Plasticity of Human Microglia and Brain Perivascular Macrophages in Aging and Alzheimer's Disease

Authors

Donghoon Lee[†1,2,3,4], Christian Porras[#1,2,3,4], Collin Spencer[#1,2,3,4], Milos Pjanic[#1,2,3,4], Philipp Weiler[#5,6], Roman Kosoy[1,2,3,4], Jaroslav Bendl[1,2,3,4], Prashant N M[1,2,3,4], Xinyi Wang[1,2,3,4], Shiwei Zheng[3,7], Karen Therrien[1,2,3,4], Deepika Mathur[1,2,3,4], Steven P. Kleopoulos[1,2,3,4], Zhiping Shao[1,2,3,4], Stathis Argyriou[1,2,3,4], Marcela Alvia[1,2,3,4], Clara Casey[1,2,3,4], Aram Hong[1,2,3,4], Kristin G. Beaumont[3], Robert Sebra[3], Christopher P. Kellner[8], David A. Bennett[9], Guo-Cheng Yuan[3,7], George Voloudakis[1,2,3,4], Fabian J. Theis[5,6,10], Vahram Haroutunian[2,4,11,12], Gabriel E. Hoffman[1,2,3,4], John F. Fullard[#1,2,3,4], Panos Roussos[#†1,2,3,4,12]

Affiliations

[1] Center for Disease Neurogenomics, Icahn School of Medicine at Mount Sinai, New York, NY, USA.
[2] Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA.
[3] Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA.
[4] Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA.
[5] Institute of Computational Biology, Helmholtz Center Munich, Munich, Germany.
[6] Department of Mathematics, Technical University of Munich, Munich, Germany.
[7] Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.
[8] Department of Neurosurgery, Icahn School of Medicine at Mount Sinai, New York, NY, USA.
[9] Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, USA.
[10] TUM School of Life Sciences Weihenstephan, Technical University of Munich, Munich, Germany.
[11] Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY, USA.
[12] Mental Illness Research, Education and Clinical Centers, James J. Peters VA Medical Center, Bronx, New York.

[#]These authors contributed equally to this work.
[†]Corresponding author: donghoon.lee@mssm.edu; panagiotis.roussos@mssm.edu

Abstract

Microglia and perivascular macrophages, myeloid-origin resident immune cells in the human brain, play crucial roles in Alzheimer's disease (AD)[1–4]. However, the field lacks a unified taxonomy describing their heterogeneity and plasticity[5]. To address this, we applied single-cell profiling to two independent, demographically diverse cohorts. The first comprises 543,012 viable myeloid cells from 137 unique postmortem brain specimens, while the second consists of 289,493 myeloid nuclei from 1,470 donors. Collectively, they cover the human lifespan and varying degrees of AD neuropathology. We identify 13 transcriptionally distinct myeloid subtypes, including the “GPNMB” subtype that proliferates with AD. We distinguish two contrasting homeostatic microglial states in AD and with aging: the first (“FRMD4A”) wanes over time, while the second (“PICALM”) becomes more prevalent. By prioritizing AD-risk genes, including PTPRG, DPYD, and IL15, and placing them into a regulatory hierarchy, we identify common upstream transcriptional regulators, namely MITF and KLF12, that regulate the expression of AD-risk genes in the opposite directions. Through the construction of cell-to-cell interaction networks, we identify candidate ligand-receptor pairs, including APOE:SORL1 and APOE:TREM2, associated with AD progression. We show polygenic risk for AD predisposes and prioritize the GPNMB subtype as a therapeutic target of early intervention. Our findings delineate the relationship between distinct functional states of myeloid cells and their pathophysiological response to aging and AD, providing a significant step toward the mechanistic understanding of the roles of microglia in AD and the identification of novel therapeutics.

Main

Despite the quantifiable neuropathology of β-amyloid plaques (Aβ) and neurofibrillary tangles (NFTs)[6], the exact neurobiological mechanisms underlying Alzheimer's disease (AD) remain elusive. Brain myeloid-origin immune cells, including microglia and perivascular macrophages (PVMs), play crucial roles in the pathogenesis of AD[1–4,7–10], providing neuroprotective benefits by clearing lesions, but also exacerbating the disease through the induction of excessive neuroinflammation[11]. While previous studies utilizing single-nucleus/-cell RNA sequencing (snRNA-seq/scRNA-seq) have made significant progress describing complex functional roles of murine and human microglia in AD[4,12–15], challenges with characterizing the wide spectrum of microglial heterogeneity and identifying more nuanced AD-associated subtypes still remain, largely due to limited sample sizes and differences in the single-cell technologies used. Among the issues that arise is the failure of nuclear fractions in snRNA-seq from frozen tissue to capture key genes related to microglial adaptation and response to pathogenic lesions[16]. Moreover, microglia are highly reactive cells, and describing their adaptive nature using scRNA-seq in cells isolated from fresh tissue is challenging[17]. To overcome those limitations, we present two independent human myeloid cohorts generated at single-cell resolution from the prefrontal cortex (PFC). In the first cohort, we isolated viable *ex-vivo* human myeloid cells from fresh postmortem PFC and deeply profiled both nuclear and cytoplasmic RNA. The second cohort focused on the breadth of the transcriptome, profiling human myeloid nuclei from a large number of demographically diverse frozen cortical tissues. By considering both the depth and the breadth of the human myeloid transcriptome, we establish a reproducible taxonomy and demonstrate the importance of microglia and PVM plasticity throughout the lifespan, across different stages of AD pathological and clinical severity, and genetic liability.

Uniformly processed atlas of human myeloid cells

In total, we profiled 832,505 human myeloid cells from the PFC of 1,607 unique donors. The first dataset, named FreshMG, includes samples from fresh autopsy tissue specimens of 137 unique postmortem donors recruited from two brain banks and contains individuals displaying varying degrees of AD neuropathology as well as controls (**Figure 1A**). FreshMG donors are aged between 26 and 107 years (average 80.7 years), comprising 76 females and 61 males (**Figure S1A-J**). To enrich for myeloid cells, viable CD45+ cells were isolated via fluorescence-activated cell sorting (FACS). In addition, for a subset (n=3 donors, each with 8 technical replicates), we profiled surface-level protein markers using CITE-seq[18], using a panel of 163 unique antibodies, resulting in a total of 161 scRNA-seq libraries from fresh brain specimens. Following rigorous QC and initial clustering, we found a large, relatively homogeneous, cluster of myeloid cells along with small subsets of co-purified immune cells, such as monocytes, neutrophils, T, NK, and B cells. The myeloid cluster consisted of 543,012 microglia and PVMs robustly expressing 23,740 genes (**Figure S2A**).

The second dataset, named PsychAD, consists of frozen prefrontal cortex specimens and includes cases and controls from a cohort of 1,470 unique donors (**Figure 1B**). PsychAD donors were aged between 0 and 108 years, (average 71.3 years), comprising 761 females and 709 males (**Figure S1M-V**). Frozen samples were subject to snRNA-seq profiling from which microglia and PVMs were sorted

in silico after basic clustering. After rigorous QC, we identified 289,493 microglia and PVM nuclei robustly expressing 34,890 genes (**Figure S2B**). Next, we aligned and harmonized the scale of clinical variables to facilitate annotation of both datasets (**Methods**) and saw a strong positive correlation with measures of the severity of AD neuropathology, namely diagnostic certainty of AD, Consortium to Establish a Registry for Alzheimer's Disease (CERAD)[19], and Braak stage[20] (**Figure S1K, W**). In contrast, the clinical measures of dementia severity were less well correlated with AD.

Establishing a reproducible taxonomy of human myeloid cells

Our primary objective was to establish a comprehensive cellular taxonomy that is robust and reproducible; however, cross-validating these independent and large-scale single-cell datasets, each with a distinct transcriptomic origin (whole cell vs. nuclei), posed technical challenges. To overcome these, we devised an iterative cross-validation strategy, which involved establishing a reference state and validating it independently until both datasets were in agreement (**Methods**). Utilizing the FreshMG dataset, which provides comprehensive transcriptomic profiles from both nuclear and cytosolic fractions, we identified functionally distinct phenotypes of microglia and PVMs. Subsequently, we cross-validated the presence of these reference subtypes in the frozen specimen snRNA-seq PsychAD dataset. Our iterative process converged on 13 functionally distinct subtypes of human myeloid cells (**Figure 1C, S2C**), and comparison between FreshMG and PsychAD revealed a high degree of consistency between the two cohorts, as evidenced by an average Pearson correlation of 0.77 across all identified subtypes (**Figure 1D**). This rigorous methodology ensured the accuracy and reliability of our cellular taxonomy, laying a solid foundation for further analyses.

We then grouped the cells using two levels of taxonomic hierarchy; the 13 distinct subtypes under six broad functional subclasses of human myeloid cells: Homeostatic (green), Adaptive (blue), Proliferative (yellow), AD-Associated or ADAM (red), *ex-vivo* Activated Microglia or exAM (pink), and PVM (orange) (**Figure 1C**). Each subtype is associated with specific markers that not only aid in their identification but also hint at their functional significance (**Figure 1E, S3A**). Within the homeostatic microglia subclass, we highlight two subtypes, FRMD4A and PICALM, both of which are associated with the regulation of GTPase activity. Homeostatic microglia make up the largest proportion of myeloid cells (**Figure 1F**) and express microglia-specific canonical markers such as P2RY12 and CX3CR1. The FRMD4A subtype uniquely expresses CECR2 and NAV2, with other genes pointing towards cellular maintenance, phagocytosis, cell migration, and adhesion. The PICALM subtype shows elevated expression of PICALM and ELMO1, suggesting roles in the regulation of the immune response. We also identified 7 specialized microglial subtypes, each exhibiting unique adaptive responses to neuro-environmental cues. In general, the gene signatures across these adaptive microglia underscored an enhancement in antigen processing and presentation programs and the facilitation of MHC protein complex assembly. The CCL3 subset is characterized by the upregulation of chemotactic genes, most notably the inflammatory cytokines CCL3, CCL4, and interleukin 1 beta (IL1B). In addition, the IFI44L subtype is enriched in interferon-inducible genes, like IFIT1, IFIT2, and IFIT3, suggesting a role in the antiviral innate immune response. The AIF1, HIF1A, and HIST clusters share a common gene program related to immunoglobulin-mediated immune response, while the TMEM163 cluster focuses on antigen processing and presentation via MHC II. The final adaptive cluster, HSPA1A, is enriched for gene signatures responsible for adaptive response to unfolded

protein, which is characterized by elevated activity of heat shock proteins and cellular stress response, with a potential role in AD neuropathology[21]. In addition, we identified a subtype, the GPNMB, which is predominantly observed in individuals with AD[22,23]. These AD-associated microglia (ADAM) feature elevated expression of GPNMB, MITF, and PTPRG genes, and functional enrichment analysis suggests increased phagocytic activity is a hallmark of these cells. Consistent with previous studies[12], we also identified a cluster of proliferative cells, MKI67, that is highly enriched in cell-cycle dependent genes (STMN1, MKI67, TOP2A). Lastly, we report a cluster, ERN1, showing specific expression of ERN1 and PLK2 genes that resemble activation patterns of exAM[17] (**Supplementary Information**). In addition to microglial subtypes, we identified a PVM cluster, named CD163, expressing a unique set of known PVM-specific markers, notably CD163 and F13A1. The CD163 cluster displayed a significant enrichment of genes involved in endocytic processes, emphasizing its priming for receptor-mediated endocytosis and phagocytosis.

While we observe a close similarity between ADAM and PVM clusters (**Figure S2C**), we found a clear separation between the two when we enriched for conserved murine disease-associated microglia (DAM) signatures as well as human DAM signature from iPSC-derived microglia[23,24] (**Figure S2F**).

Validation of the myeloid taxonomy using CITE-seq, human brain biopsies, and spatial phenotyping analysis

Since the cellular taxonomy is primarily derived from the transcriptome, we assessed the preservation of the functional hierarchical structure at the protein level using CITE-seq, jointly quantifying the transcriptome and 163 unique cell-surface proteins. Using this approach, we identified unique patterns of protein markers for each myeloid subtype (**Figure S2D**). For example, within the homeostatic microglia subclass, the FRMD4A subtype expressed CD99 and ITGB3, while the PICALM subtype expressed CLEC4C and TNFRSF13C. Likewise, the PVM cluster showed distinct surface markers, CD163 and CCR4, while the ADAM cluster was specific for CD9 and CD44. To validate the spatial context of PVMs, we conducted deep single-cell phenotyping and spatial analysis using multiplexed imaging assay (Akoya PhenoCycler) and demonstrated, via staining for CD163, that PVMs colocalize around blood vessels (**Figure 1G**).

To further validate the robustness of our human myeloid taxonomy, in addition to the FreshMG and PsychAD cohorts, we generated an additional, independent, scRNA-seq dataset, called LivinMG. For this dataset, brain biopsies were obtained from 25 unique human donors (26 libraries; 97,828 cells after QC) diagnosed with spontaneous intracerebral hemorrhage (ICH)[25]. The brain tissue was collected during treatment and processed in an identical manner to the fresh autopsy material. It's important to note that cortical biopsy samples were obtained from a site distal to the site of the hemorrhage and, in the absence of a secondary diagnosis, are considered neurotypical controls. To assign microglial subtype labels, we employed scANVI[26] to transfer the established myeloid taxonomy from the FreshMG dataset (**Figure S2E**). We confirmed that all 13 subtypes were present in all three datasets, thus validating that the taxonomy is consistent, irrespective of the tissue source (**Figure 1F**).

