Genetic variants associated with Alzheimer's disease confer different cerebral cortex cell-type population structure

Ms. Zeran Li[1], Dr. Jorge L Del-Aguila[1], Mr. Umber Dube[2], Mr. John Budde[1], Dr. Rita Martinez[1], Ms. Kathleen Black[1], Dr. Qingli Xiao[3], Prof. Nigel J. Cairns[3,4,5], the Dominantly Inherited Alzheimer Network (DIAN), Dr. Joseph D. Dougherty[1,7], Prof. Jin-Moo Lee[3], Prof. John C Morris[3,5,6], Prof. Randall J. Bateman[3,5,6], Dr. Celeste M. Karch[1], Dr. Carlos Cruchaga[1,5,6,Π] and Dr. Oscar Harari[1,Π]

Affiliations:
1. Department of Psychiatry, Washington University School of Medicine, 660 S. Euclid Ave. B8134, St. Louis, MO 63110, USA.
2. Medical Scientist Training Program, Washington University School of Medicine, 660 S. Euclid Ave., St. Louis, MO 63110, USA.
3. Department of Neurology, Washington University School of Medicine, 660 S. Euclid Ave., St. Louis, MO 63110, USA.
4. Department of Pathology & Immunology, Washington University in St. Louis, School of Medicine, 510 S. Kingshighway, MC 8131, Saint Louis, MO 63110, USA.
5. Knight Alzheimer's Disease Research Center, Washington University School of Medicine, 660 S. Euclid Ave., St. Louis, MO 63110, USA.
6. Hope Center for Neurological Disorders. Washington University School of Medicine, 660 S. Euclid Ave. B8111, St. Louis, MO 63110, USA.
7. Department of Genetics, Washington University School of Medicine, 660 S. Euclid Ave., St. Louis, MO 63110, USA.

[Π] To whom correspondence should be addressed:
Carlos Cruchaga, PhD
Associate Professor
Department of Psychiatry
The Hope Center Program on Protein Aggregation and Neurodegeneration
Washington University, School of Medicine
425 S. Euclid Ave.
BJC Institute of Heath. Box 8134
St. Louis, MO 63110
Tel: 314-286-0546 // Fax: 314-362-2244
Email: ccruchaga@wustl.edu

Oscar Harari, PhD
Assistant Professor
Department of Psychiatry
Washington University School of Medicine
660 South Euclid Avenue B8134, St. Louis, MO 63110
Tel. 314-286-0546 // Fax. 314-362-2244
Email: harario@wustl.edu

Abstract

Alzheimer's disease (AD) is characterized by neuronal loss and astrocytosis in the cerebral cortex. However, the effects of brain cellular composition are often ignored in high-throughput molecular studies. We developed and optimized a cell-type specific expression reference panel and employed digital deconvolution methods to determine brain cellular distribution in three independent transcriptomic studies. We found that neuronal and astrocyte proportions differ between healthy and diseased brains and also among AD cases that carry specific genetic risk variants. Brain carriers of pathogenic mutations in *APP*, *PSEN1* or *PSEN2* presented lower neurons and higher astrocytes proportions compared to sporadic AD. Similarly, the *APOE* ε4 allele also showed decreased neurons and increased astrocytes compared to AD non-carriers. On the contrary, carriers of variants in *TREM2* risk showed a lower degree of neuronal loss than matched AD cases in multiple independent studies. These findings suggest that genetic risk factors associated with AD etiology have a specific imprinting in the cellular composition of AD brains. Our digital deconvolution reference panel provides an enhanced understanding of the fundamental molecular mechanisms underlying neurodegeneration, enabling the analysis of large bulk RNA-seq studies for cell composition, and suggests that correcting for the cellular structure when performing transcriptomic analysis will lead to novel insights of AD.

Keywords

Digital deconvolution, Alzheimer's disease, cellular composition, bulk RNA-seq, autosomal dominant AD, *TREM2*.

Introduction

Alzheimer's Disease (AD) is a neurodegenerative disorder characterized clinically by gradual and progressive memory loss and pathologically by the presence of senile plaques (Aβ deposits) and neurofibrillary tangles (NFTs, Tau deposits) in the brain [41]. AD has a substantial but heterogeneous genetic component. Mutations in the amyloid-beta precursor protein (*APP*) and *Presenilin genes* (*PSEN1* and *PSEN2*) [21, 59] cause autosomal dominant AD (ADAD) which is typically associated with early-onset (<65 years). In contrast, the most common manifestation of AD presents late-onset (LOAD) and accounts for the majority of the cases (90-95%). Despite appearing sporadic in nature, a complex genetic architecture underlies LOAD risk. *APOE* ε4 is the most common genetic risk factor, increasing the risk in 3- to 8-fold [19]. In addition, recent whole genome and whole exome analysis have identified rare coding variants in *TREM2* [9, 32], *PLD3* [20], *ABCA7* [22, 63] and *SORL1* [26, 56] that are associated with AD and confer risk comparable to that of carrying one *APOE* ε4 allele. Besides age at onset, the clinical presentations of LOAD and ADAD are remarkably similar with an amnestic and cognitive impairment phenotype [57, 66]. A minor fraction of cases of ADAD have additional neurological findings, sometimes also seen in LOAD [57, 66].

Altered cellular composition is associated with AD progression and decline in cognition. Neuronal loss in the hippocampus is characteristic in the initial stages of AD, which could explain early memory disturbances [52, 71]. As the disease progresses, neuronal death is observed throughout the cerebral cortex. Furthermore, ~25% of individuals who die by ~75 years of age who were cognitively normal also presented substantial cerebral lesions that resemble AD pathology, including amyloid plaque, NFTs, and neuronal loss [37]. Thus, the identification of the brain cellular population structure is essential for understanding neurodegenerative disease progression [30]. However, stereology protocols for counting neurons can be tedious, require extensive training and are susceptible to technical artifacts which may lead to biased quantification of cell-type distributions [30].

Recently there has been a growing interest in understanding the transcriptomic changes attributed to AD [8, 16, 27, 46, 50, 53, 62, 72], as these may point to underlying molecular mechanisms of disease. These studies are typically designed to analyze the expression profiles of large cohorts ascertained from homogenized regions of the brain (e.g. bulk RNA-seq) of affected and control donors. However, bulk RNA-seq captures the gene expression of all of the constituent cells in the sampled tissue, and the altered cellular composition associated with AD has been reported to confound downstream analyses [62].

Digital deconvolution approaches enhance the interrogation of expression profiles to identify the cellular population structure of individual samples, alleviating the requirement of additional neurostereology procedures. These approaches have been developed, tested and applied to ascertain cellular composition altered in many traits [40, 51, 61, 75]. However, digital deconvolution has not been applied to identify the cellular population structure from RNA-seq from human brain tissue. Technical constraints restrict the dissociation of cells from the brains for very specific conditions [13, 73, 74]. Nevertheless, a limited number of RNA-seq from isolated cell populations from the brain have been generated [13, 73, 74]. Using these resources, we are now able to generate a reference panel for digital deconvolution of human brain bulk RNA-seq data.

We sought to investigate the cellular population structure in AD by analyzing RNA-seq from multiple brain regions of LOAD participants. To do so, we assembled a novel brain reference panel and evaluated the accuracy of digital deconvolution methods by analyzing additional cell-type specific RNA-seq samples and by creating synthetic admixtures with defined cellular distributions. Then we analyzed large cohorts of pathologically confirmed AD cases and controls (N = 613) and verified that it predicts cellular distribution patterns consistent with neurodegeneration. Finally, we generated RNA-seq from the parietal lobe of participants from the Knight-ADRC [39], including non-demented controls, LOAD cases, with enriched proportions of carriers of high-risk coding variants associated with AD, and also ADAD from the Dominantly Inherited Alzheimer Network [23] (DIAN). We compared the cell composition in ADAD and LOAD; and also evaluated differences among carriers of coding high-risk variants in *PLD3, TREM2* and *APOE* ε4. Our findings indicate that cell-type composition differs among carriers of specific genetic risk factors, which might be revealing distinct pathogenic mechanisms contributing to disease etiology.

Materials and methods

Subjects and Samples

DIAN and Knight-ADRC

Parietal lobe tissue of post-mortem brain was obtained with informed consent for research use and were approved by Washington University in St. Louis review board. RNA was extracted from frozen brain using Tissue Lyser LT and RNeasy Mini Kit (Qiagen, Hilden, Germany). RNA-seq Paired end reads with read length of 2×150bp were generated using Illumina HiSeq 4000 with a mean coverage of 80 million reads per sample (**Table 1**; **Table S1**). RNA-seq was generated for 19 brains from The Dominantly Inherited Alzheimer Network (DIAN), 84 brains with late-onset AD and 16 non-demented controls from The Charles F. and Joanne Knight Alzheimer's Disease Research Center (Knight ADRC) [39]. The clinical status of participants was neuropathologically confirmed [47]. We identified three additional participants from the Knight ADRC study with PSEN1 (A79V, I143T, S170F) mutations. CDR scores were obtained during regular visits throughout the study prior to the subject's decease [48]. A range of other pathological measurement were collected during autopsy including Braak staging, as previously described [11].

RNA was extracted from frozen brain tissues using Tissue Lyser LT and RNeasy Mini Kit (Qiagen, Hilden, Germany) following the manufacturer's instruction. RIN (RNA integrity) and DV200 were measured with RNA 6000 Pico Assay using Bioanalyzer 2100 (Agilent Technologies). The RIN is determined by the software on the Bioanalyzer taking into account the entire electrophoretic trace of the RNA including the presence or absence of degradation products. The DV200 value is defined as the percentage of nucleotides greater than 200nt. RIN and DV200 for all the samples can be found on **Table S1**. The yield of each sample is determined by the Quant-iT RNA Assay (Life Technologies) on the Qubit Fluorometer (Fisher Scientific). The cDNA library was prepared with the TruSeq Stranded Total RNA Sample Prep with Ribo-Zero Gold kit (Illumina) and then sequenced by HiSeq 4000 (Illumina) using 2×150 paired end reads at McDonnell Genome Institute, Washington University in St. Louis with a mean of 58.14 ± 8.62 million reads. Number of reads and other QC metrics can be found in **Table S1**.

Mayo Clinic Brain Bank

Mayo Clinic Brain Bank RNA-seq was accessed from the AMP-AD portal (synapse ID = 5550404; accessed January 2017) (**Table 1**). Paired end reads of 2×101 base pairs were generated by Illumina HiSeq 2000 sequencers for an average of 134.9 million reads per sample. Neuropathology criteria, quality control procedures, RNA extraction and sequencing details are explained elsewhere [8].

RNA-seq based transcriptome data was generated from post-mortem brain tissue collected from cerebellum (189 samples) and temporal cortex (191 samples) of Caucasian subjects [2, 8]. RNA was extracted using Trizol® reagent and cleaned with Qiagen RNeasy. RIN measurement was performed with Agilent Technologies 2100 Bioanalyzer. Samples with RIN greater than 5 were included. Library was prepared by Mayo Clinic Medical Genome Facility Gene Expression and Sequencing Cores with TruSeq RNA Sample Prep Kit (Illumina).

Mount Sinai Brain Bank

Mount Sinai Brain Bank RNAseq study was downloaded from the AMP-AD portal (synapse ID = 3157743; accessed January 2017) (**Table 1**). Single end reads of 100 nucleotides was generated by Illumina HiSeq 2500 System (Illumina, San Diego, CA) for an average of 38.7 million reads per sample [5].

This dataset contains 1030 samples collected from four post-mortem brain regions of 300 subjects: anterior prefrontal cortex (BA10), superior temporal gyrus (BA22), parahippocampal gyrus (BA36), and inferior frontal gyrus (BA44). RNAseq was generated using the TruSeq RNA Sample Preparation Kit v2 and Ribo-Zero rRNA removal kit (Illumina, San Diego, CA) [3].

iPSC-derived neurons

We have generated and characterized human iPSC made from human fibroblasts using non-integrating Sendai virus carrying OCT3/4, SOX2, KLF4, and cMYC [65, 68]. iPSCs were plated in a v-bottom plate in neural induction media (StemCell Technologies; 65,000 per well) to form highly uniform neural aggregates. After 5 days, neural aggregates were transferred onto PLO/laminin-coated tissue culture plates. Neural rosettes formed over 5-7 days. The resulting neural rosettes were then isolated by enzymatic selection (StemCell Technologies) and cultured as neural progenitor cells (NPCs). NPCs were then differentiated culturing in neural maturation medium (neurobasal medium supplemented with B27, GDNF, BDNF, cAMP).

