, comparisons by TASTPM status were repeated within the hippocampus and cortex, the results of which are shown in supplementary Figure S7A. Differences by TASTPM status were considered significant for the hippocampus and cortex at  $P = 3.1 \times 10^{-3}$ , accounting for 16 genes being tested.

### Additional Information

Analyses were performed in R and plots were generated using Seurat and ggplot2. For the statistical analysis of overlap between marker gene lists, we used DynaVenn [24].

## Availability of Data

The scRNA-seq dataset described here has been posted to the FAIR Data Sharing Portal within the Alzheimer's Disease Workbench, which is supported by the Alzheimer's Disease Data Initiative, and is accessible at: [https://fair.addi.ad-datainitiative.org/#/data/datasets/single\\_cell\\_rna\\_seq\\_data\\_derived\\_from\\_early\\_onset\\_ad\\_cases\\_and\\_controls](https://fair.addi.ad-datainitiative.org/#/data/datasets/single_cell_rna_seq_data_derived_from_early_onset_ad_cases_and_controls). The familial tauopathy scRNA-seq dataset is also available on the FAIR Data Sharing Portal and can be accessed at: [https://fair.addi.ad-datainitiative.org/#/data/datasets/single\\_cell\\_rna\\_seq\\_data\\_derived\\_from\\_mapt\\_carriers\\_and\\_controls](https://fair.addi.ad-datainitiative.org/#/data/datasets/single_cell_rna_seq_data_derived_from_mapt_carriers_and_controls).

## References

- [1] Sirkis DW, Warly Solsberg C, Johnson TP, Bonham LW, Sturm VE, Lee SE, et al. Single-cell RNA-seq reveals alterations in peripheral CX3CR1 and nonclassical monocytes in familial tauopathy. *Genome Med* 2023;15:53.
- [2] Haque A, Engel J, Teichmann SA, Lönnberg T. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Med* 2017;9:75.
- [3] Nguyen QH, Pervolarakis N, Nee K, Kessenbrock K. Experimental Considerations for Single-Cell RNA Sequencing Approaches. *Front Cell Dev Biol* 2018;6:108.
- [4] Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM 3rd, et al. Comprehensive Integration of Single-Cell Data. *Cell* 2019;177:1888–902.e21.
- [5] Hao Y, Hao S, Andersen-Nissen E, Mauck WM 3rd, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. *Cell* 2021;184:3573–87.e29.
- [6] McGinnis CS, Murrow LM, Gartner ZJ. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst* 2019;8:329–37.e4.
- [7] Xi NM, Li JJ. Benchmarking Computational Doublet-Detection Methods for Single-Cell RNA Sequencing Data. *Cell Syst* 2021;12:176–94.e6.
- [8] Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol* 2019;20:296.
- [9] Choudhary S, Satija R. Comparison and evaluation of statistical error models for scRNA-seq. *Genome Biol* 2022;23:27.
- [10] Ilicic T, Kim JK, Kolodziejczyk AA, Bagger FO, McCarthy DJ, Marioni JC, et al. Classification of low quality cells from single-cell RNA-seq data. *Genome Biol* 2016;17:29.
- [11] Law CW, Chen Y, Shi W, Smyth GK. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* 2014;15:R29.
- [12] Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015;43:e47.
- [13] Soneson C, Robinson MD. Bias, robustness and scalability in single-cell differential expression analysis. *Nat Methods* 2018;15:255–61.
- [14] Mou T, Deng W, Gu F, Pawitan Y, Vu TN. Reproducibility of Methods to Detect Differentially Expressed Genes from Single-Cell RNA Sequencing. *Front Genet* 2019;10:1331.
- [15] Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, Gassmann M, et al. The RIN: an



- RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol Biol* 2006;7:1–14.
- [16] Gate D, Saligrama N, Leventhal O, Yang AC, Unger MS, Middeldorp J, et al. Clonally expanded CD8 T cells patrol the cerebrospinal fluid in Alzheimer's disease. *Nature* 2020;577:399–404.
- [17] Kolberg L, Raudvere U, Kuzmin I, Adler P, Vilo J, Peterson H. g:Profiler-interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). *Nucleic Acids Res* 2023;51:W207–12.
- [18] The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res* 2019;47:D330–8.
- [19] Gillespie M, Jassal B, Stephan R, Milacic M, Rothfels K, Senff-Ribeiro A, et al. The reactome pathway knowledgebase 2022. *Nucleic Acids Res* 2022;50:D687–92.
- [20] Heming M, Li X, Räuber S, Mausberg AK, Börsch A-L, Hartlehnert M, et al. Neurological Manifestations of COVID-19 Feature T Cell Exhaustion and Dedifferentiated Monocytes in Cerebrospinal Fluid. *Immunity* 2021;54:164–75.e6.
- [21] Matarin M, Salih DA, Yasvoina M, Cummings DM, Guelfi S, Liu W, et al. A genome-wide gene-expression analysis and database in transgenic mice during development of amyloid or tau pathology. *Cell Rep* 2015;10:633–44.
- [22] Bonham LW, Sirkis DW, Yokoyama JS. The Transcriptional Landscape of Microglial Genes in Aging and Neurodegenerative Disease. *Front Immunol* 2019;10:1170.
- [23] Howlett DR, Richardson JC, Austin A, Parsons AA, Bate ST, Davies DC, et al. Cognitive correlates of Abeta deposition in male and female mice bearing amyloid precursor protein and presenilin-1 mutant transgenes. *Brain Res* 2004;1017:130–6.
- [24] Amand J, Fehlmann T, Backes C, Keller A. DynaVenn: web-based computation of the most significant overlap between ordered sets. *BMC Bioinformatics* 2019;20:743.