Variation in human myeloid subtypes during aging

After determining 13 distinct subtypes of human brain myeloid cells, we examined the compositional variation of myeloid subtypes that are associated with aging in a subset of neurotypical donors who were free of dementia and diagnostic neuropathology from the FreshMG and PsychAD datasets. We normalized the subtype count ratio data using the centered log-ratio transformation and modeled using a linear mixed model, accounting for technical and demographic variables (**Methods**). Notably, the two homeostatic microglia subtypes displayed opposing trajectories with respect to aging (**Figure 2A, B**). The FRMD4A subtype showed progressive decline while the PICALM subtype showed a gradual increase with age. In addition, we saw an overall increase in the proportions of the ADAM and PVM subtypes with age (**Figure 2A left, B**). In contrast, we observed an age-related decline in the CCL3 subtype, indicating a possible reduction of chemotactic microglia in older brains. In parallel, we investigated sex-dependent variation in human myeloid subtypes, with or without taking age into consideration, but did not find any statistically significant compositional differences between males and females (**Figure 2A middle, right**).

Next, we investigated differential gene expression patterns associated with normal aging (**Figure 2C**). One notable finding was the up-regulation of the MS4A6A gene, a member of the MS4A family of cell membrane proteins, which are involved in the regulation of calcium signaling and have been implicated in neurodegenerative processes[27]. The age-related gene expression changes for both homeostatic subtypes were enriched with actin filament-based process and actin cytoskeleton organization pathways, supporting their proposed roles in cell adhesion and migration (**Figure S3B**). The CD163 subtype was associated with the upregulation of cell adhesion processes as well as pathways related to cell proliferation. The gene signatures in the GPNMB subtype were enriched with immune response and activation. Overall, the increased involvement of PVM and ADAM subclasses indicated an upregulation of inflammatory responses in older individuals.

Variation in transcriptional landscape of human myeloid cells during onset and progression of AD

Next, we examined the variation of gene expression during the early stages of AD and its progression. To minimize the effect of younger brains, we limited the analysis to donors 40 years and older, resulting in a dataset composed of 134 donors from the FreshMG and 1,314 donors from the PsychAD cohort. We first evaluated the involvement of myeloid subtypes using the centered log-ratio transformed count ratio data after accounting for technical and demographic variables (**Methods**). Overall, irrespective of different measures of AD phenotypes (dx_AD, CERAD, Braak, and Dementia), we observed robust changes in subtype proportions in both FreshMG and PsychAD cohorts (**Figure 3A**). Similar to normal aging, two homeostatic subtypes showed opposing trends, where the FRMD4A subtype showed a progressive decline with increasing AD burden while the PICALM subtype showed a gradual increase (**Figure 3B**). The deviation point where the two subtypes diverged was observed during the early stages of Braak, between 0 and 1, likely before the onset of AD. Likewise, we observed a consistent increase in the proportion of the PVM subtype.

The overall compositional variation of AD phenotypes was comparable to aging except for the GPNMB subtype (**Figure 3C**). The GPNMB subtype was an outlier and showed the largest effect size across all 4 AD phenotypes, suggesting that proliferation of the GPNMB subtype is a hallmark of AD.

Next, we evaluated genes exhibiting differential expression patterns across four different measures of AD neuropathology (**Figure 3D**). Our analysis led us to discover candidate AD risk genes, including PTPRG, DPYD, and IL15, which displayed upregulation across all measures of AD severity. Pathway enrichment analysis revealed the PICALM and GPNMB subtypes share common pathways related to the regulation of cell adhesion (**Figure S3C**). In contrast, both the FRMD4A and CD163 subtypes appear to be associated with negative regulation of cell projection organization.

Given the strong compositional shifts and gene signatures for AD phenotypes, we tested the presence of AD signatures using bulk microglia RNA-seq data (BulkMG; **Methods**). First, we created myeloid subtype signatures from both the FreshMG and PsychAD datasets by aggregating gene expression by subtype. We then compared the resulting subtype signatures to BulkMG gene expression data, stratified by AD case and control status. Interestingly, the Pearson correlation between subtypes and AD diagnosis clearly reflected the compositional shifts we observed across multiple AD phenotypes (**Figure 3E**). The FRMD4A, TMEM163, CCL3, and HSPA1A signatures closely correlated with the BulkMG from controls, while the PICALM, CD163, GPNMB, and HIF1A signatures closely matched those from AD cases. These results independently validate the observed changes in the myeloid transcriptome during the onset and progression of AD.

To understand the dynamic changes that take place during the onset and progression of AD at a molecular level, we expanded our analysis from using discrete donor-level clinical variables, such as CERAD scores or Braak staging, to a continuous pseudotime measure by ordering cells along a disease trajectory. Here, we estimated Braak-stage-informed ancestor-progenitor relations between observations through transport maps between neighboring disease stages using Moscot[28], a toolbox for optimal transport-based analyses of single-cell data. We then quantified cell-cell transition probabilities, computed putative drivers, and constructed the disease-stage-informed pseudotime with CellRank 2[29] (**Figure S5A**; **Methods; Supplementary Information**). Overall, we observed an increase in pseudotime with disease progression (**Figure S5E**). Stratified by subtypes, we observed that PICALM homeostatic microglia were assigned larger pseudotime values, compared to FRMD4A homeostatic cells (p-value < 0.001, **Figure 3F**), indicating their association with disease progression and aligning with the compositional variation of AD phenotypes observed earlier. To identify potentially critical time points for disease progression, we compared changes in pseudotime across disease stages for each myeloid subtype (**Methods**). This analysis revealed that the change was most pronounced starting from Braak stage 3 (**Figure 3G**). Similarly, when the disease pseudotime was stratified by CERAD stages, the largest change occurred between CERAD stages 3 and 4.

Upstream master regulators of human myeloid cells in AD

After identifying potential AD risk genes, we analyzed myeloid gene regulatory networks (GRNs) in an attempt to discover key upstream transcriptional regulators. Using SCENIC[30,31], we constructed GRNs based on expression data and known transcription factor (TF) binding motifs and defined units of regulatory hierarchy (regulons) (**Figure 4A**). Subsequently, we assessed the enrichment of the

regulon for each myeloid subtype independently (**Methods**), revealing high concordance between the FreshMG and PsychAD cohorts (**Figure 4B**). We then derived combined regulon enrichment scores using meta-analysis (**Methods**) and observed strong regulon subtype-specificity (**Figure 4C**). The FRMD4A and PICALM homeostatic subtypes were defined by enrichment of KLF12, GLIS3, and BACH2 regulons, while the PICALM, CD163, and GPNMB subtypes displayed exclusive enrichment of the MITF regulon. To link inferred regulons to differentially expressed AD genes, we performed enrichment tests using 4 different types of AD risk signatures (**Methods**). Notably, the target genes of MITF, KLF12, and GLIS3 TFs were significantly associated with AD risk profiles in the PICALM, FRMD4A, GPNMB, and HIF1A subtypes (**Figure 4D**). MITF was preferentially enriched with upregulated AD signatures, whereas KLF12 and GLIS3 were more preferentially associated with downregulated AD signatures. Visualization of the joint MITF-KLF12-GLIS3 regulon network with AD risk genes revealed coordinated modulation of both up and down-regulated candidate risk genes (**Figure 4E**). These findings collectively suggest the coordinated activity of MITF, KLF12, and GLIS3 in regulating AD risk gene expression in disease-associated microglia states. Functional enrichment analysis revealed MITF, KLF12, and GLIS3 target genes involved in key biological processes for microglia function such as phagocytosis, cytokine production, and cellular response (**Figure 4F**). We find that MITF distinctly regulates phagocytic-related pathways in line with previous findings from *in-vitro* models[23]. In summary, by integrating SCENIC network inference with the analysis of differentially expressed genes in AD, we nominate MITF, KLF12, and GLIS3 as potential upstream master regulators of gene expression changes relevant to AD pathogenesis.

Cellular crosstalk between human myeloid subtypes

To gain mechanistic insights into how different human myeloid subtypes communicate with each other and mediate AD risks, we investigated the change of cell-to-cell interactions (CCIs) at different stages of AD using the LIANA framework[32] (**Figure S6C**). For each individual, we inferred the magnitude, specificity, and directionality of cell-to-cell communication using gene expression profiles and known ligand-receptor interactions. As expected, we observed strong concordance between the magnitude of CCI activities from FreshMG and PsychAD cohorts (**Figure 5A**). The correlation analysis revealed homeostatic, PVM, and ADAM subtypes had better consistency and sufficient statistical power, whereas rare subtypes like MKI67 and CCL3 had lower activity and inconsistency possibly due to limited statistical power. By evaluating the CCI magnitude scores as a function of all 4 AD phenotypes using a linear mixed model, we identified differential CCIs associated with AD (**Figure 5B, S6D**). We found contrasting candidate CCIs (APOE→SORL1 and MRC1→PTPRC) working in opposing directions. The APOE→SORL1 interactions were more frequent and up-regulated in AD and were prioritized as the top AD-relevant CCI, while MRC1→PTPRC interactions were down-regulated in AD. Notably, we also observed the APOE→TREM2 interaction among the top CCIs associated with AD. To test for genetic association, we stratified the CCIs by AD case- and control-associated ligands and receptors and performed the gene-set analyses using GWAS data (**Figure 5C**). Interestingly, we found a strong association of AD-associated receptors with AD risk but not with ligands. Visualizing the CCIs as directional networks in the context of different myeloid subtypes placed the GPNMB subtype as the hub mediating AD risks (**Figure 5D**). Finally, the pathway enrichment analysis indicated

interactions involved with the ADAM_GPNMB subtype were enriched with lipid metabolism and regulation of proteolysis (**Figure 5E**).

Polygenic disease risk scoring of human myeloid subtypes

We estimated the polygenic disease risk at single-cell resolution using GWAS disease-associated genes and scored myeloid subtypes responsible for heritable risks (`scDRS`; **Methods; Figure S7A**). We extended the analysis to a set of the brain related diseases beyond AD including schizophrenia (SCZ), bipolar disorder (BD), major depressive disorder (MDD), autism spectrum disorder (ASD), multiple sclerosis (MS), amyotrophic lateral sclerosis (ALS), and Parkinson's disease (PD). The polygenic risk scores for each trait were highly reproducible between the FreshMG and PsychAD cohorts with AD and MS having the greatest correlation (**Figure S7B**). The meta-analysis of both FreshMG and PsychAD cohorts indicated that the 9 subtypes of myeloid cells were significantly associated with heritable AD risk, which was the largest of all brain diseases followed by MS, SCZ, and MDD (**Figure 6A**). Notably, the GPNMB subtype had the widest coverage showing significant heritable risks for all 8 diseases. When we evaluated the association of polygenic AD risk scores with the compositional variation observed in AD, we observed a positive correlation (Pearson's r=0.55) indicating cells with greater heritable risks become enriched in AD (**Figure 6B**). This also suggested heritable risks might play a role in driving the compositional changes of myeloid subtypes. To ensure that the heritable risk estimates were not driven from the AD cases, we also calculated the AD risk scores after splitting donors by AD cases and controls (case-control concordance, **Methods; Figure S7A**). We showed that the polygenic AD risk scores were concordant between AD cases and controls, and thus showing that the estimates were independent from AD status suggesting the heritable risks could be inherent to myeloid subtypes (**Figure S7C**).

Next, we calculated per-donor AD polygenic risk scores (PRS) and evaluated their association with changes in myeloid subtype composition using a linear mixed model (PRS, **Methods; Figure S7A**). The GPNMB subtype was among the most positively and significantly changed (**Figure S7D**), and we observed a similar compositional variation between AD phenotype (dx_AD) and PRS, suggesting that changes in myeloid subtypes were partially driven by AD genetic risk predisposition (**Figure 6C circle**). We further showed this association was not purely driven by the AD status alone as the same compositional variation was observed using a disease-free subset (**Figure 6C triangle**).

Discussion

The human myeloid cell atlas presented here underscores the importance of their functional plasticity throughout life, reflecting their ability to adapt to their microenvironment progressively and reversibly. Of particular interest is their behavior in the context of normal aging and disease. The contrasting regulatory interplay between two homeostatic microglial subtypes exemplifies the dual role of myeloid cells, providing neuroprotective benefits in the early stages of AD, while exacerbating the phenotype, via neuroinflammation, as the disease progresses. We find that the shift in balance between FRMD4A and PICALM homeostatic subtypes is associated with the onset of AD and suggest that this imbalance might play a role in compromising the efficacy of their respective neuroprotective

roles. Identifying these early changes in gene expression could be crucial for the early diagnosis and treatment of AD.

Our comprehensive analyses uncover striking similarities between normal aging and AD pathology. We speculate that the natural aging process is accelerated in AD, and follows a similar trend for all subtypes, with the exception of ADAM. We show that the ADAM subset is enriched with AD-associated gene signatures and the related transcriptomic changes are unique to AD and go beyond normal age-related changes. Through our GRN analysis, we prioritize MITF as the master regulator of AD risk signatures, governing the expression of numerous AD-associated genes, including APOE, DPYD, TREM2, and PTPRG[13,33,34]. The MITF network is specifically enriched for phagocytic activity markers and has previously been studied as a key regulator of homeostatic microglia functions, in particular for driving autophagic states, and for allowing microglia to move, sense, and clear Aβ/Tau proteinopathies[22,23,35–37]. Moreover, through our CCI analysis, we prioritize ligand-receptor interactions mediating the aggravation of AD progression and place ADAM as key players. The significant enrichment of AD genetic risk loci (APP, TREM2, SORL1, SORT1, ABCA1, TSPAN14) within the prioritized receptors of AD-associated CCIs suggests potential mechanisms behind the genetic susceptibility to AD. Similarly, our heritability analysis further strengthens the hypothesis that there might be a genetic basis for the observed molecular changes. Finally, based on polygenic risk scores, we show that a portion of AD-associated transcriptomic changes are inherent, independent of AD diagnosis.