TRAP-seq mice

All animal procedures were performed in accordance with the guidelines of Washington University's Institutional Animal Care and Use Committee. The Rosa26^{fsTRAP} mice (Gt(ROSA)26Sor$^{tm1(CAG-EGFP/Rpl10a,-birA)Wtp}$) [76] (The Jackson Laboratory) were crossed with PV^{Cre} mice (Pvalb$^{tm1(cre)Arbr}$) [35] (The Jackson Laboratory) to produce PV-TRAP mice directing expression of EGFP-L10a ribosomal fusion protein in parvalbumin (PV) expressing cells.
Purification of cell-type specific mRNA by translating ribosome affinity purification (TRAP) was described previously [34] with modifications. Briefly, PV-TRAP mouse brain was removed and quickly washed in ice-cold dissection buffer (1× HBSS, 2.5 mM HEPES-KOH (pH 7.3), 35 mM glucose, and 4 mM $NaHCO_3$ in RNase-free water). Barrel cortex was rapidly dissected and flash-frozen in liquid nitrogen, and then stored at -80 °C until use. Affinity matrix was prepared with 150 µl of Streptavidin MyOne T1 Dynabeads, 60 µg of Biotinylated Protein L, and 25 µg of each of GFP antibodies 19C8 and 19F7. The tissue was homogenized on ice in 1 ml of tissue-lysis buffer (20 mM HEPES KOH (pH 7.4), 150 mM KCl, 10 mM $MgCl_2$, EDTA-free protease inhibitors, 0.5 mM DTT, 100 µg/ml cycloheximide, and 10 µl/ml rRNasin and Superasin). Homogenates were centrifuged for 10 min at 2,000 × *g*, 4 °C, and 1/9 sample volume of 10% NP-40 and 300 mM DHPC were added to the supernatant at final concentration of 1% (vol/vol). After incubation on ice for 5 min, the lysate was centrifuged for 10 min at 20,000 × *g* to pellet insolubilized material. Then 200 µl of freshly resuspended affinity matrix was added to the supernatant and incubated at 4 °C for 16–18 hours with gentle end-over-end mixing in a tube rotator. After incubation, the beads were collected with a magnet and resuspended in 1000 µl of high-salt buffer (20 mM HEPES KOH (pH 7.3), 350 mM KCl, 10 mM $MgCl_2$, 1% NP-40, 0.5 mM DTT and 100 µg/ml cycloheximide), and collected with magnet as above. After 4 times of washing with high-salt buffer, RNA was extracted using Absolutely RNA Nanoprep Kit (Agilent Technologies) following manufacturer's instruction. RNA quantification was measured using Qubit RNA HS Assay Kit (Life Technologies) and the integrity was determined by Bioanalyzer 2100 using an RNA Pico chip (Agilent Technologies). The cDNA library was prepared with Clontech SMARTer and then sequenced by HiSeq3000. Single end reads of 50 base pairs were generated for an average of 29.2 million reads per sample (24 samples).

iPSC-derived microglia

The data was accessed from the AMP-AD portal (Synapse ID syn7203233). Myeloid progenitors expressing CD14/CX3CR1 were generated within 30 days of differentiation. iPSC-derived microglia were able to phagocytose and responded to ADP by producing intracellular Ca^{2+} transients, whereas macrophages lacked such response. The differentiation protocol was highly reproducible across several induced pluripotent stem cell (iPSC) lines.

RNA-seq QC and Alignment

FastQC was applied to DIAN and Knight-ADRC RNAseq data to perform quality check on various aspects of sequencing quality [58]. The DIAN and Knight-ADRC dataset was aligned to human GRCh37 primary assembly using Star (ver 2.5.2b) [24]. We used the primary assembly and aligned reads to the assembled chromosomes, un-localized and unplaced scaffolds, and discarded alternative haploid sequences. Sequencing metrics, including coverage, distribution of reads in the genome [4], ribosomal and mitochondrial contents and alignment quality, were further obtained by applying Picard CollectRnaSeqMetrics (ver 2.8.2) to detect sample deviation. Additional QC metrics can be found in **Table S1**.

Aligned and sorted bam files were loaded into IGV [55] to perform visual inspection of target variants. Samples carrying unexpected variants or missing expected variants were labeled as potential swapped samples. In addition, variants were called from RNA-seq following BWA/GATK pipeline [44, 45]. The identity of the samples was later verified by performing IBD analysis against genomic typing from GWAS chipsets.

Expression quantification

We applied Salmon transcript expression quantification (ver 0.7.2) [54] to infer the gene expression for all samples included in the reference panel and participants in the Mayo, MSBB, DIAN and Knight-ADRC. We quantified the coding transcripts of *Homo Sapiens* included in the GENCODE reference genome (GRCh37.75). Similarly, we quantified the expression of the mice samples included in the reference panel using the Mus Musculus reference genome (mm10).

Reference Panel

We assembled a cell-type specific reference panel from publicly available RNA-seq datasets comprised of both immunopanning collected or iPSC derived neurons, astrocytes, oligodendrocytes, and microglial cells from human and murine samples. For immunopanning collected cells, antibodies for cell-type specific antigens were utilized to bind and immobilize their targeted cell types in order to immunoprecipitate and purify each cell type from the suspensions [73]. cDNA synthesis was accomplished using Ovation RNA-seq system V2 (Nugen 7102) and library prepared with Next Ultra RNA-seq library prep kit from Illumina (NEB E7530) and NEBNext® multiplex oligos from Illumina (NEB E7335 E7500). TruSeq RNA Sample Prep Kit (Illumina) was used to prepare library for paired-end sequence on 100ng of total RNA extracted from each sample. Illumina HiSeq 2000 Sequencer was used to sequence all libraries [73].

Both human adult temporal cortex tissue, collected from patients receiving neurological surgeries, and mice cells were disassociated, sorted and sequenced as described elsewhere [74], and deposited in the Gene Expression Omnibus GSE73721 and GSE52564. We also accessed neural progenitor cells (day 17) and mature human neurons (day 57 and 100) from Broad iPSC deposited in the AMP-AD portal [6] and neural progenitor cells and iPSC-derived neurons from [12]. Broad iPSC derived neurons accessed from AMP-AD portal were generated using an embryoid body-based protocol to differentiate into forebrain neurons [1]. Wild-type cells used in the protocol were obtained from UConn StemCell Core. RNA was purified using PureLink RNA mini-kit (Life Technologies) and libraries were prepared by Broad Institute's Genomics Platform using TruSeq protocol. Please refer to **Table S2** for additional information.

Gene markers

The reference panel was assembled with samples from four distinct cell types. A redundant set of well-known cell-type markers was selected from the literature [74] (**Table S3**). Principal component analysis was performed on the reference panel using R function *prcomp* (version 3.3.3) to verify that the expressions of these gene were clustering samples by their cell types (**Fig S1b; Fig S2a**).

Inference of the cellular population structure

We ascertained alternative computation deconvolution algorithms implemented in the CellMix package (ver 1.6). Based on accuracy and robustness evaluation results we compared and reported the following three algorithms that outperformed the others: Digital Sorting Algorithm (named "DSA") [75], which employs linear modeling to infer cell distributions; the method population-specific expression analysis (PSEA, also named meanProfile in CellMix implementation) [40] that calculates estimated expression profiles relative to the average of the marker gene list for each cell type [40]; and a semi-supervised learning method that employs non-negative matrix factorization (ssNMF in CellMix implementation) [29]. We tested additional methods which provided considerably lower accuracy (least-squares fit [7], quadratic programing [31]) or no significant difference (support vector regression [51] or latent variable analysis [17]) to the methods presented.

We selected the samples that provide the most faithful transcriptomic profile for their respective cell types by following a leave-one-out cross validation approach. We trained iteratively deconvolution models using all but one of the samples that was tested. Only samples predicted with a composition higher than 80% were kept for the reference panel (**Table S2; Fig S2b**).

Accuracy and Robustness Evaluation

Chimeric validation

To emulate heterogeneous tissue with known and controlled cellular composition, we generated chimeric libraries pooling reads (to a total of 400,000) contributed from cell-type specific human donor samples (See **Table S2**). This process was repeated 720 times, using alternative samples from the reference panel to model each cell type. The proportion of reads that the libraries of neurons, astrocytes, oligodendrocytes and microglia provided to the chimeric libraries varied in predefined ranges (**Fig S3**). As a result, each of the chimeric libraries contained reads that followed 32 different distributions (neuronal reads contributed between 2 to 36% of reads, astrocytes between 22 to 76%, oligodendrocytes between 6 to 62% and microglia between 1 to 5%). Refer to **Table S4** for detailed description of the 32 different distributions. We quantified the chimeric reads using Salmon (v0.7.2) [54], and employed the samples that did not contribute reads to the chimeric library as reference panel for the deconvolution methods.

Overall, we quantified the expression of 23,040 (720 × 32) chimeric libraries. We evaluated the accuracy using the root-mean-square error (RMSE, **Equation 1** to compare the digital deconvolution cellular proportion estimates (method ssNMF) versus the defined proportion of reads specific to each of the chimeric libraries:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}i - yi)^2}{n}} \qquad \textbf{(Equation 1)}$$

$$\hat{y}i - estimated\ value, yi - observed\ value$$

We also tested whether the deconvolution results were dominated by the expression of any specific gene, and ascertained the robustness of the transcriptomic signature we modeled for each cell type to any possibly altered gene expression. To do so, we performed the deconvolution analysis discarding one gene of the reference panel at a time and evaluated how these distributions differed in comparison to the full gene reference panel.

Statistical Analysis

We employed linear regression models to test the association between cell-type proportions and disease status (R Foundation for Statistical Computing, ver.3.3.3). We used stepwise discriminant analysis (stepAIC function of R package MASS, version 7.3-45) to determine significant covariates, and correct for confounding effects. We included RNA integrity number (RIN), batch, age at death and post-mortem interval (PMI) as covariates for the Mayo Clinic analyses. For Mount Sinai Brain Bank analyses, we corrected for RIN, PMI, race, batch and age at death. We also used linear-mixed models to perform multiple-region association analysis, employing random slopes and random intercepts grouping by observations and by donors [64], and correcting for the same covariates previously described.

To analyze the DIAN and Knight-ADRC studies we applied linear-mixed models (function lmer and Anova, R packages lme4 ver.1.1 and car ver.2.1, respectively), clustering at family level to ascertain the effect of the neuropathological status in the cell proportion, and corrected for RIN and PMI. For late-onset specific analyses we also corrected for age at death.

Cellular composition shown as proportions were plotted using R package ggplot2 (ver 2.2.1)

RESULTS

Study design

To infer cellular composition from RNA-seq, we firstly assembled a gene reference panel for neurons, astrocytes, oligodendrocytes and microglia. The panel was created by analyzing expression data from purified cell lines. We evaluated alternative digital deconvolution methods and selected the best performing for our primary analyses. We tested the digital deconvolution accuracy on induced pluripotent stem cell (iPSC) derived neurons/microglia cells and neuronal Translating Ribosome Affinity Purification followed by RNA-seq (TRAP-seq; **Fig 1**). Finally, we verified its accuracy by creating artificial admixture with pre-defined cellular proportions.

Once the deconvolution approach was optimized, we calculated the cell proportion in AD cases and controls from the different brain regions of Mayo and MSBB datasets. The RNA-seq data for the Mayo Clinic study (N = 191) [8] and Mount Sinai (MSSM) Brain Bank (MSBB; N = 300) [5] are deposited in the Advanced Medicines Partnership-AD (AMP-AD) knowledge portal (Synapse ID: syn5550404 and syn3157743; **Table 1**). The Mayo study includes RNA-seq from the temporal cortex and cerebellum for AD affected and non-demented controls, in addition to pathological aging participants (**Fig 1**). The MSBB also profiled four additional cerebral cortex areas: anterior prefrontal cortex - APC, superior temporal gyrus

- STG, parahippocampal gyrus – PHG, and inferior frontal gyrus – IFG; **Table 1; Fig 1**). We restricted the case-control analysis to subjects with definite AD and autopsy confirmed controls. In addition, we generated RNA-seq from parietal lobe for participants of the Knight-ADRC (84 late-onset cases, carriers of genetic risk factors and 16 controls; **Table S1**) and the Dominantly Inherited Alzheimer Network (DIAN; 19 carriers of mutations in *APP*, *PSEN1*, *PSEN2*) (**Table 1; Fig 1).** We employed the same pipeline to process all of the samples in order to avoid any bias. Furthermore, RNA-seq from the Knight-ADRC and DIAN studies allowed us to compare the cell composition from ADAD vs LOAD brains, and similarly to test for differences in brain of controls, sporadic AD who do not carry any known high-risk variant, carriers of high-risk variants in *TREM2* (N = 20), *PLD3* (N = 33), and *APOE* ε4 allele.

Development of a reference panel to estimate brain cellular population structure

Due to limited availability of brain cell-type specific transcriptomic data, we compiled samples from different sources, including single-population RNA-seq from mice and human (immunopan-purified oligodendrocytes, neurons, astrocytes and microglia and iPSC-derived neurons and astrocytes) (**Table S2**).

We first tried to create a transcriptome wide reference panel by selecting the genes that are differentially expressed among cell types [17, 28, 51]. However, the species heterogeneity of the reference samples we compiled ruled out this attempt, as the principal component analyses (PCA) showed that differences between the human and mice donor samples dominated the transcriptome-wide effect (**Fig S1a**). For this reason, we curated a list of genes that have been described to tag these distinct cell types [14, 36, 74]. A visual inspection of the expression of these genes in the samples we compiled suggested a divergent transcriptomic profile among the cell types (**Fig S2a**). The PCA showed that their expression was sufficient to cluster samples of neurons, astrocytes, oligodendrocytes and microglia with their respective cell types, regardless of the species of the reference samples (**Fig S1b; Table S3**). We observed that some samples did not cluster with their expected cell types, and coincidently the leave-one-out cross-validation indicated that these samples had an expression signature that differed from the other samples of the same cell type. However, we found that all of these outliers correspond to samples not correctly purified or that were sequenced in early stages of differentiation (**Supplementary Results**). After discarding these samples, we assessed six digital deconvolution algorithms implemented in the CellMix package [28] and found that the semi-supervised non-negative matrix factorization [29] (ssNMF) calculated the most accurate estimates (see **Materials and methods).** Our final reference panel had a very high confidence to predict cell types with a mean predicted accuracy = 95.2%; s.d. = 4.3 (**Fig S2b**), and a root-mean-square error (RMSE) = 0.06 (**Table S5**).

Optimization, validation and accuracy estimation of the reference panel and digital deconvolution method

Once we identified the optimal approach to perform digital deconvolution from brain RNA-seq, we benchmarked it by using three sets of independent pure cell populations and simulated chimeric libraries.

We firstly validated the accuracy to predict neuronal composition by generating RNA-seq for eight iPSC-derived cortical neurons (see **Materials and methods**). We observed an accurate prediction in these independent cell lines (mean neuronal proportion = 94.8% and s.d. = 1.1%; **Fig S4a**). We also ascertained the cellular composition of mRNA extracted from the barrel cortex neurons isolated by Translating Ribosome Affinity Purification (TRAP) in 24 mice. TRAP is a method that captures cell-type specific mRNA translation by purifying tagged ribosomal subunit and capturing the mRNA it bound to [34]. We observed an average of neuronal proportion = 96.7% and s.d. = 1.2% (**Fig S4b**). Similarly, we assessed the RNA-seq data generated for iPSC-derived microglia (N = 10) deposited in the AMP-AD portal (Synapse ID: -syn7203233) and inferred their cellular population structure, and observed a mean microglia proportion = 86.6% and s.d. = 7.1% (**Fig S4c**).

To evaluate the accuracy of digital deconvolution for measuring cell-type proportion from cell-type admixtures, we simulated RNA-seq libraries by pooling reads from individual cell types into well-defined proportions. We combined randomly sampled reads from neurons, astrocytes, oligodendrocytes and microglia to create chimeric libraries that mimic bulk RNA-seq from brain, but with a range of pre-defined cell-type distributions (**Fig S3**). We then quantified the gene expression for the chimeric libraries and inferred the cell-type distribution (employing for the reference panel samples that did not contribute reads to the chimeric libraries). This process was repeated 23,040 times, choosing distinct human samples to represent each cell type and varying the proportions in 32 alternative distributions (See methods and **Table S4**). The overall error (RMSE) compared to known proportions = 0.08.

Finally, we evaluated whether any gene included in the reference panel was dominating the inference of cell proportions. We re-calculated the cell-type distributions of the chimeric libraries, but dropping each of the genes from the reference panel one at a time. We observed a negligible difference between the cellular population structure inferred using the full reference and the gene-dropped panels (average RMSE = 0.022, s.d. < 0.01). In this way, we verified that the proportions inferred using the reference panel are not driven by the expression of a single gene. This reassured us the inference should be robust to any bias introduced by the potential association of a single gene included in the reference panel with a particular trait.

Deconvolution of bulk RNA-seq of non-demented and AD brains shows a characteristic signature for neurodegeneration

Pathologically, AD is associated with neuronal death and gliosis specifically in the cerebral cortex. We evaluated whether we could exploit deconvolution methods using our reference panel to detect altered cellular population structure from the bulk RNA-seq, and whether this corresponded to known pathological alterations.

We initially analyzed the RNA-seq from the Mayo Clinic Brain Bank that includes bulk RNA-seq from the temporal cortex (TC) and cerebellum (CB) for 191 participants [8] (**Table 1**). In the TC, we observed a significant increase of astrocyte ($\beta = 0.23$; $p = 5.01\times10^{-09}$; **Table 2; Fig 2; Table S6**) in AD brains compared to controls brains. We also found a significant decrease of neurons ($\beta = -0.17$; $p = 1.58\times10^{-07}$; **Table 2; Fig 2; Table S6)** and oligodendrocytes ($\beta = -0.07$; $p = 1.8\times10^{-02}$; **Table 2**; **Fig S5; Table S6**). As expected, given the absence of pathology, we did not observe a significant change in the cell-type composition in the CB (**Table 2**).

The distribution of microglia was similar in the TC and CB from AD and control brains (**Table 2**; **Fig S5**). The proportion of microglia was lower than any other cell types. The Mayo dataset also includes brains from individuals with pathological aging (PA; **Table 1**); which is neuropathologically defined by amyloid-beta (Aβ) senile plaque deposits but little or no neurofibrillary tau pathology [8, 49]. We observed a significant decrease of microglia proportion of PA brains compared to AD in both TC and CB (**Table S7; Fig S6**) [43]. Therefore, we speculated that the lack of changes in the AD microglial population was neither due to low statistical power nor the inability of our method to estimate the microglial proportions, but reflected unaltered neuropathological observations in AD brains.

We also analyzed data from the MSBB, which contains bulk RNA-seq for four additional cerebral cortex areas (APC, STG, PHG, IFG). Replicating our findings from the Mayo dataset we observed a significant decrease in neurons and increase in astrocytes in all four areas (**Table 2; Fig 2; and Table S6**). The strongest effect size was detected in the parahippocampal gyrus and superior temporal gyrus ($p < 3.49\times10^{-07}$) (**Table 2; Table S8**). Neuropathological studies have described that the parahippocampal gyrus in one of the first brain areas in which AD pathology occurs [10, 25, 69]. We also observed a significant and strong correlation between neuronal and astrocyte proportions and last ascertained clinical status (Clinical Dementia Rating - CDR), and number of amyloid plaques and Braak staging (**Table 2; Fig 2; Fig S7**).

The cellular population structure differs between ADAD vs LOAD

While the loss of neurons is a common feature of AD, it is not clear whether the mechanism holds true across different forms of AD or AD cases carrying different genetic risk variants. Therefore, we investigated whether AD with distinct etiologies showed different cellular compositions. We generated RNA-seq data from the parietal lobe of participants enrolled in Knight-ADRC (84 LOAD, 3 ADAD, and 16 neuropath-free controls) and DIAN (19 ADAD) studies (**Table 1; Table S1**). We selected the LOAD and ADAD participants to match for CDR at death, brain weight and sex distributions (See **Table S1**).

Using digital deconvolution, we determined the cellular composition for these brains. We observed a significant decrease in neurons ($\beta = -0.02$, $p = 2.66\times10^{-02}$) and significant increase in astrocytes in AD ($\beta = 0.03$, $p = 5.48\times10^{-03}$) for the combined LOAD and ADAD brains compared to controls (**Table 3; Fig 3; Table S9**), consistent with our findings in the Mayo and MSBB datasets. Similarly, the joint analysis of the brains from Knight-ADRC and DIAN showed a significant association between the neuronal and astrocyte proportions and neuropathological measures (Braak staging: $\beta = -0.03$, $p = 8.51\times10^{-06}$ for neurons and $\beta = 0.03$, $p = 3.83\times10^{-06}$ for astrocytes; **Table 3**; **Fig 3b**) as well as for clinical measures (CDR: $\beta = -0.02$, $p = 2.66\times10^{-02}$ for neurons and $\beta = 0.03$ and $p = 5.48\times10^{-03}$ for astrocytes; **Table 3**; **Fig 3c**). We did not observe a significant difference in the compositions of microglia or oligodendrocytes (**Table 3**; **Fig S8**).

Next, we compared the cell proportion of LOAD vs ADAD and found that the cell composition differs between them. We firstly selected the LOAD brains (N = 25) to match the Braak staging distribution of ADAD brains (N = 17). The ADAD brains showed a significant decreased neuronal proportion compared to LOAD brains ($\beta = -0.08$; $p = 1.03\times10^{-02}$; **Table 3**), and increased astrocytes ($\beta = 0.11$; $p = 9.26\times10^{-04}$; **Table 3**). Then, we analyzed the entire Knight-ADRC LOAD brains, by extending the model to correct for Braak stages. We also observed significant decreased neurons ($\beta = -0.09$; $p = 4.71\times10^{-03}$; **Table 3; Fig 3a; Table S9)** and increased astrocytes ($\beta = 0.11$; $p = 5.24\times10^{-04}$; **Table 3; Fig 3a; Table S9** in ADAD brains compared to LOAD. We observed the same cellular differences when we corrected for CDR at death ($\beta = -0.12$; $p = 2.11\times10^{-03}$ for neurons and $\beta = 0.13$; $p = 6.29\times10^{-04}$ for astrocytes; **Table 3; Fig 3bc**). In summary, our results indicate that ADAD individuals present a higher neuronal death even in the same stage of the disease, suggesting that in ADAD neuronal death play a more important role in pathogenesis than sporadic AD, in which other factors such as inflammation or immune response may be involved.

Specific genetic variants confer a distinctive cell composition profile

A variety of genetic variants increase risk of LOAD; however, it is unclear if the cellular mechanisms are the same across these distinct risk factors. Therefore, we tested the hypothesis that distinct genetic causes of LOAD have characteristic cellular population signatures.

We initially ascertained the effect of *APOE* ε4 on the cell-type composition. We observed a significant decrease in neurons ($\beta = -0.06$ for each of the ε4 alleles; $p = 9.91\times10^{-03}$) and increase of astrocytes ($\beta = 0.10$; $p = 4.15\times10^{-02}$) from the TC included in the Mayo Clinic dataset (**Table S10; Fig 4a; Fig S9a**). This finding was replicated when we performed a multi-area analysis of the MSBB dataset ($\beta = -0.03$; $p = 2.75\times10^{-04}$ and $\beta = 0.04$; $p = 8.06\times10^{-06}$ for neurons and astrocytes respectively; **Table 4; Fig 4a; Table S10; Fig S9a**). Given the strong risk conferred by the *APOE* ε4 allele [19], we studied its effects on the cell-type composition by restricting our analysis to AD brains. We observed a significant association in the multi-area analysis of the MSBB dataset, with the same effect size for the neurons as the observed when we analyzed both affected and control brains ($p = 1.60\times10^{-02}$; **Table 4; Fig 4b; Table S11; Fig S9b**) and also a significant increase in astrocytes ($\beta = 0.03$; $p = 1.03\times10^{-02}$; **Table 4; Fig 4b; Table S11; Fig S9b**). We also observed a significant decrease in neurons proportion ($\beta = -0.06$; $p = 2.11\times10^{-02}$; **Table 4**; **Fig 4c**) when we analyzed the LOAD and control brains from the Knight-ADRC. When we restricted the analysis to AD brains from the Knight-ADRC and compared the *APOE* ε4 carriers (N = 46) to non-carriers (N = 41) we also observed decreased neurons ($\beta = -0.06$; $p = 2.69\times10^{-02}$; **Table 4**; **Fig 4d**).

Next, we analyzed the cellular composition in *PLD3* carriers (N = 33). *PLD3* carriers exhibited significantly decrease of neurons compared to controls ($\beta = -0.10$; $p = 1.60\times10^{-04}$; **Fig 3d**) and a significant increase in astrocytes ($\beta = 0.13$; $p = 2.84\times10^{-03}$; **Table 4; Fig 3d**). Sporadic AD non-carriers cases also exhibited significantly decrease of neurons compared to controls ($\beta = -0.11$; $p = 5.45\times10^{-03}$) and significant increase of astrocytes ($\beta = 0.13$; $p = 2.95\times10^{-04}$; **Table 4; Fig 3d**). The cell proportion between sporadic AD non-carriers and *PLD3* carriers did not show any significantly difference ($p > 0.05$).