Taken together, our findings shed light on the intricate molecular changes underlying the plasticity of human myeloid cells that predispose them to AD. As the number of people living with AD is expected to double by 2050, there is an urgent need to address the enormous societal impact and serious challenges that AD poses to our healthcare and economic systems. The data presented here are a valuable resource for identifying novel therapeutic strategies to combat this debilitating neurodegenerative disorder.

Methods

Sources and description of human biosamples

All brain specimens were obtained through informed consent via brain donation programs at the respective organizations. All procedures and research protocols were approved by the respective ethical committees of our collaborator's institutions. The FreshMG samples (n=137) were taken from 96 fresh postmortem autopsy samples obtained at the Mount Sinai/JJ Peters VA Medical Center NIH Brain and Tissue Repository (NBTR) in the Bronx, NY. An additional set of 41 fresh postmortem autopsy samples was obtained from participants in the Religious Orders Study or Rush Memory and Aging Project (ROSMAP) at Rush Alzheimer's Disease Center (RADC) in Chicago, IL. Both studies were approved by an Institutional Review Board of Rush University Medical Center and all participants signed informed and repository consents and an Anatomic Gift Act[38]. The PsychAD cohort comprises 1,470 donors from three brain banks, Mount Sinai NIH Brain Bank and Tissue Repository (MSSM; 1,023 samples), NIMH Human Brain Collection Core (HBCC; 295 samples), and ROSMAP (RUSH; 152 samples). Finally, LivingMG biopsies were collected from patients undergoing procedures for intracerebral hemorrhage evacuation (STUDY-18-01012A), as described previously[10].

Collection and harmonization of clinical, pathological, and demographic metadata

Since the brain tissue specimens were collected from three different sites, the available clinical data varies as a function of source. As such, we used the following scheme to harmonize available clinical, pathological, and demographic metadata: the CERAD scoring scheme for neuritic plaque density[19] was harmonized for consistency across multiple brain banks, where the scores range from 1 to 4, with increasing CERAD number corresponding to an increase in AD burden; 1=no neuritic plaque (normal brain), 2=sparse (possible AD), 3=moderate (probable AD), 4=frequent (definite AD). Samples from ROSMAP used consensus summary diagnosis of no cognitive impairment (NCI), mild cognitive impairment (MCI), and dementia and its principal cause, Alzheimer's dementia[39–41]. MSSM/VA samples used clinical dementia rating (CDR), which was based on a scale of 0-5; 0=no dementia, 0.5=questionable dementia (very mild), 1=mild dementia, 2=moderate dementia, 3=severe dementia, 4=profound dementia, 5=terminal dementia. After consulting with clinicians, we created a harmonized ordinal variable where dementia is categorized into three levels of cognitive decline, independent of AD diagnosis; 0=no cognitive impairment, 0.5=MCI (mild cognitive impairment), and 1=dementia. In addition to AD phenotype, we collected comprehensive demographic (age, sex, and genetic ancestry) and technical variables (brain bank, sequencing facility, sequence pooling information, postmortem interval (PMI; measured in minutes), APOE genotype) to describe each cohort (**Figure S1**, **Table S1**).

Clinical diagnosis of AD

For analysis comparing donors with AD cases and neurotypical controls, a binary clinical diagnosis variable for AD, **dx_AD**, was defined as follows. Individuals with CERAD 2, 3, or 4, Braak ≥ 3, and CDR ≥ 1 for MSSM/VA or Alzheimer's dementia for ROSMAP were classified as AD cases. Controls were defined as individuals with CERAD 1 or 2 and Braak 0, 1, or 2.

Measuring AD neuropathology

For analysis comparing donors with pathologic AD, the following variables were used to measure the severity of AD neuropathology. **CERAD score**[19]. A quantitative measure of Aβ plaque density where 1 is normal, 2 is possible AD, 3 is probable AD, and 4 is definite AD[39]. **Braak AD-staging score** measuring progression of neurofibrillary tangle neuropathology (Braak & Braak-score, or BBScore). A quantitative measure of the regional patterns of neurofibrillary tangle (NFT) density across the brain, where 0 is normal and asymptotic, 1-2 indicate initial stages where NFT begins to appear in the locus coeruleus and the transentorhinal region, 3-4 indicate progression to limbic regions, such as the hippocampus and amygdala, and 5-6 indicate NFT are widespread, affecting multiple cortical regions[42–44].

Measuring cognitive impairment

For analysis comparing donors with AD-related dementia, the following variable was used to measure the severity of cognitive impairment. **Clinical assessment of dementia.** A harmonized variable of cognitive status based on CDR scale for MSSM/VA or NCI, MCI, Alzheimer's dementia for ROSMAP. We used the three-level ordinal categories of clinical dementia to measure the severity of dementia, in which 0 indicates no dementia, 0.5 indicates minor cognitive impairment, and 1.0 indicates definite clinical dementia.

Isolation and fluorescence-activated cell sorting (FACS) of microglia from fresh brain specimens (FreshMG and LivingMG)

Fresh brain tissue specimens were placed in tissue preservation buffer (Miltenyi Biotech) and stored at 4˚C for ≤48hrs before processing using the Adult Brain Dissociation Kit (Miltenyi Biotech), according to manufacturer's instructions. RNase inhibitors (Takara Bio) were used throughout cell preparation. Following de-myelination (Miltenyi de-myelination kit - Miltenyi Biotech) cells were incubated in antibody (CD45: BD Pharmingen, Clone HI30 and CD11b: BD Pharmingen, Clone ICRF44) at 1:500 for 1 hour in the dark at 4˚C with end-over-end rotation. Prior to fluorescence-activated cell sorting (FACS), DAPI (Thermoscientific) was added at 1:1000 to facilitate the selection of viable cells. Viable (DAPI negative) CD45/CD11b positive cells were isolated by FACS using a FACSAria flow cytometer (BD Biosciences). Following FACS, cellular concentration and viability were confirmed using a Countess automated cell counter (Life technologies).

Isolation and fluorescence-activated nuclear sorting (FANS) of nuclei from frozen brain specimens (PsychAD), with hashing

All buffers were supplemented with RNAse inhibitors (Takara). 25mg of frozen postmortem human brain tissue was homogenized in cold lysis buffer (0.32M Sucrose, 5 mM $CaCl_2$, 3 mM Magnesium acetate, 0.1 mM, EDTA, 10 mM Tris-HCl, pH8, 1 mM DTT, 0.1% Triton X-100) and filtered through a 40 µm cell strainer. The flow-through was underlaid with sucrose solution (1.8 M Sucrose, 3 mM Magnesium acetate, 1 mM DTT, 10 mM Tris-HCl, pH8) and centrifuged at 107,000 g for 1 hour at 4˚C. Pellets were resuspended in PBS supplemented with 0.5% bovine serum albumin (BSA). 6 samples were processed in parallel. Up to 2M nuclei from each sample were pelleted at 500 g for 5 minutes at 4˚C. Nuclei were re-suspended in 100 µl staining buffer (2% BSA, 0.02% Tween-20 in PBS) and

incubated with 1 µg of a unique TotalSeq-A nuclear hashing antibody (Biolegend) for 30 min at 4˚C. Prior to FANS, volumes were brought up to 250 µl with PBS and 7aad (Invitrogen) added according to the manufacturer's instructions. 7aad positive nuclei were sorted into tubes pre-coated with 5% BSA using a FACSAria flow cytometer (BD Biosciences).

scRNA-seq and CITE-seq library preparation (FreshMG and LivingMG)

Following FACS, 10,000 cells were processed using 10x Genomics single cell 3' capture reagents (10x Genomics), according to the manufacturer's instructions. In parallel, CITE-seq was performed on a subset of samples (n=3 donors, n=8 replicates per donor) using the TotalSeq™-A Human Universal Cocktail (BioLegend) with 154 unique cell surface antigens, including principal lineage antigens, and includes 9 isotype control antibodies to survey surface antigens. CITE-seq was performed according to the manufacturer's instructions. For the CITE-seq experiment, a total of 80,000 cells were loaded on 10x Genomics B chips (10,000 of each uniquely barcoded sample aliquot per B chip lane), with a total targeted recovery of around 40,000 cells.

snRNA-seq and hashing library preparation (PsychAD)

Following FANS, nuclei were subjected to 2 washes in 200 µl staining buffer, after which they were re-suspended in 15 µl PBS and quantified (Countess II, Life Technologies). Concentrations were normalized and equal amounts of differentially hash-tagged nuclei were pooled. A total of 60,000 (10,000 each) pooled nuclei were processed using 10x Genomics single cell 3' v3.1 reagents (10x Genomics). Each pool was run across x2 10x Genomics lanes to create a technical replicate. At the cDNA amplification step (step 2.2) during library preparation, 1 µl 2 µm HTO cDNA PCR "additive" primer v3.1 was added (PMID: 30567574). After cDNA amplification, supernatant from 0.6x SPRI selection was retained for HTO library generation. cDNA library was prepared according to the 10x Genomics protocol. HTO libraries were prepared as previously described (PMID: 30567574). cDNA and HTO libraries were sequenced at NYGC using the Novaseq platform (Illumina).

Processing of scRNA-seq data (FreshMG)

We developed a tracking platform to record all technical covariates (such as 10x Genomics lot kit number, dates of different preparations, viable cell counts, etc.) and quality metrics derived from data preprocessing. **Alignment**. Paired-end scRNA-seq reads were aligned to the hg38 reference genome and the count matrix was generated using 10x cellranger count (v7.0.0). Subsequently, we used the CellBender[45] to carefully separate out true cells from empty droplets with ambient RNA from raw unfiltered cellranger output. **QC**. We performed the downstream analysis by aggregating gene-count matrices of multiple samples. A battery of QC tests was performed to filter low-quality libraries and non-viable cells within each library using Pegasus (v1.7.0)[46]. Viable cells were retained based on UMI ($1{,}000 \leq n_UMI \leq 40{,}000$), gene counts ($500 \leq n_genes \leq 8{,}000$), and percentage of mitochondrial reads ($percent_mito \leq 20$). We also checked for possible contamination from ambient RNA, a fraction of reads mapped to non-mRNA like rRNA, sRNA, pseudogenes, and known confounding features such as lncRNA MALAT1. Further filtering was carried out by removing doublets using the Scrublet method[47]. After filtering, the retained count matrix was normalized and log-transformed. Batch correction. We assessed the correlation between all pairs of technical and biological variables using

Canonical Correlation Analysis and used the Harmony method[48] to regress out unwanted confounding variables such as the source of brain tissue. Clustering. From the kNN graph calculated from the PCA, we clustered cells in the same cell state using Leiden[49] clustering. We use UMAP[50] for the visualization of resulting clusters. Cells identified as T cells, NK cells, monocytes, neutrophils, oligodendrocytes, and astrocytes were removed, and those identified as microglia and PVMs were carried forward for subsequent taxonomic analysis.

Processing of snRNA-seq data (PsychAD)

Alignment. Samples were multiplexed by combining 6 donors in each nuclei pool using hashing, and each biosample was processed in duplicate to produce technical replicates. Paired-end snRNA-seq libraries were aligned to the hg38 reference genome using STAR solo[51,52] and multiplexed pools were demultiplexed using genotype matching via vireoSNP[53]. After per-library count matrices were generated, the downstream processing was performed using pegasus v1.7.0[46] and scanpy v1.9.1[54]. **QC**. We applied rigorous three-step QC to remove ambient RNA and retain nuclei for subsequent downstream analysis. First, the QC is applied at the individual nucleus level. A battery of QC tests was performed to filter low-quality nuclei within each library. Poor-quality nuclei were detected by thresholding based on UMI (1,179 ≤ n_UMI ≤ 200,000; determined based on median absolute deviation of n_UMI distribution), gene counts (986 ≤ n_genes ≤ 15,000; determined based on median absolute deviation of n_genes distribution), and percentage of mitochondrial reads (percent_mito ≤ 1). We also checked for possible contamination from ambient RNA, the fraction of reads mapped to non-mRNA like rRNA, sRNA, and pseudogenes, as well as known confounding features, such as the lncRNA MALAT1. Second, the QC was applied at the feature level. We removed features that were not robustly expressed in at least 0.05% of nuclei. Lastly, the QC was applied at the donor level. We removed donors with very low nuclei counts, which can introduce more noise to the downstream analysis. We also removed donors with low genotype concordances. Further filtering was carried out by removing doublets using the Scrublet method[47]. **Batch correction**. We assessed the correlation between all pairs of technical variables using Canonical Correlation Analysis and used the Harmony method[48] to regress out unwanted variables such as the effect of brain tissue sources. **Clustering**. Highly variable features were selected from mean and variance trends, and we used the k-nearest-neighbor (kNN) graph calculated on the basis of harmony-corrected PCA embedding space to cluster nuclei in the same cell type using Leiden[49] clustering algorithms. We used UMAP[50] for the visualization of the resulting clusters. **Isolation of myeloid cells**. Identified cell-type clusters were annotated based on manual curation of known gene marker signatures obtained from Human Cell Atlas and human DLPFC study[55]. Classes of immune cells, including Microglia and PVM, were isolated and subjected to myeloid subtype annotation and downstream analysis.