Finally, we performed similar analyses with *TREM2* carriers. *TREM2* is involved in the immune response and its role in amyloid-β deposition or clearance remain controversial [67]. Our analysis on the Knight-ADRC data showed significantly increased astrocytes in AD affected *TREM2* carriers (N = 20) compared to controls ($\beta = 0.11$; $p = 1.05\times10^{-02}$; **Table 4**; **Fig 3d**). Despite *TREM2* carriers presented lower neuron proportion compared to controls, this difference was not statistically significant ($p>0.05$; **Table 4**; **Fig 3d**). We analyzed whether the *TREM2* carriers provided sufficient power to detect a significant association. Our empirical estimates showed that *TREM2* sample size provides 96% of power to detect an association with an effect size comparable to that observed for sporadic AD ($\beta = -0.11$). We also investigated whether the cellular proportion of the eleven *TREM2* carriers in the MSBB dataset. The multi-region analysis showed *TREM2* carriers do not show a significant difference in neurons compared to controls ($p > 0.05$; **Table 4; Fig 4e**), whereas in the AD *TREM2* non-carriers the neuronal and astrocytic proportions are significantly different from controls ($\beta = -0.07$; $p = 1.91\times10^{-08}$ and $\beta = 0.08$; $p = 1.25\times10^{-08}$ respectively; **Table 4**; **Fig 4e**).

In fact, our analyses indicate that *TREM2* carriers have a unique cellular brain composition distinct than the other AD cases. *TREM2* brains showed significantly higher neurons ($\beta = 0.05$; $p = 1.98\times10^{-02}$) and significantly decreased astrocytes than the AD *non*-carries ($\beta = -0.05$; $p = 1.58\times10^{-02}$; **Table 4**). The distribution of CDR, mean number of amyloid plaques and Braak staging do not differ between strata. Nonetheless, we verified that the cellular proportions were still significant after correcting for each of those variables (**Table 4**). These results suggested that the mechanism that lead to disease in *TREM2* carriers is less neuron-centric than in the general AD population.

Discussion

We have developed, optimized and validated a digital deconvolution approach to infer cell composition from bulk brain gene expression that integrates publicly available cell-type specific expression data while addressing the heterogeneity of the phenotypic differences of samples and technical characteristics of transcriptome ascertainment. We acknowledge that the accuracy of this platform might be affected by the phenotypic diversity of the reference panel or the disease-induced dysregulation of genes it includes. However, the deconvolution approach proved to be robust to the genes included in the reference panel, as we demonstrated that the proportions it inferred are not driven by the expression of any single gene. This platform produced reliable cell proportion estimates, as was shown by the evaluation of independent datasets of iPSC-derived neurons and microglia, mice cortical neurons (**Fig S4**) and simulated chimeric libraries.

We used this approach to deconvolve studies that include large number of neuropathologically defined AD and control brains with their transcriptome ascertained in distinct brain regions, and observed consistently significant neuronal loss and astrocytosis in the cerebral cortex. Compatible with other studies, we also identified that the altered cellular proportion is also significantly associated with decline in cognition and Braak staging [60]. In contrast, we did not identify a significant difference in the cellular population structure in the cerebellum, a region not affected in AD (**Table 2**; **Fig 2a**).

We generated RNA-seq data from brains carrying pathogenic mutations in *APP*, *PSEN1*, *PSEN2*, which cause alterations in Aβ processing and lead to ADAD, and also generated RNA-seq from brains of LOAD and neuropath-free controls. We observed altered cell composition in both ADAD and LOAD compared to controls. However, we identified that ADAD brains have a different cell-type composition than disease-stage-matched LOAD, as the ADAD has a significantly lower neuronal proportion and more pronounced astrocytosis. Based on our results, we would hypothesize that this change in Aβ processing of ADAD would leads to more direct to neuronal death than the pathological processes of LOAD. Similarly, decreased neurons and increased astrocytes were significantly associated with *APOEε4* allele. It has been reported APOE ε4 allele increase the risk for AD by affecting APP metabolism or Aβ clearance [15, 38], suggesting a direct link between APP metabolism and neuronal death.

In contrast, the analysis of the Knight-ADRC brains showed that the neuronal loss is less pronounced in *TREM2* carriers than in other LOAD cases. We replicated this finding in a multi-area analysis from the MSBB dataset. These results may implicate that *TREM2* risk variants lead to a cascade of pathological events that differ from those occurring in sporadic AD cases, which is also consistent with the known biology of *TREM2*. *TREM2* is involved in AD pathology through microglia mediated pathways, implicated on altered immune response and inflammation [18]. Recent studies in *TREM2* knock-out animals showed that fewer microglia cells were found surrounding Aβ plaques with impaired microgliosis [70]. Furthermore, *TREM2* deficiency was reported to attenuate tauopathy against brain atrophy [42]. We found no significant difference in the proportion of microglia between AD cases and controls. However, we found significantly decreased microglia in brains exhibiting pathological aging (**Table S7; Fig S6**), proving that these studies are sufficiently powered to identify significant differences. In any case, we cannot rule out the possibility of a change in the activation stage of microglia in these individuals. Overall, these results suggest that *TREM2* affects AD risk through a slightly different mechanism to that of ADAD or LOAD in general. Therefore, other pathogenic mechanisms should contribute to disease. We believe that a detailed modeling of immune response cells, reflecting the alternative microglia activation states, will generate more accurate profiles to elucidate the immune cell distribution in AD.

There is a large interest in the scientific community to use brain expression studies to try to identity novel pathogenic mechanism in AD and to identify novel therapeutic targets. These efforts are generating a large amount of bulk RNA-seq data, as single-cell RNA (scRNA-seq) from human brain tissue in large sample size is not feasible. Single-cell sorting needs to be performed with fresh tissue [33], which restrains the analysis of highly characterized fresh-frozen brains collected by AD research centers. Our results indicate that digital deconvolution methods can accurately infer relative cell distributions from brain bulk RNA-seq data. Having this approach validated for AD can have an important impact in the community, because digital deconvolution analyses 1) can reveal distinct cellular composition patterns underlying different disease etiologies 2) can provide additional insights about the overall pathologic mechanisms underlying different mutations carriers for variants as in genes such as *TREM2*, *APOE, APP, PSEN1* and *PSEN2*) can correct the effect that altered cell composition and genetic statuses have in addition to downstream transcriptomic analyses and lead to novel and informative results. 4) can help the analysis of highly informative frozen brains collected over the years.

In conclusion, our study provides a reliable approach to enhance our understanding of the fundamental cellular mechanisms involved in AD and enable the analysis of large bulk RNA-seq data that may lead to novel discoveries and insights into neurodegeneration.

Acknowledgements

We thank all participants and their families for their commitment and dedication to helping advance research into the early detection and causation of AD; and the Knight-ADRC and DIAN research and support staff at each of the participating sites for their contributions to this study.

This work was supported by grants from the National Institutes of Health (R01-AG044546, P01-AG003991, RF1AG053303, R01-AG035083, and R01-NS085419), and the Alzheimer's Association (NIRG-11-200110). This research was conducted while CC was a recipient of a New Investigator Award in Alzheimer's disease from the American Federation for Aging Research. CC is a recipient of a BrightFocus Foundation Alzheimer's Disease Research Grant (A2013359S). This work was supported in part by NIH K01AG046374 awarded to CMK, and also the Tau Consortium (CMK). JDD is supported by the Brain and Behavior Research Foundation and the NIH (R01NS102272). The recruitment and clinical characterization of research participants at Washington University were supported by NIH P50 AG05681, P01 AG03991, and P01 AG026276.

Data collection and sharing for this project was supported by The Dominantly Inherited Alzheimer's Network (DIAN, U19AG032438) funded by the National Institute on Aging (NIA), the German Center for Neurodegenerative Diseases (DZNE), Raul Carrea Institute for Neurological Research (FLENI), Partial support by the Research and Development Grants for Dementia from Japan Agency for Medical Research and Development, AMED, and the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI).This manuscript has been reviewed by DIAN Study investigators for scientific content and consistency of data interpretation with previous DIAN Study publications. We acknowledge the altruism of the participants and their families and contributions of the DIAN research and support staff at each of the participating sites for their contributions to this study.

We would like to thank the operations staff at the Elizabeth H. and James S. McDonnell III Genome Institute at Washington University with their assistance in constructing the RNA-seq libraries and generating sequence data for our project. This work was also supported by accessing to equipment made possible by the Hope Center for Neurological Disorders and the Departments of Neurology and Psychiatry at Washington University School of Medicine. We also thank Allison M. Lake for her comments and suggestions.

The results published here are in whole or in part based on data obtained from the AMP-AD Knowledge Portal accessed at doi:10.7303/syn2580853. Study data were provided by the following sources: The Mayo Clinic Alzheimer's Disease Genetic Studies, led by Dr. Nilufer Taner and Dr. Steven G. Younkin, Mayo Clinic, Jacksonville, FL using samples from the Mayo Clinic Study of Aging, the Mayo Clinic Alzheimer's Disease Research Center, and the Mayo Clinic Brain Bank. Data collection was supported through funding by NIA grants P50 AG016574, R01 AG032990, U01 AG046139, R01 AG018023, U01 AG006576, U01 AG006786, R01 AG025711, R01 AG017216, R01 AG003949, NINDS grant R01 NS080820, CurePSP Foundation, and support from Mayo Foundation. Study data includes samples collected through the Sun Health Research Institute Brain and Body Donation Program of Sun City, Arizona. The Brain and Body Donation Program is supported by the National Institute of Neurological Disorders and Stroke (U24 NS072026 National Brain and Tissue Resource for Parkinson's Disease and Related Disorders), the National Institute on Aging (P30 AG19610 Arizona Alzheimer's Disease Core Center), the Arizona Department of Health Services (contract 211002, Arizona Alzheimer's Research Center), the Arizona Biomedical Research Commission (contracts 4001, 0011, 05-901 and 1001 to the Arizona Parkinson's Disease Consortium) and the Michael J. Fox Foundation for Parkinson's Research. These data were generated from postmortem brain tissue collected through the Mount Sinai VA Medical Center Brain Bank and were provided by Dr. Eric Schadt from Mount Sinai School of Medicine. These data were generated by Kristen Brennand, a New York Stem Cell Foundation - Robertson Investigator. This work was supported by the Brain and Behavior Research Foundation, NIH grant R01 MH101454 and the New York Stem Cell Foundation. We analyzed iPSC-derived microglia RNA-seq data funded by NIH U01AG046170.

References

1 AMPAD Knowledge Portal BroadiPSC RNAseq https://www.synapse.org/ - !Synapse:syn3607401

2 AMPAD Knowledge Portal Mayo Clinic RNAseq https://www.synapse.org/ - !Synapse:syn5550404

3 AMPAD Knowledge Portal Mount Sinai Brain Bank RNAseq https://www.synapse.org/ - !Synapse:syn3157743

4 Broad Institute The Picard Pipeline http://broadinstitute.github.io/picard/

5 Mount Sinai Brain Bank (MSBB) RNA-seq data deposited in the AMP-AD https://www.synapse.org/ - !Synapse:syn3157743

6 UConn StemCell Core Broad iPSC deposited in the AMP-AD https://www.synapse.org/ - !Synapse:syn36074012016

7 Abbas AR, Wolslegel K, Seshasayee D, Modrusan Z, Clark HF (2009) Deconvolution of blood microarray data identifies cellular activation patterns in systemic lupus erythematosus. PLoS One 4: e6098 Doi 10.1371/journal.pone.0006098

8 Allen M, Carrasquillo MM, Funk C, Heavner BD, Zou F, Younkin CS et al (2016) Human whole genome genotype and transcriptome data for Alzheimer's and other neurodegenerative diseases. Sci Data 3: 160089 Doi 10.1038/sdata.2016.89

9 Benitez BA, Cruchaga C (2013) TREM2 and neurodegenerative disease. N Engl J Med 369: 1567-1568 Doi 10.1056/NEJMc1306509#SA4

10 Braak H, Braak E (1990) Neurofibrillary changes confined to the entorhinal region and an abundance of cortical amyloid in cases of presenile and senile dementia. Acta Neuropathol 80: 479-486

11 Braak H, Braak E (1995) Staging of Alzheimer's disease-related neurofibrillary changes. Neurobiol Aging 16: 271-278; discussion 278-284

12 Brennand KJ The hiPSC Neurons and NPCs study (MSSMiPSC) deposited in the AMP-AD.

13 Brennand KJ, Simone A, Jou J, Gelboin-Burkhart C, Tran N, Sangar S et al (2011) Modelling schizophrenia using human induced pluripotent stem cells. Nature 473: 221-225 Doi 10.1038/nature09915

14 Cahoy JD, Emery B, Kaushal A, Foo LC, Zamanian JL, Christopherson KS et al (2008) A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. J Neurosci 28: 264-278 Doi 10.1523/jneurosci.4178-07.2008

15 Castellano JM, Kim J, Stewart FR, Jiang H, DeMattos RB, Patterson BW et al (2011) Human apoE isoforms differentially regulate brain amyloid-beta peptide clearance. Sci Transl Med 3: 89ra57 Doi 10.1126/scitranslmed.3002156

16 Chan G, White CC, Winn PA, Cimpean M, Replogle JM, Glick LR et al (2015) CD33 modulates TREM2: convergence of Alzheimer loci. Nat Neurosci 18: 1556-1558 Doi 10.1038/nn.4126

17 Chikina M, Zaslavsky E, Sealfon SC (2015) CellCODE: a robust latent variable approach to differential expression analysis for heterogeneous cell populations. Bioinformatics 31: 1584-1591 Doi 10.1093/bioinformatics/btv015

18 Colonna M (2003) TREMs in the immune system and beyond. Nat Rev Immunol 3: 445-453 Doi 10.1038/nri1106

19 Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GW et al (1993) Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science 261: 921-923

20 Cruchaga C, Karch CM, Jin SC, Benitez BA, Cai Y, Guerreiro R et al (2014) Rare coding variants in the phospholipase D3 gene confer risk for Alzheimer's disease. Nature 505: 550-554 Doi 10.1038/nature12825

21 De Strooper B, Annaert W (2010) Novel research horizons for presenilins and gamma-secretases in cell biology and disease. Annu Rev Cell Dev Biol 26: 235-260 Doi 10.1146/annurev-cellbio-100109-104117

22 Del-Aguila JL, Fernandez MV, Jimenez J, Black K, Ma SM, Deming Y et al (2015) Role of ABCA7 loss-of-function variant in Alzheimer's disease: a replication study in European-Americans. Alzheimers Research & Therapy 7: Doi ARTN 7310.1186/s13195-015-0154-x

23 DIAN Dominantly Inherited Alzheimer Network http://www.dian-info.org/. Accessed 2017-05-10

24 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S et al (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29: 15-21 Doi 10.1093/bioinformatics/bts635

25 Echavarri C, Aalten P, Uylings HB, Jacobs HI, Visser PJ, Gronenschild EH et al (2011) Atrophy in the parahippocampal gyrus as an early biomarker of Alzheimer's disease. Brain Struct Funct 215: 265-271 Doi 10.1007/s00429-010-0283-8

26 Fernandez MV, Black K, Carrell D, Saef B, Budde J, Deming Y et al (2016) SORL1 variants across Alzheimer's disease European American cohorts. Eur J Hum Genet 24: 1828-1830 Doi 10.1038/ejhg.2016.122

27 Gaiteri C, Mostafavi S, Honey CJ, De Jager PL, Bennett DA (2016) Genetic variants in Alzheimer disease - molecular and brain network approaches. Nat Rev Neurol 12: 413-427 Doi 10.1038/nrneurol.2016.84

28 Gaujoux R, Seoighe C (2013) CellMix: a comprehensive toolbox for gene expression deconvolution. Bioinformatics 29: 2211-2212 Doi 10.1093/bioinformatics/btt351

29 Gaujoux R, Seoighe C (2012) Semi-supervised Nonnegative Matrix Factorization for gene expression deconvolution: a case study. Infect Genet Evol 12: 913-921 Doi 10.1016/j.meegid.2011.08.014

30 Golub VM, Brewer J, Wu X, Kuruba R, Short J, Manchi M et al (2015) Neurostereology protocol for unbiased quantification of neuronal injury and neurodegeneration. Front Aging Neurosci 7: 196 Doi 10.3389/fnagi.2015.00196

31 Gong T, Hartmann N, Kohane IS, Brinkmann V, Staedtler F, Letzkus M et al (2011) Optimal deconvolution of transcriptional profiling data using quadratic programming with application to complex clinical blood samples. PLoS One 6: e27156 Doi 10.1371/journal.pone.0027156

32 Guerreiro R, Wojtas A, Bras J, Carrasquillo M, Rogaeva E, Majounie E et al (2013) TREM2 variants in Alzheimer's disease. N Engl J Med 368: 117-127 Doi 10.1056/NEJMoa1211851

33 Habib N, Avraham-Davidi I, Basu A, Burks T, Shekhar K, Hofree M et al (2017) Massively parallel single-nucleus RNA-seq with DroNc-seq. Nat Methods 14: 955-958 Doi 10.1038/nmeth.4407

34 Heiman M, Kulicke R, Fenster RJ, Greengard P, Heintz N (2014) Cell type-specific mRNA purification by translating ribosome affinity purification (TRAP). Nature protocols 9: 1282-1291 Doi 10.1038/nprot.2014.085

35 Hippenmeyer S, Vrieseling E, Sigrist M, Portmann T, Laengle C, Ladle DR et al (2005) A developmental switch in the response of DRG neurons to ETS transcription factor signaling. PLoS biology 3: e159 Doi 10.1371/journal.pbio.0030159

36 Holtman IR, Raj DD, Miller JA, Schaafsma W, Yin Z, Brouwer N et al (2015) Induction of a common microglia gene expression signature by aging and neurodegenerative conditions: a co-expression meta-analysis. Acta Neuropathol Commun 3: 31 Doi 10.1186/s40478-015-0203-5

37 Holtzman DM, Morris JC, Goate AM (2011) Alzheimer's disease: the challenge of the second century. Sci Transl Med 3: 77sr71 Doi 10.1126/scitranslmed.3002369

38 Kim J, Basak JM, Holtzman DM (2009) The role of apolipoprotein E in Alzheimer's disease. Neuron 63: 287-303 Doi 10.1016/j.neuron.2009.06.026

39 KnightADRC Knight-Alzheimer’s Disease Research Center http://alzheimer.wustl.edu/

40 Kuhn A, Thu D, Waldvogel HJ, Faull RL, Luthi-Carter R (2011) Population-specific expression analysis (PSEA) reveals molecular changes in diseased brain. Nat Methods 8: 945-947 Doi 10.1038/nmeth.1710

41 LaFerla FM, Oddo S (2005) Alzheimer's disease: Abeta, tau and synaptic dysfunction. Trends Mol Med 11: 170-176

42 Leyns CEG, Ulrich JD, Finn MB, Stewart FR, Koscal LJ, Remolina Serrano J et al (2017) TREM2 deficiency attenuates neuroinflammation and protects against neurodegeneration in a mouse model of tauopathy. Proceedings of the National Academy of Sciences: 201710311 Doi 10.1073/pnas.1710311114

43 Leyns CEG, Ulrich JD, Finn MB, Stewart FR, Koscal LJ, Remolina Serrano J et al (2017) TREM2 deficiency attenuates neuroinflammation and protects against neurodegeneration in a mouse model of tauopathy. Proc Natl Acad Sci U S A 114: 11524-11529 Doi 10.1073/pnas.1710311114

44 Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25: 1754-1760 Doi 10.1093/bioinformatics/btp324

45 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A et al (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20: 1297-1303 Doi 10.1101/gr.107524.110

46 Miller JA, Woltjer RL, Goodenbour JM, Horvath S, Geschwind DH (2013) Genes and pathways underlying regional and cell type changes in Alzheimer's disease. Genome Med 5: 48 Doi 10.1186/gm452

47 Mirra SS, Heyman A, McKeel D, Sumi SM, Crain BJ, Brownlee LM et al (1991) The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part II. Standardization of the neuropathologic assessment of Alzheimer's disease. Neurology 41: 479-486

48 Morris JC (1997) Clinical dementia rating: a reliable and valid diagnostic and staging measure for dementia of the Alzheimer type. Int Psychogeriatr 9 Suppl 1: 173-176; discussion 177-178

49 Murray ME, Dickson DW (2014) Is pathological aging a successful resistance against amyloid-beta or preclinical Alzheimer's disease? Alzheimers Res Ther 6: 24 Doi 10.1186/alzrt254

50 Narayanan M, Huynh JL, Wang K, Yang X, Yoo S, McElwee J et al (2014) Common dysregulation network in the human prefrontal cortex underlies two neurodegenerative diseases. Molecular systems biology 10: 743 Doi 10.15252/msb.20145304

51 Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y et al (2015) Robust enumeration of cell subsets from tissue expression profiles. Nat Methods 12: 453-457 Doi 10.1038/nmeth.3337

52 Paduraru M, Ciobica A, Mavroudis I, Fotiou D, Baloyannis S (2012) Hippocampal neuronal loss in the CA1 and CA3 areas of Alzheimer's disease patients. Psychiatria Danubina 24: 152-158

53 Parikshak NN, Gandal MJ, Geschwind DH (2015) Systems biology and gene networks in neurodevelopmental and neurodegenerative disorders. Nat Rev Genet 16: 441-458 Doi 10.1038/nrg3934

54 Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C (2017) Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods 14: 417-419 Doi 10.1038/nmeth.4197

55 Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G et al (2011) Integrative genomics viewer. Nature biotechnology 29: 24-26 Doi 10.1038/nbt.1754

56 Rogaeva E, Meng Y, Lee JH, Gu Y, Kawarai T, Zou F et al (2007) The neuronal sortilin-related receptor SORL1 is genetically associated with Alzheimer disease. Nat Genet 39: 168-177

57 Ryan NS, Nicholas JM, Weston PS, Liang Y, Lashley T, Guerreiro R et al (2016) Clinical phenotype and genetic associations in autosomal dominant familial Alzheimer's disease: a case series. Lancet Neurol 15: 1326-1335 Doi 10.1016/S1474-4422(16)30193-4

58 S. A (2010) FastQC: a quality control tool for high throughput sequence data. City

59 Selkoe DJ (2001) Alzheimer's disease: genes, proteins, and therapy. Physiol Rev 81: 741-766

60 Serrano-Pozo A, Frosch MP, Masliah E, Hyman BT (2011) Neuropathological alterations in Alzheimer disease. Cold Spring Harbor perspectives in medicine 1: a006189 Doi 10.1101/cshperspect.a006189

61 Shen-Orr SS, Gaujoux R (2013) Computational deconvolution: extracting cell type-specific information from heterogeneous samples. Current opinion in immunology 25: 571-578 Doi 10.1016/j.coi.2013.09.015

62 Srinivasan K, Friedman BA, Larson JL, Lauffer BE, Goldstein LD, Appling LL et al (2016) Untangling the brain's neuroinflammatory and neurodegenerative transcriptional responses. Nature communications 7: 11295 Doi 10.1038/ncomms11295

63 Steinberg S, Stefansson H, Jonsson T, Johannsdottir H, Ingason A, Helgason H et al (2015) Loss-of-function variants in ABCA7 confer risk of Alzheimer's disease. Nat Genet 47: 445-447 Doi 10.1038/ng.3246

64 Sul JH, Han B, Ye C, Choi T, Eskin E (2013) Effectively identifying eQTLs from multiple tissues by combining mixed model and meta-analytic approaches. PLoS Genet 9: e1003491 Doi 10.1371/journal.pgen.1003491

65 Takahashi K, Yamanaka S (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126: 663-676 Doi 10.1016/j.cell.2006.07.024

66 Tang M, Ryman DC, McDade E, Jasielec MS, Buckles VD, Cairns NJ et al (2016) Neurological manifestations of autosomal dominant familial Alzheimer's disease: a comparison of the published

literature with the Dominantly Inherited Alzheimer Network observational study (DIAN-OBS). Lancet Neurol 15: 1317-1325 Doi 10.1016/S1474-4422(16)30229-0

67 Ulrich JD, Ulland TK, Colonna M, Holtzman DM (2017) Elucidating the Role of TREM2 in Alzheimer's Disease. Neuron 94: 237-248 Doi 10.1016/j.neuron.2017.02.042

68 van de Leemput J, Boles NC, Kiehl TR, Corneo B, Lederman P, Menon V et al (2014) CORTECON: a temporal transcriptome analysis of in vitro human cerebral cortex development from human embryonic stem cells. Neuron 83: 51-68 Doi 10.1016/j.neuron.2014.05.013

69 Van Hoesen GW, Augustinack JC, Dierking J, Redman SJ, Thangavel R (2000) The parahippocampal gyrus in Alzheimer's disease. Clinical and preclinical neuroanatomical correlates. Ann N Y Acad Sci 911: 254-274

70 Wang Y, Ulland TK, Ulrich JD, Song W, Tzaferis JA, Hole JT et al (2016) TREM2-mediated early microglial response limits diffusion and toxicity of amyloid plaques. J Exp Med 213: 667-675 Doi 10.1084/jem.20151948

71 Wright AL, Zinn R, Hohensinn B, Konen LM, Beynon SB, Tan RP et al (2013) Neuroinflammation and neuronal loss precede Abeta plaque deposition in the hAPP-J20 mouse model of Alzheimer's disease. PLoS One 8: e59586 Doi 10.1371/journal.pone.0059586

72 Zhang B, Gaiteri C, Bodea LG, Wang Z, McElwee J, Podtelezhnikov AA et al (2013) Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. Cell 153: 707-720 Doi 10.1016/j.cell.2013.03.030

73 Zhang Y, Chen K, Sloan SA, Bennett ML, Scholze AR, O'Keeffe S et al (2014) An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. J Neurosci 34: 11929-11947 Doi 10.1523/JNEUROSCI.1860-14.2014

74 Zhang Y, Sloan SA, Clarke LE, Caneda C, Plaza CA, Blumenthal PD et al (2016) Purification and Characterization of Progenitor and Mature Human Astrocytes Reveals Transcriptional and Functional Differences with Mouse. Neuron 89: 37-53 Doi 10.1016/j.neuron.2015.11.013

75 Zhong Y, Wan YW, Pang K, Chow LM, Liu Z (2013) Digital sorting of complex tissues for cell type-specific gene expression profiles. BMC Bioinformatics 14: 89 Doi 10.1186/1471-2105-14-89

76 Zhou P, Zhang Y, Ma Q, Gu F, Day DS, He A et al (2013) Interrogating translational efficiency and lineage-specific transcriptomes using ribosome affinity purification. Proc Natl Acad Sci U S A 110: 15395-15400 Doi 10.1073/pnas.1304124110

Table 1. Demographics and disease status of cohorts from four brain bank resources.