Processing of bulk RNA-seq data (BulkMG)

Fresh brain tissue specimens were placed in tissue preservation buffer (Miltenyi Biotech) and stored at 4°C for ≤48hrs before processing using the Adult Brain Dissociation Kit (Miltenyi Biotech), according to manufacturer's instructions. RNase inhibitors (Takara Bio) were used throughout the cell prep. Following de-myelination (Miltenyi de-myelination kit - Miltenyi Biotech) cells were incubated in antibody (CD45: BD Pharmingen, Clone HI30 and CD11b: BD Pharmingen, Clone ICRF44) at 1:500 for

1 hour in the dark at 4°C with end-over-end rotation. Prior to fluorescence-activated cell sorting (FACS), DAPI (Thermoscientific) was added at 1:1000 to facilitate the selection of viable cells. Viable (DAPI negative) CD45/CD11b positive cells were isolated by FACS using a FACSAria flow cytometer (BD Biosciences). Following FACS, cellular concentration and viability were confirmed using a Countess automated cell counter (Life technologies).

RNA was extracted from aliquots of up to 100,000 FACS-sorted microglia using the Arcturus PicoPure RNA isolation kit (Applied Biosystems). RNA-sequencing libraries were generated using the SMARTer Stranded Total RNA-Seq Kit v2 (Takara Bio USA). Libraries were quantified by Qubit HS DNA kit (Life Technologies) and by quantitative PCR (KAPA Biosystems) before sequencing on the Hi-Seq2500 (Illumina) platform obtaining 2x100 paired-end reads.

Count matrices were generated using Kallisto pseudo-mapping[56] using the standard Genecode v38 reference (starting with 235,227 transcripts for 60,535 unique genes). For gene-level analyses, 21,856 features were retained for downstream analyses after filtering for features with CPM>1 in at least 15% of samples. Correct identity of the samples was confirmed by concordance between the genetic variants obtained from RNA-seq with those obtained from ATAC-seq, or directly available genotypes, as available.

Spatial validation using Akoya PhenoCycler

FFPE sections from both AD and control cases were used for the Akoya PhenoCycler experiment. The experiments were performed according to the manufacturer's protocol, with the Neuroinflammation Module, Neuroscience Core Panel and Immune Module provided by Akoya. Briefly, samples were deparaffinized and hydrated. For antigen retrieval, samples were boiled in Tris-EDTA pH 9 for 20 minutes in a programmable pressure cooker. Samples were stained in Antibody Cocktail Solution containing antibodies (**Table S5**) and PhenoCycler Blocking Buffer. Following staining, samples were washed, fixed, and loaded on the PhenoCycler, with data generated using the automatic workflow. Akoya PhenoCycler results were saved as .qpproj files. and protein expression quantified using QuPath[57]. After the sections were annotated, cells were segmented with the QuPath extension StarDist fluorescent cell detection script, with dsb2018_paper.pb as a training model. Protein expression was quantified using raw channel intensity with spatial boundaries of cells inferred by export measurement and export detection commands using QuPath.

Compositional variation analysis

We applied `Crumblr` method (https://diseaseneurogenomics.github.io/crumblr) for testing the variation of cell type composition. Analysis of count ratio data (i.e., fractions) requires special consideration since data is non-normal, heteroskedastic, and spans a low-rank space. While counts can be considered directly using Poisson, negative binomial, or Dirichlet-multinomial models for simple regression applications, these can be problematic since they 1) can be very computationally expensive, 2) can produce poorly calibrated hypothesis tests, and 3) are challenging to extend to other applications. The widely used centered log-ratio (CLR) transform from compositional data analysis makes count ratio data more normal and enables use the of linear models and other standard methods. Yet CLR-transformed data is still highly heteroskedastic: the precision of measurements varies widely. This important factor is not considered by existing methods. `crumblr` uses a fast

asymptotic normal approximation of CLR-transformed counts from a Dirichlet-multinomial distribution to model the sampling variance of the transformed counts. `crumblr` enables incorporating the sampling variance as precision weights to linear (mixed) models in order to increase power and control the false positive rate. `crumblr` also uses a variance stabilizing transform (vst) based on the precision weights to improve the performance of PCA and clustering.

Differential gene expression analysis

We applied `Dreamlet` for differential expression analysis. Building from the previously developed statistical tool Dream[58], it applies linear mixed models to the differential expression problem in single-cell omics data. It starts by aggregating cells by the donor using a pseudobulk approach[59,60] and fits a regression model and cell. For each feature and cell cluster, the following mixed model was applied: Gene expression ~ Intercept + age + (1|sex) + (1|ancestry) + PMI + (1|batch) + (1|source) + phenotype, where categorical and numerical variables were modeled as random and fixed effects, respectively. We ran gene set analysis using the full spectrum of gene-level t-statistics[61].

Constructing gene regulatory networks

We inferred gene regulatory networks with the pySCENIC (v 0.12.1)[30,31] pipeline using a concatenated dataset of FreshMG and PsychAD microglia cohorts. We followed the standard SCENIC expression preprocessing; log normalizing expression counts and selecting highly variable genes (3,192 total) while accounting for batch correction between datasets with scanpy (v1.9.3). The pySCENIC's GRNboost2 (arboreto v0.1.6) method was utilized for gene regulatory network inference. The pySCENIC's cisTarget function with Human motif database v10 (https://resources.aertslab.org/cistarget/motif2tf/motifs-v10nr_clust-nr.hgnc-m0.001-o0.0.tbl) was used to enrich for gene signatures and pruned based on *cis*-regulatory cues using default settings. The "aucell" positional argument was utilized to find the enrichment of regulons across single cells.

To compare regulon enrichment between subtypes, the resulting AUCell matrix was z-score normalized. To assess the concordance between regulon enrichment across subtypes between FreshMG and PsychAD cohorts, we computed the Pearson correlations between normalized z-scores using the pandas `corrwith` function. For each dataset, we computed the normalized regulon enrichment Z-score, and performed a meta-analysis between two datasets using Stouffer's method.

To test whether there is a significant association between the target genes of a TF and disease signatures, we performed Fisher's exact tests between SCENIC GRN target genes and AD risk gene signatures, based on `Dreamlet` DEG analysis. Disease-associated genes with FDR < 0.05 were selected based on four different measures of AD severity, namely, case-control diagnosis, Braak, CERAD, and dementia status. The top 3 regulons were prioritized based on the overall enrichment of the AD DEG signature. The similarity between target genes of regulons was evaluated using the Jaccard similarity index. We found the target genes of the top 3 TFs shared high similarity and clustered distinctly from other regulons (**Figure S4A**). Node centrality was calculated using PageRank analysis, which measures a ranking of the nodes in the graph based on the structure of the incoming links (**Figure S4B**). The mean estimate and -log10FDR of target risk genes in SCENIC regulons are visualized using the importance score edge weights from GRNboost2 with networkX (v3.1). For each regulon, a list of SCENIC GRN TF target genes after cisTarget pruning were obtained and, using

gseapy (v1.0.5)[62], tested for gene-set enrichment based on the Gene Ontology Biological Processes 2023 reference[62]. GO terms were clustered based on their Ward distance between -log10FDR values.

Constructing a pseudotemporal trajectory of AD

We followed the same steps when constructing Braak-stage-informed pseudotimes for the FreshMG and PsychAD datasets, respectively. First, cells that were not assigned a Braak stage were excluded and the data subsetted to contain either only cells annotated as adaptive, homeostatic, ADAM, or PVM. We then computed a k-nearest-neighbor graph using SCANPY[54] (`scanpy.pp.neighbors`, k=30, n_pcs=15) on the PCA embedding, regressed with Harmony. Finally, transport maps were computed between pairs of consecutive disease stages using moscot's TemporalProblem[28].

Identification of driver genes with CellRank 2

To identify drivers of disease progression within each cell type, we first estimated a cell-cell transition matrix T using the `RealTimeKernel` implemented in the CellRank framework[29]. Disease stage heterogeneity within each Braak stage was accounted for by including cell-cell similarity (`conn_weight=0.1`). Since moscot relies on entropical regularization to solve the underlying optimal transport problem, it returns dense transport maps, leading to negligible transition probabilities and large memory requirements when computing the transition matrix. To render our downstream analysis computationally feasible, we thresholded the constructed transition matrix using the `RealTimeKernel`'s automatic thresholding scheme (`threshold="auto"`).

Based on the `RealTimeKernel`-derived transition matrix, we estimated terminal states with the GPCCA estimator[29,63]. In the case of adaptive and homeostatic cells, we computed terminal states at the resolution of subclasses (MG_Adapt, MG_Homeo), for the PVM subset based on the subtype (PVM_CD163, ADAM_GPNMB) annotation. We confirmed the estimated terminal states as biologically plausible by quantifying the cell type and disease stage purity, respectively. Given n cells (n=30 in this study) identifying a terminal state, the disease stage purity is defined as

$$\frac{|C \cap B_j|}{n},$$

with C denoting the set of cells identifying the terminal state, B_j the set of cells with Braak stage j, and $|\cdot|$ the cardinality of the set. Cell type purity is defined similarly.

We computed fate probabilities towards each terminal state next and identified driver genes by correlating gene expression with fate probabilities as outlined in the corresponding CellRank tutorial. To compare the gene ranking between the FreshMG and PsychAD cohorts, we subsetted to genes present in both datasets and computed the Pearson correlation coefficient between the gene-specific correlations $r^{(g)}_{\mathrm{FMG}}$ and $r^{(g)}_{\mathrm{PAD}}$ by CellRank:

$$r = \frac{\sum_g \left(r^{(g)}_{\mathrm{FMG}} - \bar{r}^{(g)}_{\mathrm{FMG}}\right)\left(r^{(g)}_{\mathrm{PAD}} - \bar{r}^{(g)}_{\mathrm{PAD}}\right)}{\sqrt{\sum_g \left(r^{(g)}_{\mathrm{FMG}} - \bar{r}^{(g)}_{\mathrm{FMG}}\right)^2}\sqrt{\sum_g \left(r^{(g)}_{\mathrm{FMG}} - \bar{r}^{(g)}_{\mathrm{FMG}}\right)^2}},$$

with sample mean $\bar{r}$.

Pseudotime construction

DPT[64] is traditionally calculated on a symmetric connectivity matrix, the constructed transition matrix T is, however, non-symmetric. To compute DPT based on our Braak-stage-informed transition matrix, we, thus, symmetrized T via $\frac{1}{2}(T + T^{\mathrm{T}})$ and row-normalized entries. Following, we computed diffusion pseudotime using SCANPY's `scanpy.tl.dpt` function with default values. The corresponding initial cells were identified via extreme points in diffusion components[65]. To verify that the constructed pseudotime recapitulates our findings that homeostatic FRMD4A cells decline with disease progression, while the number of homeostatic PICALM cells increases, we stratified pseudotime by subtype (**Figure 5C**). Next, we performed a Welch's t-test to validate that homeostatic FRMD4A cells are assigned significantly smaller pseudotime values compared to the set of homeostatic PICALM. We performed similar Welch's t-tests to assess the change between consecutive Braak stages. For each dataset, we first computed the T-statistics from Welch's t-test and then combined them across datasets using Stouffer's method: Considering T-statistics T^{freshMG} and T^{psychAD}, the reported T-statistic T given by

$$T = \frac{T^{\mathrm{freshMG}} + T^{\mathrm{psychAD}}}{2}.$$

As a final analysis, we correlated fate probabilities with gene expression for each lineage to identify putative driver genes for each lineage. Here, we included genes present in both datasets. To confirm concordance between the two independent cohorts, we computed the Spearman correlation between the two rankings.

Inferring cell-to-cell interactions

CCI analysis relies on inference of ligand-receptor interactions given a gene expression matrix with annotated cell types. By comparing to a known set of ligand-receptor (LR) pairs (reference "resource"), CCI can be inferred and scored through a variety of methods. LR pairs across cell types that are expressed above a set threshold and are differentially co-expressed, are inferred to represent interactions between the cell types. The `LIANA`[32] framework combines several established CCI inference methods and reference resources. We used the Python implementation of `LIANA` (v0.1.8). For each donor in FreshMG and PsychAD, we separately ran the standard rank_aggregate pipeline (resource_name='consensus', expr_prop=0.1, min_cells=5, n_perms=1000). This utilizes the RobustRankAggregate[66] algorithm for aggregating scores from several methods (CellPhoneDB, Connectome, log2FC, NATMI, SingleCellSignalR, and CellChat)[32,67–71] into a single magnitude score for each CCI. We used the standard "consensus" CCI reference resource provided by `LIANA`. For a given

pair of cell types and pair of genes to be considered for CCI scoring, each gene must be found in the reference resource, and be expressed in at least 10% of each involved cell type and in at least 5 cells. We evaluated CCI using the rank aggregate magnitude score for cell types annotated at the subtype level. We negative log-transformed these scores so that low magnitude CCI are scored close to 0 and normalized the score distribution. Additional post-processing was then performed to filter out CCI that were not expressed in both FreshMG and PsychAD.

To assess the concordance between FreshMG and PsychAD CCI scores, we first computed the average CCI score for each CCI across donors within each cohort. We then grouped the average scores by each possible pair of subtypes and measured the Spearman rank correlation. We additionally computed an activity z-score for the frequency of CCI involving each subtype pair (**Figure 5A**).