	Mayo[a]	MSBB[b]	DIAN	Knight-ADRC
Sample Size	191	300	19	103
Age	83 ± 7.77	83.3 ± 7.55	50.6 ± 7.06	85.1 ± 9.78
% Male	45.5	36	68.4	38.8
% *APOE* ε4+	33.2	31.7	14.3	45.6
Brain weight	-	-	1187.7 ± 184.5	1138.1 ± 142.5
AD[c]	82	135	19	87
PA[d]	29	0	0	0
Control	80	85	0	16
CDR[e] = 0	-	40	0	13
CDR = 0.5	-	40	0	9
CDR = 1	-	30	2	11
CDR = 2	-	44	4	14
CDR = 3	-	146	1	56

[a] Mayo stands for Mayo Clinic.

[b] MSBB stands for Mount Sinai Brain Bank.

[c] AD stands for Alzheimer's Disease.

[d] PA stands for pathological aging (amyloid plaques but no tau tangles).

[e] CDR stands for clinical dementia rating for available samples.

Table 2. Comparison of the cellular population structure (AD vs. neuropath-free controls) from the brains in the Mayo Clinic and Mount Sinai Brain Bank.

	Brain Regions	Sample Size	Neuron		Astrocyte		Oligodendrocyte		Microglia	
Mayo	**AD vs Control**	N	Effect	P-value	Effect	P-value	Effect	P-value	Effect	P-value
	Cerebellum	119	-0.03	2.74×10^{-01}	0.05	8.65×10^{-02}	-0.02	1.07×10^{-01}	-3.19×10^{-04}	9.19×10^{-01}
	Temporal Cortex	119	-0.17	1.58×10^{-07}	0.23	5.01×10^{-09}	-0.07	1.8×10^{-02}	-2.03×10^{-03}	5.48×10^{-01}
Mount Sinai Brain Bank	**AD vs Control**									
	Anterior Prefrontal Cortex	184	-0.04	8.14×10^{-04}	0.06	8.11×10^{-05}	-0.01	3.36×10^{-02}	-3.18×10^{-03}	1.12×10^{-02}
	Superior Temporal Gyrus	167	-0.08	3.49×10^{-07}	0.1	1.45×10^{-07}	-0.01	5.8×10^{-02}	-3.17×10^{-03}	5.78×10^{-02}
	Parahippocampal Gyrus	160	-0.11	1.35×10^{-08}	0.13	5.48×10^{-10}	-0.02	1.79×10^{-03}	-3.18×10^{-03}	1.35×10^{-01}
	Inferior Frontal Gyrus	159	-0.04	3.12×10^{-03}	0.06	3.58×10^{-04}	-0.01	4.39×10^{-02}	-3.98×10^{-03}	1.64×10^{-02}
	Clinical Dementia Rating									
	Anterior Prefrontal Cortex	184	-0.02	9.38×10^{-04}	0.02	2.07×10^{-04}	-3.43×10^{-03}	1.25×10^{-01}	-1.46×10^{-03}	4.95×10^{-03}
	Superior Temporal Gyrus	167	-0.03	1.87×10^{-06}	0.04	3.33×10^{-07}	-0.01	2.1×10^{-02}	-1.02×10^{-03}	1.49×10^{-01}
	Parahippocampal Gyrus	160	-0.04	8.56×10^{-06}	0.04	2.85×10^{-06}	-0.01	8.7×10^{-02}	-1.94×10^{-03}	2.53×10^{-02}
	Inferior Frontal Gyrus	159	-0.02	8.29×10^{-05}	0.03	1.4×10^{-05}	-4.64×10^{-03}	6.7×10^{-02}	-1.46×10^{-03}	3.11×10^{-02}
	Braak Staging									
	Anterior Prefrontal Cortex	173	-0.01	1.21×10^{-02}	0.01	1.27×10^{-03}	-3.09×10^{-03}	2.77×10^{-02}	-7.04×10^{-04}	3.12×10^{-02}
	Superior Temporal Gyrus	158	-0.02	2.22×10^{-07}	0.02	2.77×10^{-07}	-2.91×10^{-03}	1.17×10^{-01}	-5.47×10^{-04}	1.97×10^{-01}
	Parahippocampal Gyrus	147	-0.02	1.83×10^{-06}	0.03	9.6×10^{-08}	-0.01	1.49×10^{-03}	-3.71×10^{-04}	4.97×10^{-01}
	Inferior Frontal Gyrus	152	-0.01	1.01×10^{-02}	0.01	8.56×10^{-04}	-3.55×10^{-03}	2.37×10^{-02}	-1.01×10^{-03}	1.74×10^{-02}
	Mean Amyloid Plaques									
	Anterior Prefrontal Cortex	184	-1.88×10^{-03}	3.6×10^{-03}	2.82×10^{-03}	1.03×10^{-04}	-7.99×10^{-04}	2.13×10^{-03}	-1.46×10^{-04}	1.72×10^{-02}
	Superior Temporal Gyrus	167	-4.2×10^{-03}	7.73×10^{-08}	0.01	4.63×10^{-08}	-6.08×10^{-04}	9.01×10^{-02}	-2.04×10^{-04}	1.5×10^{-02}
	Parahippocampal Gyrus	160	-4.96×10^{-03}	5.05×10^{-09}	0.01	1.26×10^{-10}	-9.99×10^{-04}	1.85×10^{-03}	-2.1×10^{-04}	2.58×10^{-02}
	Inferior Frontal Gyrus	159	-2.58×10^{-03}	3.82×10^{-04}	3.53×10^{-03}	1.96×10^{-05}	-7.41×10^{-04}	1.51×10^{-02}	-2.04×10^{-04}	1.26×10^{-02}

The cell-type proportions from AD cases and control were inferred from bulk RNA-seq using the ssNMF method. Effects of AD and associations with additional clinical and pathological phenotypes in cell-type distributions were estimated using linear regression model.

Table 3. Cellular population structure altered in the parietal lobe from AD brains in the DIAN study and Knight-ADRC brain bank.

Disease Status	**Sample Size**	**Neuron**		**Astrocyte**		**Oligodendrocyte**		**Microglia**	
AD Status	N	Effect	P-value	Effect	P-value	Effect	P-value	Effect	P-value
AD[a] vs Control	122	-0.11	5.52×10^{-04}	0.14	2.48×10^{-05}	-0.03	6.5×10^{-02}	-2.64×10^{-03}	2.49×10^{-01}
ADAD vs Control	38	-0.19	3.94×10^{-07}	0.24	1.57×10^{-10}	-0.04	8.5×10^{-03}	-0.01	7.77×10^{-05}
LOAD vs Control	100	-0.09	5.67×10^{-03}	0.12	3.34×10^{-04}	-0.02	1.06×10^{-01}	-1.70×10^{-03}	4.57×10^{-01}
ADAD vs LOAD									
Braak matched	42	-0.08	1.03×10^{-02}	0.11	9.26×10^{-04}	-0.03	7.1×10^{-02}	-1.46×10^{-03}	7.01×10^{-01}
Braak corrected	91	-0.09	4.71×10^{-03}	0.11	5.24×10^{-04}	-0.02	1.77×10^{-01}	-2.41×10^{-03}	4.25×10^{-01}
CDR corrected	94	-0.12	2.11×10^{-03}	0.13	6.29×10^{-04}	-0.02	3.8×10^{-01}	-3.11×10^{-03}	2.41×10^{-01}
Clinical Dementia Rating									
AD[a] and Controls	110	-0.02	2.66×10^{-02}	0.03	5.48×10^{-03}	-0.01	2×10^{-01}	-4.63×10^{-04}	4.77×10^{-01}
ADAD and Controls	26	-0.08	4.12×10^{-04}	0.11	1.78×10^{-07}	0.01	4.03×10^{-03}	-1.55×10^{-03}	1.75×10^{-08}
LOAD and Controls	100	-0.02	3.22×10^{-02}	0.03	7.01×10^{-03}	-0.01	1.81×10^{-01}	-4.64×10^{-04}	5.11×10^{-01}
Braak Staging									
AD[a] and Controls	106	-0.03	8.51×10^{-06}	0.03	3.83×10^{-06}	-4.24×10^{-03}	2.04×10^{-01}	-2.52×10^{-04}	6.81×10^{-01}
ADAD and Controls	33	-0.05	2.37×10^{-05}	0.06	2.45×10^{-05}	-0.01	2.29×10^{-01}	-7.2×10^{-04}	4.89×10^{-01}
LOAD and Controls	88	-0.03	7.41×10^{-04}	0.03	4.63×10^{-04}	-3.72×10^{-03}	3.29×10^{-01}	-1.66×10^{-04}	7.86×10^{-01}

[a] AD includes both autosomal dominant AD (ADAD) and late-onset AD (LOAD).

The cellular population structure was inferred using the ssNMF method. Effects and p-values for the association with disease status, clinical dementia rating and Braak staging using generalized mixed models. We identified similar trends with approximately the same significance levels.

Table 4. Gene specific cellular proportion analysis for Knight-ADRC and Mount Sinai Brain Bank studies

Variant Carriers	**Sample Size**	**Neuron**		**Astrocyte**		**Oligodendrocyte**		**Microglia**	
Knight-ADRC	N	Effect	P-value	Effect	P-value	Effect	P-value	Effect	P-value
PLD3 vs Control	49	-0.1	1.6×10^{-04}	0.13	2.84×10^{-03}	-0.03	6.17×10^{-02}	7.05×10^{-04}	7.89×10^{-01}
TREM2 vs Control	36	-0.07	7.93×10^{-02}	0.11	1.05×10^{-02}	-0.03	4.9×10^{-02}	1.65×10^{-03}	5.84×10^{-01}
Sporadic AD vs Control	45	-0.11	5.45×10^{-03}	0.13	2.95×10^{-04}	-0.02	4.55×10^{-01}	-3.48×10^{-03}	1.13×10^{-01}
*APOE*ε4+ vs *APOE*ε4- LOAD cases and controls	100	-0.06	2.11×10^{-02}	0.05	5.35×10^{-02}	0.01	3.72×10^{-01}	-8.09×10^{-04}	6.31×10^{-01}
*APOE*ε4+ vs *APOE*ε4- LOAD cases only	84	-0.06	2.69×10^{-02}	0.03	2×10^{-01}	0.03	1.4×10^{-02}	-8.31×10^{-04}	6.21×10^{-01}
Mount Sinai Brain Bank - Multi-region									
AD *TREM2* carriers vs Control	301	-0.03	3.57×10^{-01}	0.03	3.19×10^{-01}	-2.08×10^{-03}	7.87×10^{-01}	-2.68×10^{-03}	8.67×10^{-02}
AD non-carriers *TREM2* vs Control	882	-0.07	1.91×10^{-08}	0.08	1.25×10^{-08}	-3.36×10^{-03}	4.79×10^{-01}	-2.89×10^{-04}	7.97×10^{-01}
AD *TREM2* vs AD non-*TREM2*	673	0.05	1.98×10^{-02}	-0.05	1.58×10^{-02}	2.12×10^{-03}	7.76×10^{-01}	-2.13×10^{-03}	1.74×10^{-01}
CDR corrected	673	0.04	5.83×10^{-02}	-0.04	4.46×10^{-02}	1.68×10^{-03}	8.19×10^{-01}	-1.92×10^{-03}	2.22×10^{-01}
Braak corrected	642	0.05	1.3×10^{-02}	-0.05	2.7×10^{-02}	-1.82×10^{-03}	8.13×10^{-01}	-2.66×10^{-03}	1.28×10^{-01}
Mean plaque counts corrected	673	0.05	2×10^{-02}	-0.05	1.59×10^{-02}	1.73×10^{-03}	8.15×10^{-01}	-2.2×10^{-03}	1.5×10^{-01}
*APOE*ε4 counts all samples	556	-0.03	2.75×10^{-04}	0.04	8.06×10^{-06}	-4.33×10^{-04}	4.31×10^{-01}	-0.01	2.42×10^{-03}
*APOE*ε4 counts AD cases	225	-0.03	1.60×10^{-02}	0.03	1.03×10^{-02}	-2.07×10^{-04}	8.20×10^{-01}	-3.46×10^{-03}	3.29×10^{-01}

Figures Legends

Fig 1 Study Design Development of the brain cell-type transcriptomic reference panel (**left column**): the expression signatures of key cell types of the brain were curated by compiling publicly available RNA-seq data from neurons, astrocytes, oligodendrocytes and microglia. The panel was curated iteratively to retain only those samples that showed the most faithful expression signature, while evaluating alternative digital deconvolution methods. The accuracy of digital deconvolution to estimate brain cellular proportion was validated using additional cell-type specific samples, and also by generating chimeric libraries. To study cellular population structure in AD (**right column**), we accessed publicly available datasets from the Advanced Medicines Partnership-AD knowledge portal (AMP-AD), including Mayo Clinic and Mount Sinai Brain Bank datasets. In addition, we generated RNA-seq from participants of the Knight-ADRC and the Dominantly Inherited Alzheimer (DIAN) studies. These three studies generated RNA-seq data from pathological aging brains, Alzheimer Disease cases, and neuropath-free controls for a total of six cerebral cortex regions and cerebellum. We quantified the gene expression for all of the samples included in these studies using the same RNA-seq processing pipeline. Using digital deconvolution methods, we estimated the brain cellular proportions of the samples and compared the proportion between AD cases and controls. We study the cell structure of brains carriers of Mendelian pathological mutations and variants that confer high-risk to AD. Anterior prefrontal cortex – APC; superior temporal gyrus – STG; parahippocampal gyrus – PHG; inferior frontal gyrus – IFG; Mount Sinai Brain Bank – MSBB; Alzheimer Disease – AD; pathological aging – PA.

Fig 2 Cell-type distributions of the samples included in the Mayo Clinic and Mount Sinai Brain Bank Mean neuronal (blue) and astrocytic proportion (red) for **a)** Alzheimer disease affected brains (AD) and controls (bars indicate standard deviations). The numbers of subjects for each group are shown below the x-axis. Distribution for additional clinical and pathological phenotypes reported for the Mount Sinai Brain Bank (MSBB): **b)** clinical dementia rating scores (CDR) and **c)** Braak and Braak staging. **d)** Brain cell-type proportions (x-axis) plotted against the mean number of amyloid plaque (values greater than 0; y-axis). Standard errors were depicted in shaded area with LOESS smooth curve fitted to cell-type proportions derived from deconvolution. (** $P< 0.01$; *** $P< 1.0\times10^{-3}$; and **** $P< 1.0\times10^{-4}$).

Fig 3 Neuron and astrocyte distributions from the DIAN and Knight-ADRC brains a) Mean neuronal (blue) and astrocytic (red) proportions for carriers of pathogenic mutations in APP, PSEN1 or PSEN2 (ADAD), late-onset AD (LOAD) and neuropath-free controls (bars indicate standard deviations). Neuronal and astrocytic proportions plotted against **b)** Braak Staging; **c)** by Clinical Dementia Rating. **d)** Cell-type distributions for carriers of AD genetic risk factors. Lines indicate significance levels (*$P< 0.05$; ** $P< 0.01$; *** $P< 1.0\times10^{-3}$; **** $P< 1.0\times10^{-4}$).

Fig 4 Effect of the APOE ε4 allele and TREM2 coding variants on the cellular population structure Mean neuronal (blue) and astrocytic (red) proportions for **a)** AD cases and controls in the Knight-ADRC brains categorized by *APOE* ε4 carriers vs. non-carriers and **b)** AD cases of Knight-ADRC brain bank (bars indicate standard deviations). **c)** AD cases and controls in the Mayo Clinic and MSBB **d)** AD cases in the Mayo Clinic and MSBB. **e)** Neuronal (blue) and astrocyte (red) distributions for samples included in the Mount Sinai brain bank stratified by *TREM2* genetic status. APC: Anterior Prefrontal Cortex; STG: Superior Temporal Gyrus; PHG: Parahippocampal Gyrus; IFG: Inferior Frontal Gyrus; (n.s. $P > 0.05$; * $P< 0.05$; **** $P< 1.0\times10^{-4}$).

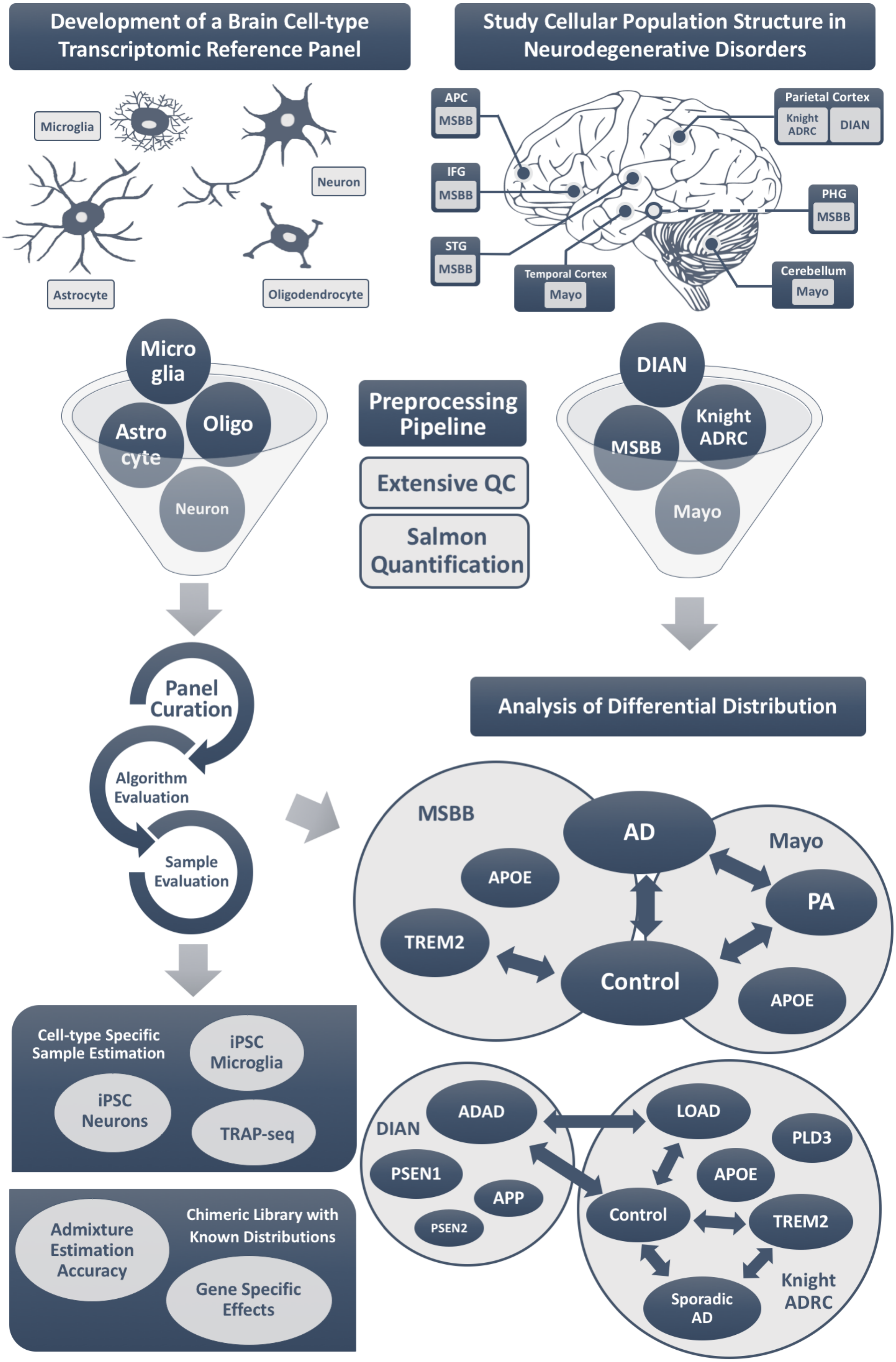
Development of a Brain Cell-type Transcriptomic Reference Panel
Microglia
Neuron
Astrocyte
Oligodendrocyte
Micro glia
Astro cyte
Oligo
Neuron
Panel Curation
Algorithm Evaluation
Sample Evaluation
Cell-type Specific Sample Estimation
iPSC Microglia
iPSC Neurons
TRAP-seq
Admixture Estimation Accuracy
Chimeric Library with Known Distributions
Gene Specific Effects
Preprocessing Pipeline
Extensive QC
Salmon Quantification
Study Cellular Population Structure in Neurodegenerative Disorders
APC
MSBB
IFG
MSBB
STG
MSBB
Temporal Cortex
Mayo
Parietal Cortex
Knight ADRC
DIAN
PHG
MSBB
Cerebellum
Mayo
DIAN
MSBB
Knight ADRC
Mayo
Analysis of Differential Distribution
MSBB
AD
Mayo
APOE
PA
TREM2
Control
APOE
DIAN
ADAD
LOAD
PLD3
PSEN1
APP
APOE
PSEN2
Control
TREM2
Sporadic AD
Knight ADRC