To determine whether CCI are differentially expressed in AD, we applied Dream[13] to construct a linear mixed model and account for technical and biological covariates (batch, age, sex). Linear mixed model regression was performed separately for FreshMG and PsychAD across four measures of AD progression: diagnosis (dx_AD), CERAD score, Braak stage, and dementia status. We then also used Dream to meta-analyze CCI that are significantly differentially expressed across both cohorts and to estimate the log fold change effect of each CCI for each AD progression measure (**Figure 5B, Supplementary Figure 6D**).

To evaluate whether significantly differentially expressed CCI are enriched in AD genome-wide association studies (GWAS), we utilize the Multi-marker Analysis of GenoMic Annotation (`MAGMA`)[72]. The `MAGMA` gene set was constructed from significantly differentially expressed CCI LR pairs. These include CCI with at least a false discovery rate (FDR) of 0.05 significance, as determined by the Dream regression meta-analyses across each AD progression measure. We normalized the log2 fold change (log2FC) values for these CCI for each progression measure and split the CCI into AD (log2FC > 0) and CTRL (log2FC < 0) groups. Top LR pairs for AD and CTRL were then ranked by selecting the LR pairs involved in the CCI with the largest absolute normalized log2FC values. The ranked AD and CTRL LR pair gene sets were analyzed by the `MAGMA` for enrichment across several GWAS (**Figure 5C**). We generated a network diagram to highlight the top-ranked three AD and CTRL CCI in our gene set (**Figure 5D**).

To investigate the biological processes AD-relevant CCI are involved in, we performed a gene set enrichment analysis using the cell2cell[73] package and the human Gene Ontology Biological Processes 2023 reference[74,75] The AD-relevant CCI gene set was constructed in the same way as the `MAGMA` analysis gene set. We aggregated these CCI by the sender and receiver subtype to improve statistical power, and then computed normalized enrichment scores for each process based on our AD-relevant LR pair gene set. We prioritized processes that pass the 0.05 FDR significance threshold (**Figure 5E**).

Analysis of genetic heritability of AD polygenic risk

We established a standardized pipeline for Multi-marker Analysis of GenoMic Annotation (`MAGMA`) followed by single-cell Disease-Relevance Scoring (`scDRS`). `MAGMA` incorporates the association p-values of genetic variants from the latest AD genome-wide association study (GWAS)[9]. We applied `MAGMA` using a standard window of 35kbp upstream and 10kbp downstream around the gene body. We executed `scDRS` using the top 1000 gene weights, sorted by Z score. The `MAGMA` and `scDRS`

pipeline was run separately on FreshMG and PsychAD single-cell cohorts using the following parameters. `MAGMA` was run using `-snp-loc g1000_eur.bim` (SNP location file corresponding to the Phase 3 1000 Genome Project) and `--gene-loc NCBI38.gene.loc` (gene location file from NCBI build 38). Both files were obtained from https://ctg.cncr.nl/software/magma.

To justify the use of the top 1000 genes for `scDRS`, genes were sorted by `MAGMA` z-score and top 200, 500, and 1000 genes were used for testing the concordance between cohorts in subsequent `scDRS` analysis. `scDRS` command `scdrs.preprocess` was run with the default parameters of `n_mean_bin=20`, `n_var_bin=20w`, while `scdrs.score_cell` was run using `n_ctrl=100`. Best concordance between FreshMG and PsychAD cohorts of average `scDRS` scores (per myeloids subtype) was achieved using top 1000 genes (Pearson's R of top n `MAGMA` genes; n=200 was 0.85, n=500 was 0.94, n=1000 was 0.97, hence we proceeded with the analysis using top 1000 `MAGMA` genes).

We tested the following GWAS summary stats in `scDRS`/`MAGMA` pipeline: AD[9], MS[76], PD[77], MDD[78], ASD[79], BD[80], SCZ[81] and ALS[82]. The `scDRS` scores were highly reproducible between the FreshMG and PsychAD cohorts, with AD having the greatest correlation (r=0.97), followed by MS (r=0.92), PD (r=0.83), MDD (r=0.79), ASD (r=0.71) and ALS (r=0.66). The lowest correlation was noted for BD (r=0.60) and SCZ (r=0.42), both of which are neuropsychiatric diseases. This emphasized a high reproducibility of myeloid cell heritability estimates in AD, MS, and PD, all of which are neurodegenerative diseases with progressive damage to neuronal cell types, and whose primary pathogenic mechanisms involve non-neuronal cells, including microglia and PVM. Neuropsychiatric diseases such as SCZ and BP, whose primary risks are enriched in synaptic dysfunctions of neurons, had poor correlations (**Figure S7B**). Per cluster association z-scores (`scDRS` assoc_mcz) were obtained using `scdrs.method.downstream_group_analysis` function. Stouffer's method for meta-analysis was used to combine FreshMG and PsychAD `scDRS` association z-scores using the number of cells per cell cluster as cluster weights. For each cohort, `scDRS` permutation p-values per cluster were combined using Stouffer's method on p-values with weights (cell counts of clusters). Meta p-values were further corrected for multiple testing using FDR correction (both per cell type and using all cell types combined). A more stringent correction was achieved using the per-cell type method hence we applied this correction.

Case-control residual analysis of heritability estimates

To test for concordance in heritability estimates, we separated cases and controls and repeated `MAGMA`/`scDRS` pipe for both FreshMG and PsychAD single-cell cohorts. We calculated meta z-scores for cases and controls separately using Stouffer's method with weights (cluster's cell counts). We correlated meta-z-scores of cases and controls using linear regression and derived residuals as a deviation from the regression line, indicative of per-cell cluster heritability. For both cohorts, per-cell pseudotime scores were correlated with per-cell `scDRS` scores for every microglia subtype and the correlation coefficient was calculated using Spearman's test. FreshMG and PsychAD coefficients were combined in a meta value by applying Stouffer's method and the number of cells per cell cluster as cluster weights.

Processing of genotypes

The FreshMG cohort genotype data consisted of samples from the Mount Sinai and Rush brain banks, as has been previously described (PMID: 35931864). The PsychAD cohort genotype data consisted of samples from the Mount Sinai brain bank. For both cohorts, DNA extraction and genotyping were performed as described previously (PMID: 35931864). In brief, genomic DNA was extracted from buffy coat or frozen brain tissue using the QIAamp DNA Mini Kit (Qiagen), according to the manufacturer's instructions. Samples were genotyped using the Infinium Psych Chip Array (Illumina) at the Mount Sinai Sequencing Core. Pre-imputation processing consisted of running the quality control script HRC-1000G-check-bim.pl from the McCarthy Lab Group (https://www.well.ox.ac.uk/~wrayner/tools/), using the Trans-Omics for Precision Medicine (TOPMed) reference (PMID: 33568819). Genotypes were then phased and imputed on the TOPMed Imputation Server (https://imputation.biodatacatalyst.nhlbi.nih.gov). Samples with a mismatch between one's self-reported and genetically inferred sex, suspected sex chromosome aneuploidies, high relatedness as defined by the KING kinship coefficient[83] (KING > 0.177), and outlier heterozygosity (+/- 3SD from mean) were removed. Additionally, samples with a sample-level missingness > 0.05 were removed and calculated within a subset of high-quality variants (variant-level missingness ≤ 0.02).

Samples of European (EUR) ancestry, as defined by assignment to the EUR superpopulation described by the 1000 Genomes Project[84,85], were isolated using a 3SD ellipsoid method. Genotypes were first merged with GRCh38 v2a 1000 Genomes Project data (https://wellcomeopenresearch.org/articles/4-50)[85] using BCFtools version 1.9[86]. PLINK 2.0[87] was then used to calculate the merged genotypes' principal components (PCs), following filtering (minor allele frequency (MAF) ≥ 0.01, Hardy-Weinberg equilibrium (HWE) p-value$\geq 1\times10^{-10}$, variant-level missingness ≤ 0.01, regions with high linkage disequilibrium (LD) removed) and LD pruning (window size = 1000 kb, step size = 10, r2 = 0.2) steps. An ellipsoid with a radius of 3SD, corresponding to the 1000 Genomes Project EUR superpopulation, was constructed using the first three genotype PCs. Only samples that fell within this ellipsoid (FreshMG: n=178, PsychAD: n=759) were retained for subsequent variant-level filtering. Autosomal bi-allelic variants with an imputation $R^2 > 0.8$, HWE p-value$\geq 1\times10^{-6}$, and variant-level missingness ≤ 0.02 were retained. Genotypes were then annotated with ancestry-specific MAF values from the National Center for Biotechnology Information's Allele Frequency Aggregator (ALFA) (https://ftp.ncbi.nih.gov/snp/population_frequency/latest_release/). Only variants with an ancestry-specific ALFA MAF ≥ 0.01 (FreshMG: n=10,828,658, PsychAD: n=18,490,352) were retained.

PRS calculation

Polygenic risk scores (PRS) were calculated on the FreshMG and PsychAD cohort samples using AD GWAS summary statistics[9]. The PRS-CS-auto method[88] was used to apply continuous shrinkage priors to the effect sizes from these summary statistics. A EUR LD reference panel provided by the developers of PRS-CS was utilized (https://github.com/getian107/PRScs), which draws from 1000 Genomes Project data[85]. The following PRS-CS default settings were used: parameter a in the γ-γ prior = 1, parameter b in the γ-γ prior = 0.5, MCMC iterations = 1000, number of burn-in iterations = 500, and thinning of the Markov chain factor = 5. The global shrinkage parameter phi was set using a fully Bayesian determination method. Individual-level PRS were calculated using PLINK 2.0[87].

Transcriptional variation with PRS

PRS analysis was performed using the `Dreamlet` package (v0.99) that applies a linear mixed model with precision weights to fit a regression model. Instead of applying a fixed contrast between two coefficients, we used AD polygenic risk scores (PRS) as a continuous variable to test for the presence of non-linear effects on variance. We subtracted admixed donors from this analysis using the PCA analysis of the first 5 genotype PCs (2 individuals were removed from PsychAD and 1 from the FreshMG cohort as clear outliers in the PCA plots). In addition, we removed a cluster of 200 individuals that were clustering separately on the PC1/PC2 plot in psychAD as admixed individuals with a percentage of a non-EUR ancestry ranging from 2.5%-10% (mainly AFR ancestry). Similarly, we removed 6 more donors from FreshMG that were outliers and showed decreased EUR and increased AFR ancestry. The removal of admixed individuals allowed us to achieve a high level of concordance in `Dreamlet` PRS analysis between FreshMG and PsychAD cohorts. For the PsychAD cohort, we considered the following covariates to model the variance: source, log(n_counts), PMI, age, AD_Bellenguez PRS, sex, dx, genotype PCs 1-5. For the FreshMG cohort, we used the same set of covariates, in addition to the sequencing batch as a covariate. We further performed a fixed-effect meta-analysis with the R metafor package using effect sizes and standard error from the `Dreamlet` output (metafor `rma` function with parameters: `yi = logFC, sei = logFC / t, method = "FE"`). Meta p-values were further corrected for multiple testing using FDR correction (using the per-cell type correction).

To test whether the observed cell proportions change with PRS as a continuous variable, we applied the `crumblr` R package to each cohort and performed analysis of count ratio data with precision-weighted linear mixed models. Similar to the `Dreamlet` analysis, we modeled PRS as a continuous variable and tested for non-linear effects. FreshMG and PsychAD effect sizes were combined in a meta value using Stouffer's method and the number of cells per cell cluster as cluster weights.

Supplementary Information

Assessment of ex-vivo activation in Microglia

Microglia are highly reactive cells and prolonged exposure to non-physiological conditions could induce unintended responses[89,90]. It has previously been shown that enzymatic dissociation of brain tissue performed at elevated temperatures can stimulate microglial cells, termed *ex-vivo* activated microglia (exAM)[17,91,92]. Failing to account for dissociation-induced changes can compromise the interpretation of microglial biology[93]. Although artificially induced, we wondered if the proportion of exAM might reflect microglial reactivity to pathological lesions and, as such, could provide valuable insights into disease mechanisms. In this study, we are uniquely positioned to interrogate the impact of exAM as myeloids from the FreshMG cohort were isolated via enzymatic dissociation at 37°C while samples from the PsychAD cohort, consisting of flash frozen tissue, were processed using an ice-cold, enzyme free, dissociation buffer. Assessing the extent and prevalence of the exAM signature in our datasets, we observed a distinct subtype in the FreshMG dataset, called MG_exAM_ERN1, which accounts for about 14.8% of all myeloid cells, and shows up-regulation of genes implicated in cell-cell adhesion, including ERN1, PLK2 (serine/threonine-protein kinase), CSKMT, and SNHG5. Notably, PLK2 is an enzyme regulating synaptic activity and has been implicated in stimulating Aβ production[94,95]. We note that the PsychAD dataset had very few cells identified as the ERN1 subtype, which belongs to the exAM subclass. This result suggests a number of possibilities. Given that the transcripts in the PsychAD dataset originate from the nuclear fraction of frozen nuclei and are free from enzymatic treatment or dissociation bias, it's possible that the exAM cluster is predominantly derived from fresh tissue and is artificially induced during processing[17]. In addition, the transcripts that define the ERN1 subtype may be predominantly cytoplasmic and are, thus, missing from the PsychAD samples. Moreover, we acknowledge that we observe a substantial compositional variation among three different cohorts. However, to properly assess the extent of compositional variation, we need to model this by accounting for various technical effects, including the dissection bias. In conclusion, this comparative analysis confirmed the robustness of the human microglial taxonomy, independent of tissue source, agonal state, and postmortem interval (PMI).