Figure 1

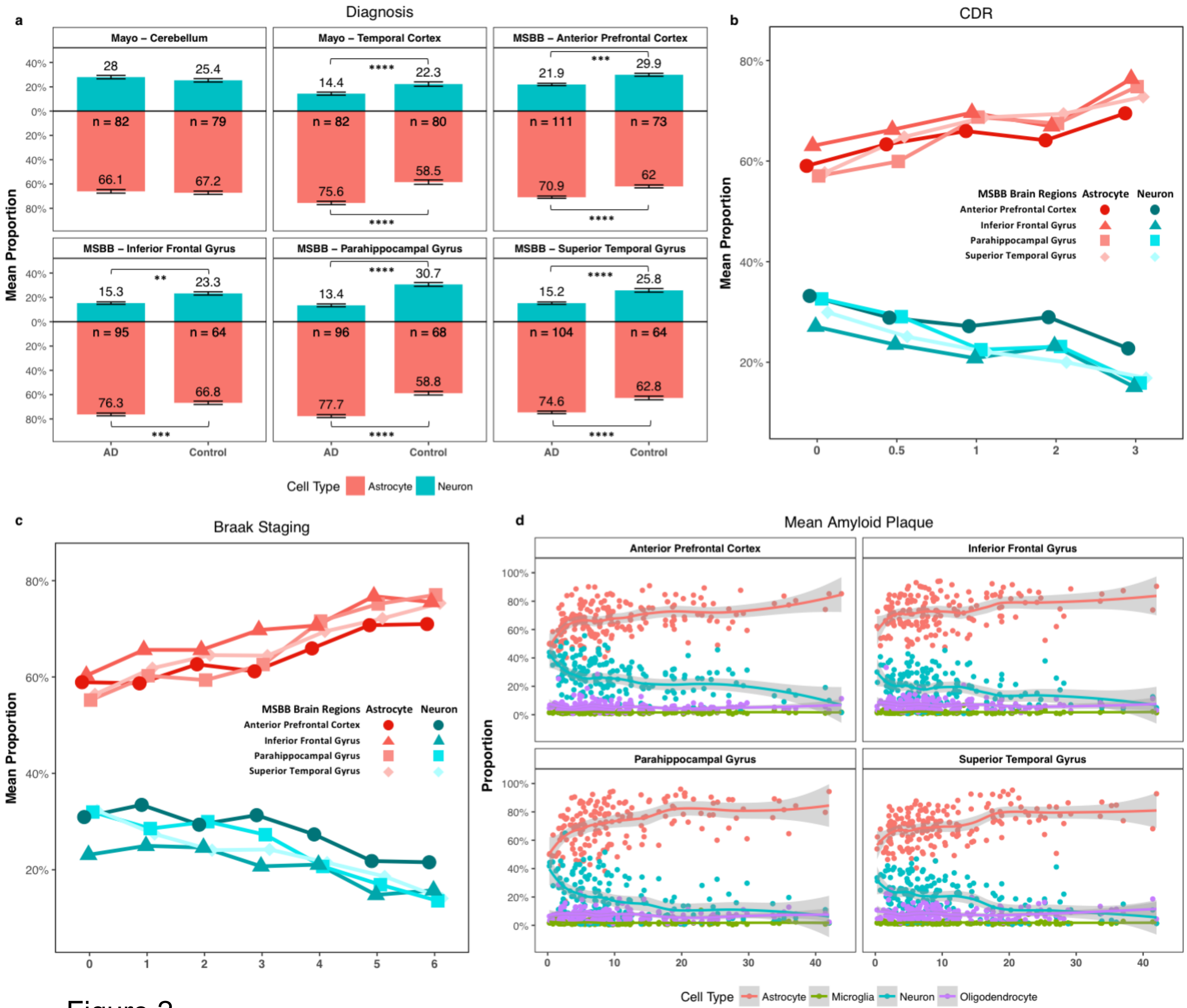
a
Diagnosis
Mean Proportion
Mayo – Cerebellum
28
25.4
n = 82
n = 79
66.1
67.2
Mayo – Temporal Cortex
****
14.4
22.3
n = 82
n = 80
75.6
58.5
****
MSBB – Anterior Prefrontal Cortex
***
21.9
29.9
n = 111
n = 73
70.9
62
****
MSBB – Inferior Frontal Gyrus
**
15.3
23.3
n = 95
n = 64
76.3
66.8
***
MSBB – Parahippocampal Gyrus
****
13.4
30.7
n = 96
n = 68
77.7
58.8
****
MSBB – Superior Temporal Gyrus
****
15.2
25.8
n = 104
n = 64
74.6
62.8
****
40%
20%
0%
20%
40%
60%
80%
AD
Control
Cell Type
Astrocyte
Neuron
b
CDR
Mean Proportion
80%
60%
40%
20%
0
0.5
1
2
3
MSBB Brain Regions
Astrocyte
Neuron
Anterior Prefrontal Cortex
Inferior Frontal Gyrus
Parahippocampal Gyrus
Superior Temporal Gyrus
c
Braak Staging
Mean Proportion
80%
60%
40%
20%
0
1
2
3
4
5
6
MSBB Brain Regions
Astrocyte
Neuron
Anterior Prefrontal Cortex
Inferior Frontal Gyrus
Parahippocampal Gyrus
Superior Temporal Gyrus
d
Mean Amyloid Plaque
Proportion
Anterior Prefrontal Cortex
Inferior Frontal Gyrus
Parahippocampal Gyrus
Superior Temporal Gyrus
100%
80%
60%
40%
20%
0%
0
10
20
30
40
Cell Type
Astrocyte
Microglia
Neuron
Oligodendrocyte

Figure 2

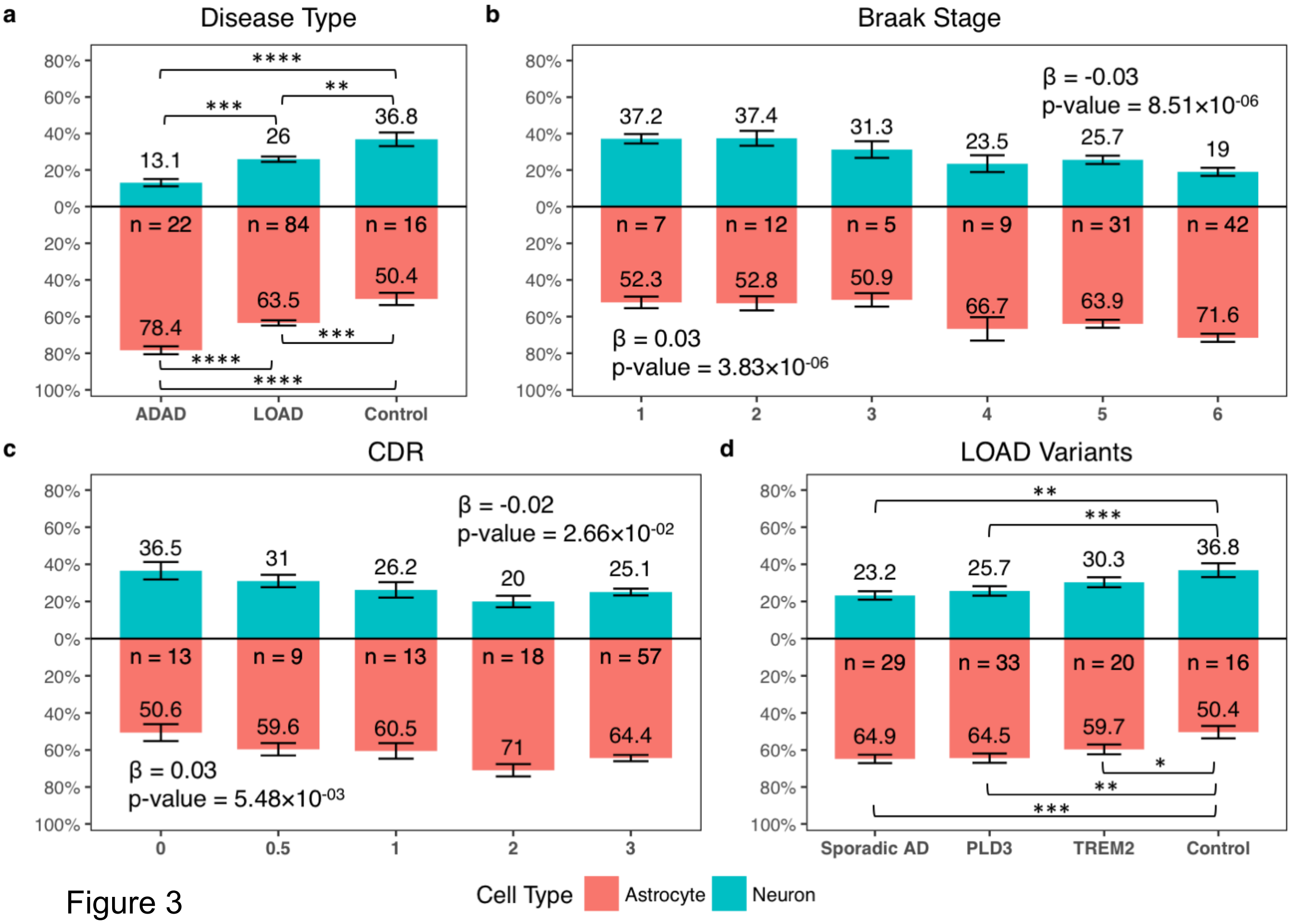
a
Disease Type
80%
60%
40%
20%
0%
20%
40%
60%
80%
100%
****
**
***
13.1
26
36.8
n = 22
n = 84
n = 16
78.4
63.5
50.4
***
****
****
ADAD
LOAD
Control
b
Braak Stage
β = -0.03
p-value = 8.51×10-06
37.2
37.4
31.3
23.5
25.7
19
n = 7
n = 12
n = 5
n = 9
n = 31
n = 42
52.3
52.8
50.9
66.7
63.9
71.6
β = 0.03
p-value = 3.83×10-06
1
2
3
4
5
6
c
CDR
β = -0.02
p-value = 2.66×10-02
36.5
31
26.2
20
25.1
n = 13
n = 9
n = 13
n = 18
n = 57
50.6
59.6
60.5
71
64.4
β = 0.03
p-value = 5.48×10-03
0
0.5
1
2
3
d
LOAD Variants
**
***
23.2
25.7
30.3
36.8
n = 29
n = 33
n = 20
n = 16
64.9
64.5
59.7
50.4
*
**
***
Sporadic AD
PLD3
TREM2
Control
Cell Type
Astrocyte
Neuron

Figure 3

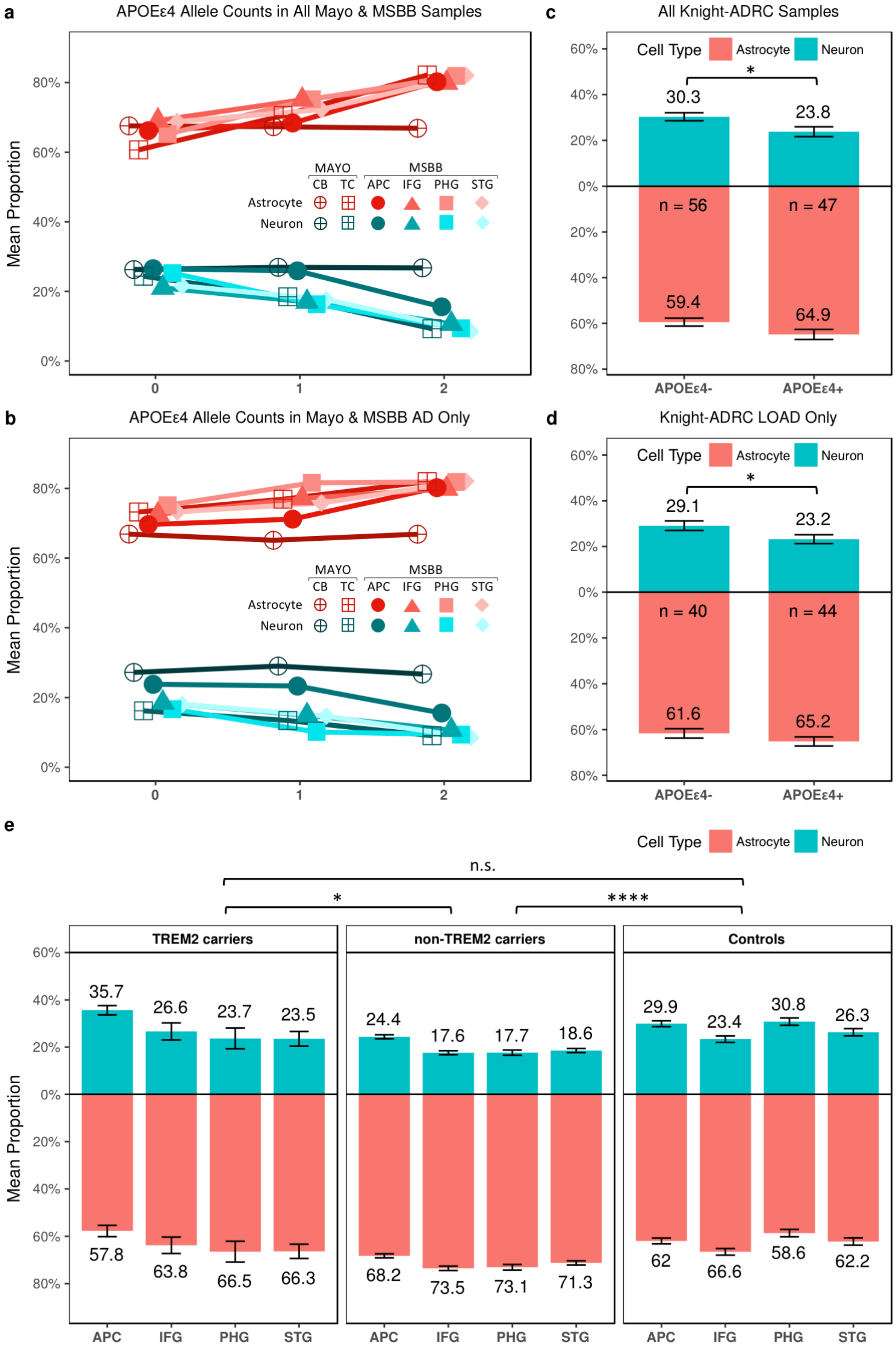
a
APOEε4 Allele Counts in All Mayo & MSBB Samples
Mean Proportion
80%
60%
40%
20%
0%
0
1
2
MAYO
MSBB
CB
TC
APC
IFG
PHG
STG
Astrocyte
Neuron
b
APOEε4 Allele Counts in Mayo & MSBB AD Only
c
All Knight-ADRC Samples
Cell Type
Astrocyte
Neuron
*
30.3
23.8
n = 56
n = 47
59.4
64.9
APOEε4-
APOEε4+
d
Knight-ADRC LOAD Only
29.1
23.2
n = 40
n = 44
61.6
65.2
e
n.s.
*
****
TREM2 carriers
non-TREM2 carriers
Controls
35.7
26.6
23.7
23.5
57.8
63.8
66.5
66.3
24.4
17.6
17.7
18.6
68.2
73.5
73.1
71.3
29.9
23.4
30.8
26.3
62
66.6
58.6
62.2
APC
IFG
PHG
STG

Figure 4