Disease trajectory of human myeloid cells

Understanding the dynamic changes that take place during the onset and progression of AD at a molecular-level requires modeling of gene expression change along the measures of disease progression and identifying corresponding putative drivers. Analyses solely based on donor-level clinical covariate, such as CERAD scores or Braak staging, are limited due to the discrete nature of these variables. However, we reasoned that each labeled disease stage contains observations spanning the range of disease development and sought to identify drivers and order cells along a disease pseudotime trajectory.

Trajectory inference allows us to expand our ability to describe molecular changes at single-cell resolution. Trajectory inference methods have been remarkably successful in describing normal developmental processes faithfully and identifying regulatory mechanisms based on single-cell sequencing data[64,96,97]. Classical pseudotime algorithms order cells along a developmental axis, with early and late cells being assigned low and high pseudotime values, respectively. To take advantage

of advances in experimental design and different sources of information, alternative methods have been developed. To link observations across experimental time points, optimal transport-based solutions have been proposed to assign each measurement from one experimental time point its likely future state in the following via a probabilistic assignment in the form of transport maps[28,98]. These couplings can then be used to quantify cellular change and determine the fate and putative drivers using CellRank 2's *RealTimeKernel*[29]. Using it together with and GPCCA estimator[29,63], we automatically inferred the terminal states of the disease dynamics. As AD onset and progression do not result in the emergence or disappearance of novel cell types, we expected to recover the major subclasses of the data, *i.e.*, adaptive, homeostatic, ADAM, or PVM. In concordance with this ground truth, we recovered the terminal states accordingly and observed a high terminal state purity defined by the fraction of cells with the correct cell type and Braak stage six (terminal state purity was 1.0 for all subclasses used). Following, we computed driver genes of these respective fates by correlating gene expression with fate probabilities (**Methods**). We replicated the same analysis using the PsychAD and observed high concordance between the correlations between gene expression and fate probabilities associated with each gene (**Figure S5B**). While CellRank 2 identifies putative lineage drivers, it does not align cells along the disease trajectory. To construct this orthogonal information, we used the transition matrix computed by the *RealTimeKernel* to compute a Braak-stage-informed pseudotime for the two subsets of the data in a similar fashion as previously proposed for experimental time points by the CellRank 2 study[29] (**Methods**). We tested if the gene expression changes over pseudotime inferred on the FreshMG cohort agrees with the corresponding change in the PsychAD dataset. As expected, the inferred change in gene expression showed high concordance between the two independent cohorts for both homeostatic and adaptive lineages (**Figure S5C, D**).

Data availability

The single-cell dataset, clinical metadata, and analysis outputs are available via the AD Knowledge Portal (https://adknowledgeportal.org). The AD Knowledge Portal is a platform for accessing data, analyses, and tools generated by the Accelerating Medicines Partnership (AMP-AD) Target Discovery Program and other National Institute on Aging (NIA)-supported programs to enable open-science practices and accelerate translational learning. The data, analyses, and tools are shared early in the research cycle without a publication embargo on secondary use. Data is available for general research use according to the following requirements for data access and data attribution (https://adknowledgeportal.org/DataAccess/Instructions). For access to data described in this manuscript see:
https://www.synapse.org/#!Synapse:syn52795287

Code availability

All the source codes used in this study are available via GitHub:
https://github.com/DiseaseNeuroGenomics/scMyeloidAD

Acknowledgments

We would like to express our deep gratitude to the patients and their families who generously donated the invaluable biological material essential for the success of this study. We are profoundly indebted to their participation and commitment to advancing scientific knowledge and improving human health. We acknowledge the National Institute on Aging for their generous support in funding this research with the following NIH grants: R01AG065582 (PR, VH), R01AG067025 (PR, VH), and R01AG082185 (PR, VH, DL). ROSMAP is supported by P30AG10161, P30AG72975, R01AG15819, R01AG17917. U01AG46152, U01AG61356. ROSMAP resources can be requested at https://www.radc.rush.edu. We also thank Stefano Marenco and Pavan Auluck from the Human Brain Collection Core (HBCC) at the National Institute of Mental Health-Intramural Research Program for their contribution to the resource. The HBCC is supported by the NIMH-IRP under project ZIC MH002903.

Author contributions

Conceptualization: P.R., J.F.F., V.H., D.L.
Methodology: P.R., J.F.F., D.L., G.H.
Investigation: S.P.K., Z.S., S.A., M.A., C.C., A.H., J.F.F.,
Formal analysis: D.L., C.P., C.S., M.P., P.W., R.K., J.B., P.N.M., S.Z., K.T., D.M.
Validation: X.W., J.F.F.
Resources: V.H., D.A.B., C.P.K., K.G.B., R.S.
Writing: D.L., J.F.F., P.R., with contributions from all authors.
Visualization: D.L.
Supervision: P.R., D.L., J.F.F., G.H., V.H., F.J.T., G.V., G.C.Y.

Project administration: D.L., P.R.
Funding acquisition: P.R., V.H., D.L.
All authors read and approved the final draft of the paper.

Competing interests

F.J.T. consults for Immunai Inc., Singularity Bio B.V., CytoReason Ltd, Omniscope Ltd, Cellarity, and has ownership interest in Dermagnostix GmbH and Cellarity. The remaining authors declare no competing interests.

Materials & Correspondence

Correspondence to Donghoon Lee and Panos Roussos.

Figure Legends

Figure 1. Overview of the two independent single-cell resources utilized to define human myeloid subtypes. **(A)** *FreshMG* discovery cohort (scRNA-seq) using live human microglial cells from postmortem autopsies and **(B)** *PsychAD* replication cohort (snRNA-seq) using flash-frozen tissues and *in-silico* sorted for microglia and PVMs. **(C)** Unified taxonomy of human microglial states. **(D)** Correlation of the cell state taxonomy between *FreshMG* and *PsychAD* datasets. Pairwise Pearson correlation of highly variable genes common in both datasets. **(E)** Human myeloid subtype-specific gene expression markers. Top-left: *FreshMG*. Bottom-right: *PsychAD*. **(F)** Compositional differences of myeloid subtypes between *FreshMG* and *PsychAD* datasets. In addition, the composition was compared to an independent cohort named *LivinMG* sourced from biopsy specimens and represents living brains. **(G)** Representative image of Akoya PhenoCycler multiplex immunofluorescence results showing $CD163^{+}$/IBA-1^{+} cells are enriched near blood vessels (outlined by gray line), labeled by Collagen IV. Scale bar 20 μm.

Figure 2. Transcriptional variation of human myeloid cells using disease-free subset. **(A)** Compositional variation of myeloid subtypes by age, sex, and the interaction between age and sex. CLR transformed composition data was modeled using a linear mixed model accounting for technical batch effects including tissue sources and sequencing pools and donor effects including age, sex, genetic ancestry, and PMI (see methods on `crumblr`). Fixed effect meta-analysis using results from FreshMG and PsychAD cohorts. **(B)** Covariate adjusted compositional variation with normal aging. **(C)** Differentially expressed genes by myeloid subtypes associated with normal aging. CLR: centered-log-ratio.

Figure 3. Transcriptional variation of human myeloid cells with AD. **(A)** Compositional variation of myeloid subtypes by four different neuropathological measures of the AD progression; diagnosis (dx_AD), CERAD, Braak staging, and dementia status, after accounting for technical and donor-level covariates. Fixed effect meta-analysis using both FreshMG and PsychAD cohorts. **(B)** Covariate adjusted compositional variation with Braak staging. **(C)** Comparison of compositional variation between normal aging and AD. Subtypes are weighted by the inverse of standard error. **(D)** Differentially expressed genes by four different measures of AD neuropathology adjusted for technical and donor-level covariates. Fixed effect meta-analysis using both FreshMG and PsychAD cohorts. **(E)** Comparison of myeloid subtypes to homogenate bulk RNA-seq expression stratified by AD diagnosis. **(F)** Distribution of disease pseudotime between two homeostatic microglia subtypes. **(G)** Changes in disease pseudotime between two adjacent disease stages measured by Braak and CERAD.

Figure 4. Prioritization of upstream master regulators of AD using GRN analysis. **(A)** Schematic overview of GRN inference and TF-gene regulon enrichment. **(B)** Concordance of normalized regulon activity scores (AUCell) between FreshMG and PsychAD cohorts. Pairwise Pearson correlation. **(C)** Enrichment of regulon by subtypes. Meta-analysis of consensus regulon enrichment Z-score with Stouffer's correction between FreshMG and PsychAD cohorts. Top 3 regulons per each subtype shown. **(D)** Enrichment of AD gene signatures by regulons. Fisher's exact tests for enrichment of

differentially expressed gene signature in regulon target genes across myeloid subtypes. **(E)** TFs that modulate AD risk genes. Gene regulatory network visualization of KLF12, MITF, and GLIS3 TFs and downstream target risk genes. Node colors represent gene expression changes from `Dreamlet` analysis. Edge weights represent importance scores inferred from the SCENIC pipeline. **(F)** Pathway enrichment analysis of KLF12, MITF, and GLIS3 regulons using GO biological process database.

Figure 5. Cell-cell interactions among human myeloid cells. **(A)** Concordance of CCI scores between the FreshMG and PsychAD cohorts. Pairwise Spearman correlation using aggregated CCI scores by subtype. Row labels correspond to the sender or ligand-producing cell. Column labels correspond to the receiver or receptor-producing cell. **(B)** Differential CCI analysis based on Braak stages. Meta-analysis of linear mixed model regression using both FreshMG and PsychAD cohorts. Estimated log fold change corresponds to increased representation in the high Braak stage (red) vs. the low Braak stage (blue). **(C)** `MAGMA` enrichment analysis on differential CCI, stratified by direction of regulation (AD vs CTRL) and role of interaction (ligands, receptors, or both). **(D)** Visualization of top CCI pairs. Top: AD, Bottom: controls. Nodes represent each subtype and directional edge weights represent the importance of interaction. The edge color represents the estimated log fold change from differential CCI analysis. **(E)** Gene set enrichment analysis of CCI pairs using Gene Ontology Biological Processes. CCIs aggregated by subtype, direction of regulation (AD vs CTRL), and role of interaction (ligands or receptors). The color scale represents the normalized enrichment score (NES). The dot size represents the FDR significance. + marks FDR < 0.05.

Figure 6. Heritability estimation of AD risk using GWAS disease-associated genes and polygenic risk scoring (`scDRS` and PRS). **(A)** Meta-analysis of `scDRS` z-scores per myeloid subtype across 8 brain diseases. The asterisk denotes FDR < 0.05. **(B)** Correlation between `scDRS` meta z-scores and `crumblr` estimate of compositional variation by dx_AD as a coefficient. Weighted Pearson's correlation using average -log10(P-value) as weights. **(C)** Correlation between `crumblr` estimate of compositional variation by PRS as a coefficient against `crumblr` estimate of compositional variation by dx_AD as a coefficient. Weighted Pearson's correlation using inverse of average of standard error as weights. Circle denotes `crumblr` analysis using all donors while triangle denotes `crumblr` analysis using controls only. SCZ: schizophrenia, BD: bipolar disorder, MDD: major depressive disorder, ASD: autism spectrum disorder, MS: multiple sclerosis, ALS: amyotrophic lateral sclerosis, and PD: Parkinson's disease.

Supplementary Figure S1. **(A-J)** Demographic and clinical metadata of the FreshMG cohort. **(K)** Pairwise correlation of clinical variables. **(L)** Variance partition of covariates used in the FreshMG cohort. **(M-V)** Demographic and clinical metadata of the PsychAD cohort. **(W)** Pairwise correlation of clinical variables. **(X)** Variance partition of covariates used in the PsychAD cohort.

Supplementary Figure S2. **(A)** Schematic overview of the FreshMG single-cell data processing. **(B)** Schematic overview of the PsychAD single-cell data processing. **(C)** Pairwise Pearson correlation between myeloid subtypes. **(D)** Subtype-specific surface protein markers, CITE-seq antibody derived tags (ADT), replicating the cell state taxonomy of RNA. **(E)** Annotation of the LivinMG (biopsies)

dataset using the FreshMG as a reference annotation. **(F)** Enrichment of gene signatures from published studies. DAM: disease-associated microglia, LDAM: lipid-droplet accumulating microglia, DIM: disease inflammatory macrophages, YAM: youth-associated microglia, iMGL: human stem-cell-differentiated microglia.

Supplementary Figure S3. **(A)** Pathway enrichment analysis of human myeloid subtypes using GO biological process database. Subtype-defining markers were prioritized using Mann–Whitney U tests of one-vs-the-rest subsets. **(B)** Pathway enrichment analysis using the GO biological process database for aging. **(C)** Pathway enrichment analysis using the GO biological process database for AD phenotypes.

Supplementary Figure S4. **(A)** Jaccard Index of SCENIC regulon target genes. **(B)** PageRank centrality scores of the GRN nodes.

Supplementary Figure S5. Pseudo-temporal change of human myeloid states during AD neurodegeneration. **(A)** Schematic overview of the pseudotime inference based on Braak staging. State transition matrix inferred using the optimal transport algorithm (Moscot). **(B)** Concordance of the putative drivers of two fates (adaptive and homeostatic) between FreshMG and PsychAD cohorts. Putative drivers are defined by correlating gene expression with fate probabilities. **(C-D)** Correlation between gene expression change and inferred disease pseudotime. The concordance of the correlation between FreshMG and PsychAD cohorts is plotted for homeostatic and adaptive lineages separately. **(E)** Spearman correlations between inferred disease pseudotime and different measures of AD phenotypes.

Supplementary Figure S6. **(A)** `MAGMA` analysis of aging signature. **(B)** `MAGMA` analysis of AD signature. **(C)** Schematic overview of the methods for inferring disease-associated CCI. **(D)** Differential CCI using 4 different measures of AD phenotypes; AD cases vs controls (dx_AD), CERAD, Braak, and dementia. **(E)** Concordance of differential CCIs among different measures of AD phenotypes.

Supplementary Figure S7. **(A)** Schematic overview of the `scDRS` and PRS methods for evaluating heritability of AD risk. **(B)** Correlation of the polygenic risk scores between the FreshMG and PsychAD cohorts across human brain disorders. **(C)** Correlation of `scDRS` z-scores between AD cases and controls. **(D)** Compositional variation of PRS, side-by-side comparison between one using all donors and another using control donors only. Weighted Pearson's correlation using inverse of average of standard error as weights. SCZ: schizophrenia, BD: bipolar disorder, MDD: major depressive disorder, ASD: autism spectrum disorder, MS: multiple sclerosis, ALS: amyotrophic lateral sclerosis, and PD: Parkinson's disease.

Supplementary Tables

Supplementary Table 1. Myeloid taxonomy of the FreshMG dataset.
Supplementary Table 2. Myeloid taxonomy of the PsychAD dataset.
Supplementary Table 3. Clinical metadata of the FreshMG dataset.
Supplementary Table 4. Clinical metadata of the PsychAD dataset.
Supplementary Table 5. List of Akoya antibodies.
Supplementary Table 6. Compositional variation meta-analysis for Aging.
Supplementary Table 7. Compositional variation meta-analysis for AD phenotypes.
Supplementary Table 8. Differential gene expression meta-analysis for Aging.
Supplementary Table 9. Differential gene expression meta-analysis for AD phenotypes.
Supplementary Table 10. Pseudotime estimates for the FreshMG dataset.
Supplementary Table 11. Pseudotime estimates for the PsychAD dataset.
Supplementary Table 12. GRN adjacencies.
Supplementary Table 13. Regulon enrichment analysis for AD signatures.
Supplementary Table 14. Differential CCI meta-analysis.
Supplementary Table 15. scDRS scores from meta-analysis.

Reference

1. Hickman, S., Izzy, S., Sen, P., Morsett, L. & El Khoury, J. Microglia in neurodegeneration. *Nat. Neurosci.* **21**, 1359–1369 (2018).
2. Leng, F. & Edison, P. Neuroinflammation and microglial activation in Alzheimer disease: where do we go from here? *Nat. Rev. Neurol.* **17**, 157–172 (2021).
3. De Schepper, S. *et al.* Perivascular cells induce microglial phagocytic states and synaptic engulfment via SPP1 in mouse models of Alzheimer's disease. *Nat. Neurosci.* **26**, 406–415 (2023).
4. Keren-Shaul, H. *et al.* A Unique Microglia Type Associated with Restricting Development of Alzheimer's Disease. *Cell* **169**, 1276–1290.e17 (2017).
5. Paolicelli, R. C. *et al.* Microglia states and nomenclature: A field at its crossroads. *Neuron* **110**, 3458–3483 (2022).
6. Long, J. M. & Holtzman, D. M. Alzheimer Disease: An Update on Pathobiology and Treatment Strategies. *Cell* **179**, 312–339 (2019).
7. Jonsson, T. *et al.* Variant of TREM2 associated with the risk of Alzheimer's disease. *N. Engl. J. Med.* **368**, 107–116 (2013).
8. Guerreiro, R. *et al.* TREM2 variants in Alzheimer's disease. *N. Engl. J. Med.* **368**, 117–127 (2013).
9. Bellenguez, C. *et al.* New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat. Genet.* **54**, 412–436 (2022).
10. Kosoy, R. *et al.* Genetics of the human microglia regulome refines Alzheimer's disease risk loci. *Nat. Genet.* **54**, 1145–1154 (2022).
11. Chen, Y. & Colonna, M. Two-faced behavior of microglia in Alzheimer's disease. *Nature neuroscience* vol. 25 3–4 (2022).
12. Olah, M. *et al.* Single cell RNA sequencing of human microglia uncovers a subset associated with Alzheimer's disease. *Nat. Commun.* **11**, 6129 (2020).

13. Mathys, H. *et al.* Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332–337 (2019).
14. Masuda, T. *et al.* Spatial and temporal heterogeneity of mouse and human microglia at single-cell resolution. *Nature* **566**, 388–392 (2019).
15. Sun, N. *et al.* Human microglial state dynamics in Alzheimer's disease progression. *Cell* **186**, 4386–4403.e29 (2023).
16. Thrupp, N. *et al.* Single-Nucleus RNA-Seq Is Not Suitable for Detection of Microglial Activation Genes in Humans. *Cell Rep.* **32**, 108189 (2020).
17. Marsh, S. E. *et al.* Dissection of artifactual and confounding glial signatures by single-cell sequencing of mouse and human brain. *Nat. Neurosci.* **25**, 306–316 (2022).
18. Stoeckius, M. *et al.* Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868 (2017).
19. Mirra, S. S. *et al.* 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 (1991).
20. Braak, H. & Braak, E. Evolution of the neuropathology of Alzheimer's disease. *Acta Neurol. Scand. Suppl.* **165**, 3–12 (1996).
21. Ajoolabady, A., Lindholm, D., Ren, J. & Pratico, D. ER stress and UPR in Alzheimer's disease: mechanisms, pathogenesis, treatments. *Cell Death Dis.* **13**, 706 (2022).
22. Smith, A. M. *et al.* Diverse human astrocyte and microglial transcriptional responses to Alzheimer's pathology. *Acta Neuropathol.* **143**, 75–91 (2022).
23. Dolan, M.-J. *et al.* Exposure of iPSC-derived human microglia to brain substrates enables the generation and manipulation of diverse transcriptional states in vitro. *Nat. Immunol.* **24**, 1382–1390 (2023).

24. Silvin, A. *et al.* Dual ontogeny of disease-associated microglia and disease inflammatory macrophages in aging and neurodegeneration. *Immunity* **55**, 1448–1465.e6 (2022).
25. Lieber, A. C. *et al.* Biopsy During Minimally Invasive Intracerebral Hemorrhage Clot Evacuation. *World Neurosurg.* (2018) doi:10.1016/j.wneu.2018.12.058.
26. Xu, C. *et al.* Probabilistic harmonization and annotation of single-cell transcriptomics data with deep generative models. *Mol. Syst. Biol.* **17**, e9620 (2021).
27. Ma, J., Yu, J.-T. & Tan, L. MS4A Cluster in Alzheimer's Disease. *Mol. Neurobiol.* **51**, 1240–1248 (2015).
28. Klein, D. *et al.* Mapping cells through time and space with moscot. *bioRxiv* 2023.05.11.540374 (2023) doi:10.1101/2023.05.11.540374.
29. Weiler, P., Lange, M., Klein, M., Pe'er, D. & Theis, F. J. Unified fate mapping in multiview single-cell data. *bioRxiv* 2023.07.19.549685 (2023) doi:10.1101/2023.07.19.549685.
30. Aibar, S. *et al.* SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).
31. Van de Sande, B. *et al.* A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nat. Protoc.* **15**, 2247–2276 (2020).
32. Dimitrov, D. *et al.* Comparison of methods and resources for cell-cell communication inference from single-cell RNA-Seq data. *Nat. Commun.* **13**, 3224 (2022).
33. Srinivasan, K. *et al.* Alzheimer's Patient Microglia Exhibit Enhanced Aging and Unique Transcriptional Activation. *Cell Rep.* **31**, 107843 (2020).
34. Schwabe, T., Srinivasan, K. & Rhinn, H. Shifting paradigms: The central role of microglia in Alzheimer's disease. *Neurobiol. Dis.* **143**, 104962 (2020).
35. Agostini, F. *et al.* The Regulation of MiTF/TFE Transcription Factors Across Model Organisms: from Brain Physiology to Implication for Neurodegeneration. *Mol. Neurobiol.* **59**, 5000–5023

(2022).

36. Miao, J. *et al.* Microglia in Alzheimer's disease: pathogenesis, mechanisms, and therapeutic potentials. *Front. Aging Neurosci.* **15**, 1201982 (2023).
37. Friedman, B. A. *et al.* Diverse Brain Myeloid Expression Profiles Reveal Distinct Microglial Activation States and Aspects of Alzheimer's Disease Not Evident in Mouse Models. *Cell Rep.* **22**, 832–847 (2018).
38. Bennett, D. A. *et al.* Religious Orders Study and Rush Memory and Aging Project. *J. Alzheimers. Dis.* **64**, S161–S189 (2018).
39. Bennett, D. A. *et al.* Neuropathology of older persons without cognitive impairment from two community-based studies. *Neurology* **66**, 1837–1844 (2006).
40. Bennett, D. A. *et al.* Natural history of mild cognitive impairment in older persons. *Neurology* **59**, 198–205 (2002).
41. Bennett, D. A. *et al.* Decision rules guiding the clinical diagnosis of Alzheimer's disease in two community-based cohort studies compared to standard practice in a clinic-based cohort study. *Neuroepidemiology* **27**, 169–176 (2006).
42. Braak, H. & Braak, E. Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathol.* **82**, 239–259 (1991).
43. Braak, H., Alafuzoff, I., Arzberger, T., Kretzschmar, H. & Del Tredici, K. Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry. *Acta Neuropathol.* **112**, 389–404 (2006).
44. Braak, H., Thal, D. R., Ghebremedhin, E. & Del Tredici, K. Stages of the pathologic process in Alzheimer disease: age categories from 1 to 100 years. *J. Neuropathol. Exp. Neurol.* **70**, 960–969 (2011).
45. Fleming, S. J. *et al.* Unsupervised removal of systematic background noise from droplet-based

single-cell experiments using CellBender. *Nat. Methods* **20**, 1323–1335 (2023).

46. Li, B. *et al.* Cumulus provides cloud-based data analysis for large-scale single-cell and single-nucleus RNA-seq. *Nat. Methods* **17**, 793–798 (2020).
47. Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst* **8**, 281–291.e9 (2019).
48. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
49. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233 (2019).
50. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv [stat.ML]* (2018).
51. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
52. Kaminow, B., Yunusov, D. & Dobin, A. STARsolo: accurate, fast and versatile mapping/quantification of single-cell and single-nucleus RNA-seq data. *bioRxiv* 2021.05.05.442755 (2021) doi:10.1101/2021.05.05.442755.
53. Huang, Y., McCarthy, D. J. & Stegle, O. Vireo: Bayesian demultiplexing of pooled single-cell RNA-seq data without genotype reference. *Genome Biol.* **20**, 273 (2019).
54. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
55. Ma, S. *et al.* Molecular and cellular evolution of the primate dorsolateral prefrontal cortex. *Science* **377**, eabo7257 (2022).
56. Bray, N., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal RNA-Seq quantification with kallisto. *Nat. Biotechnol.*
57. Bankhead, P. *et al.* QuPath: Open source software for digital pathology image analysis. *Sci. Rep.*

7, 16878 (2017).

58. Hoffman, G. E. & Roussos, P. Dream: powerful differential expression analysis for repeated measures designs. *Bioinformatics* **37**, 192–201 (2021).
59. Squair, J. W. *et al.* Confronting false discoveries in single-cell differential expression. *Nat. Commun.* **12**, 5692 (2021).
60. Crowell, H. L. *et al.* muscat detects subpopulation-specific state transitions from multi-sample multi-condition single-cell transcriptomics data. *Nat. Commun.* **11**, 6077 (2020).
61. Wu, D. & Smyth, G. K. Camera: a competitive gene set test accounting for inter-gene correlation. *Nucleic Acids Res.* **40**, e133 (2012).
62. Fang, Z., Liu, X. & Peltz, G. GSEApy: a comprehensive package for performing gene set enrichment analysis in Python. *Bioinformatics* **39**, (2023).
63. Reuter, B., Fackeldey, K. & Weber, M. Generalized Markov modeling of nonreversible molecular kinetics. *J. Chem. Phys.* **150**, 174103 (2019).
64. Haghverdi, L., Büttner, M., Wolf, F. A., Buettner, F. & Theis, F. J. Diffusion pseudotime robustly reconstructs lineage branching. *Nat. Methods* **13**, 845–848 (2016).
65. Heumos, L. *et al.* Best practices for single-cell analysis across modalities. *Nat. Rev. Genet.* 1–23 (2023).
66. Kolde, R., Laur, S., Adler, P. & Vilo, J. Robust rank aggregation for gene list integration and meta-analysis. *Bioinformatics* **28**, 573–580 (2012).
67. Efremova, M., Vento-Tormo, M., Teichmann, S. A. & Vento-Tormo, R. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat. Protoc.* **15**, 1484–1506 (2020).
68. Raredon, M. S. B. *et al.* Computation and visualization of cell-cell signaling topologies in single-cell systems data using Connectome. *Sci. Rep.* **12**, 4187 (2022).

69. Hou, R., Denisenko, E., Ong, H. T., Ramilowski, J. A. & Forrest, A. R. R. Predicting cell-to-cell communication networks using NATMI. *Nat. Commun.* **11**, 5011 (2020).

70. Cabello-Aguilar, S. *et al.* SingleCellSignalR: inference of intercellular networks from single-cell transcriptomics. *Nucleic Acids Res.* **48**, e55 (2020).

71. Jin, S. *et al.* Inference and analysis of cell-cell communication using CellChat. *Nat. Commun.* **12**, 1088 (2021).

72. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).

73. Armingol, E. *et al.* Context-aware deconvolution of cell–cell communication with Tensor-cell2cell. *Nat. Commun.* **13**, 1–15 (2022).

74. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).

75. Gene Ontology Consortium *et al.* The Gene Ontology knowledgebase in 2023. *Genetics* **224**, (2023).

76. International Multiple Sclerosis Genetics Consortium. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science* **365**, (2019).

77. Nalls, M. A. *et al.* Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet Neurol.* **18**, 1091–1102 (2019).

78. Als, T. D. *et al.* Depression pathophysiology, risk prediction of recurrence and comorbid psychiatric disorders using genome-wide analyses. *Nat. Med.* **29**, 1832–1844 (2023).

79. Grove, J. *et al.* Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* **51**, 431–444 (2019).

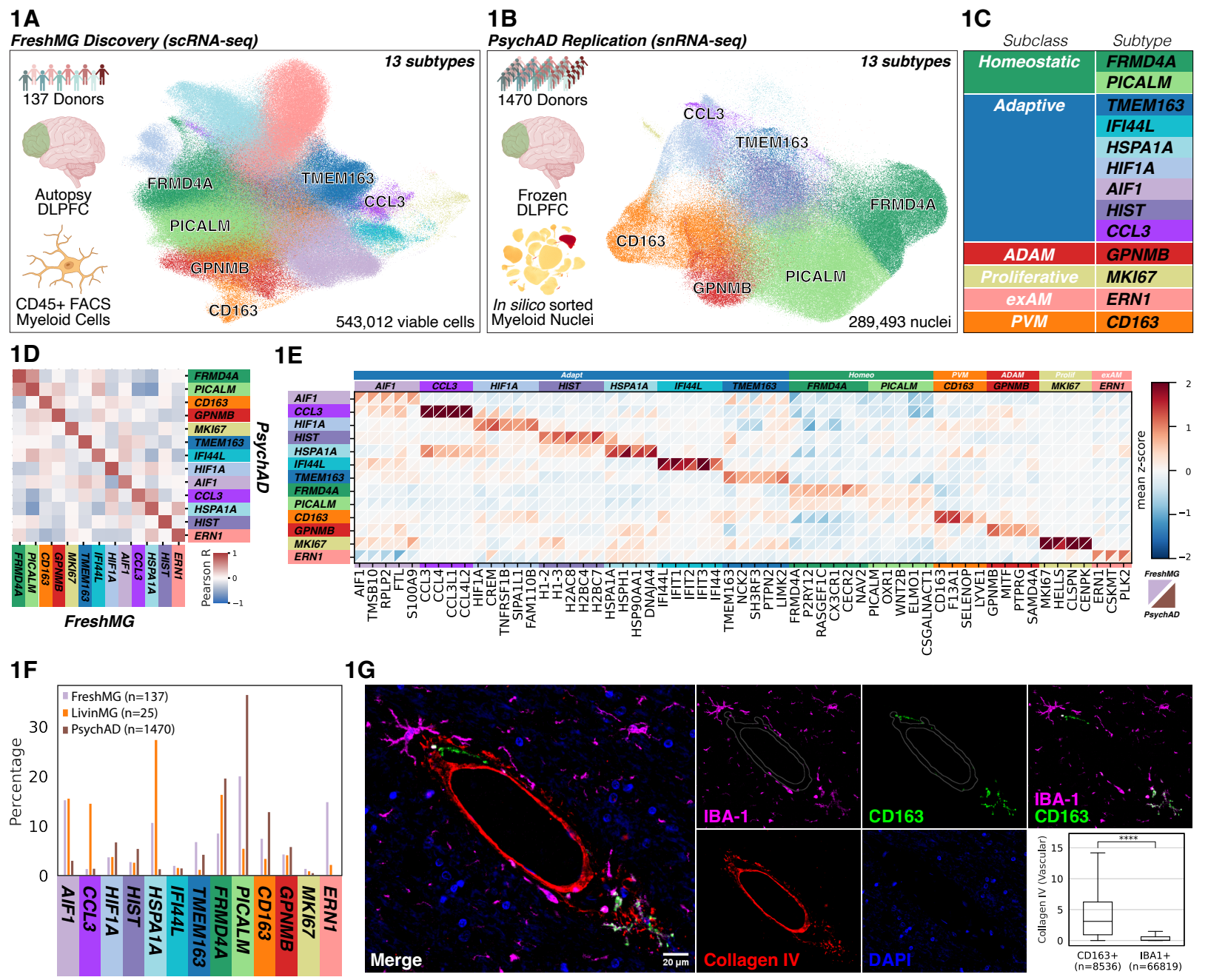
80. Mullins, N. *et al.* Genome-wide association study of more than 40,000 bipolar disorder cases

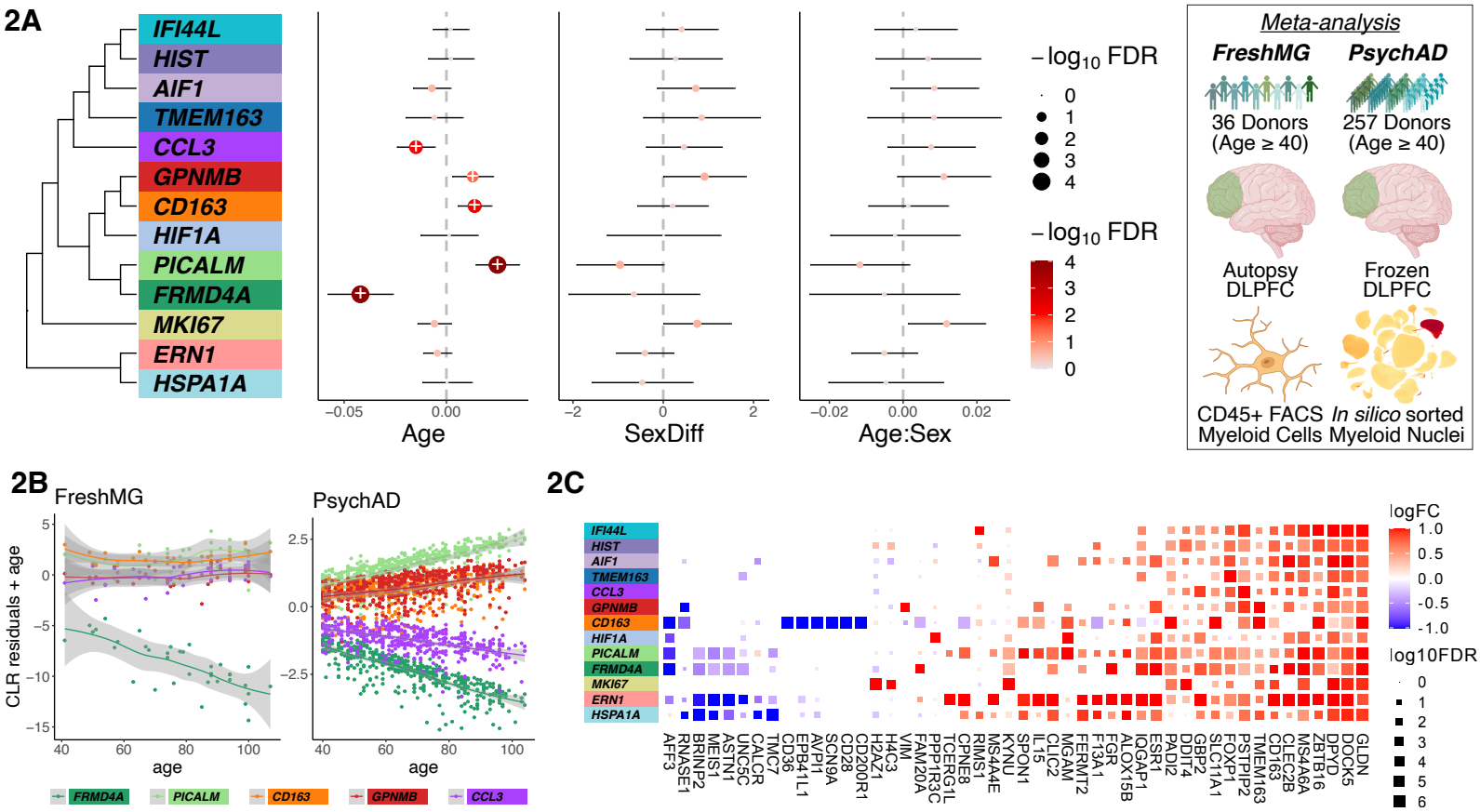
provides new insights into the underlying biology. *Nat. Genet.* **53**, 817–829 (2021).

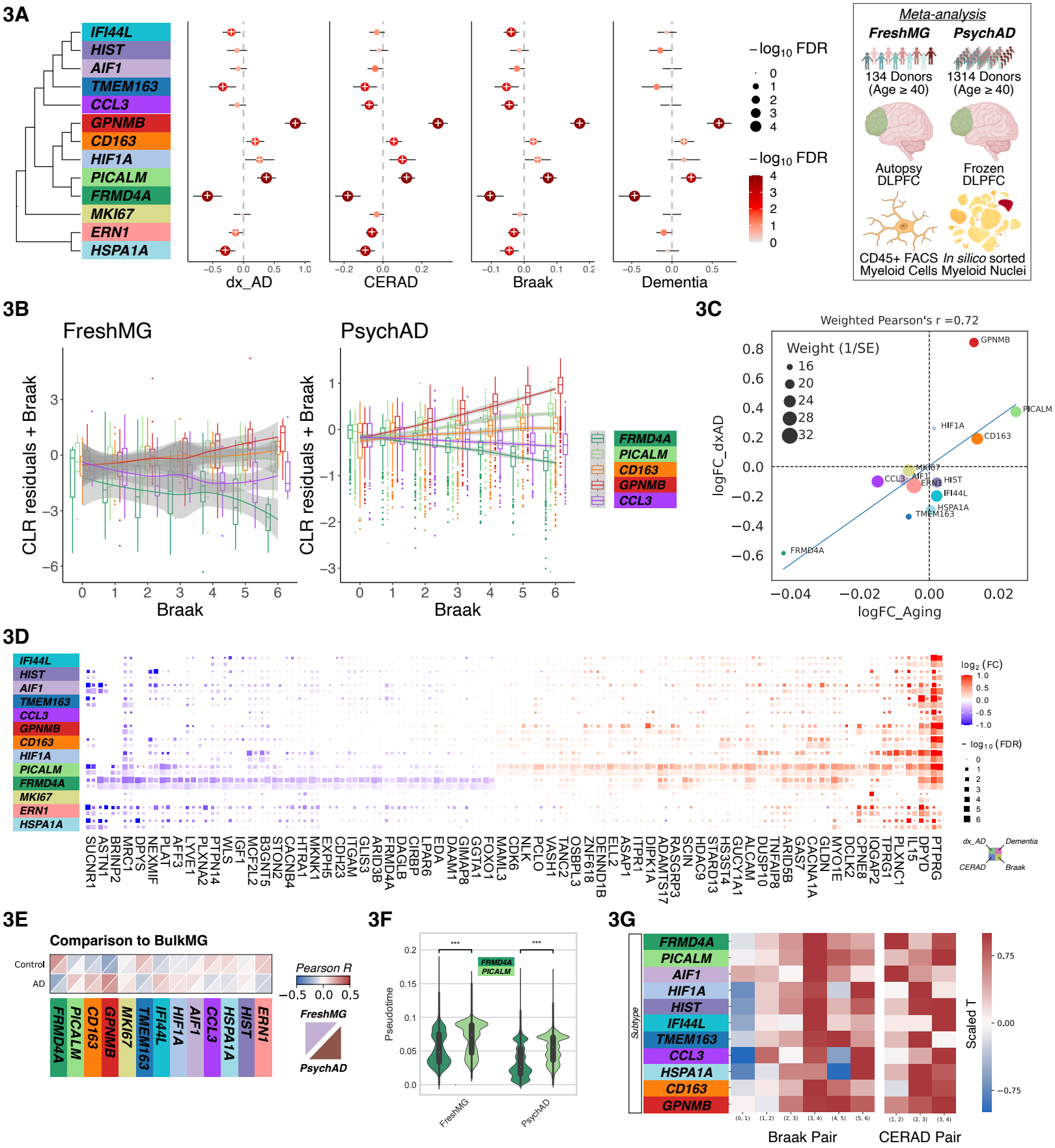
81. Trubetskoy, V. *et al.* Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* **604**, 502–508 (2022).
82. van Rheenen, W. *et al.* Common and rare variant association analyses in amyotrophic lateral sclerosis identify 15 risk loci with distinct genetic architectures and neuron-specific biology. *Nat. Genet.* **53**, 1636–1648 (2021).
83. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).
84. 1000 Genomes Project Consortium *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
85. 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
86. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, (2021).
87. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
88. Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C. A. & Smoller, J. W. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat. Commun.* **10**, 1776 (2019).
89. Bohlen, C. J. *et al.* Diverse Requirements for Microglial Survival, Specification, and Function Revealed by Defined-Medium Cultures. *Neuron* **94**, 759–773.e8 (2017).
90. Gosselin, D. *et al.* An environment-dependent transcriptional network specifies human microglia identity. *Science* **356**, (2017).
91. Mattei, D. *et al.* Enzymatic Dissociation Induces Transcriptional and Proteotype Bias in Brain Cell Populations. *Int. J. Mol. Sci.* **21**, (2020).
92. Ocañas, S. R. *et al.* Minimizing the Ex Vivo Confounds of Cell-Isolation Techniques on

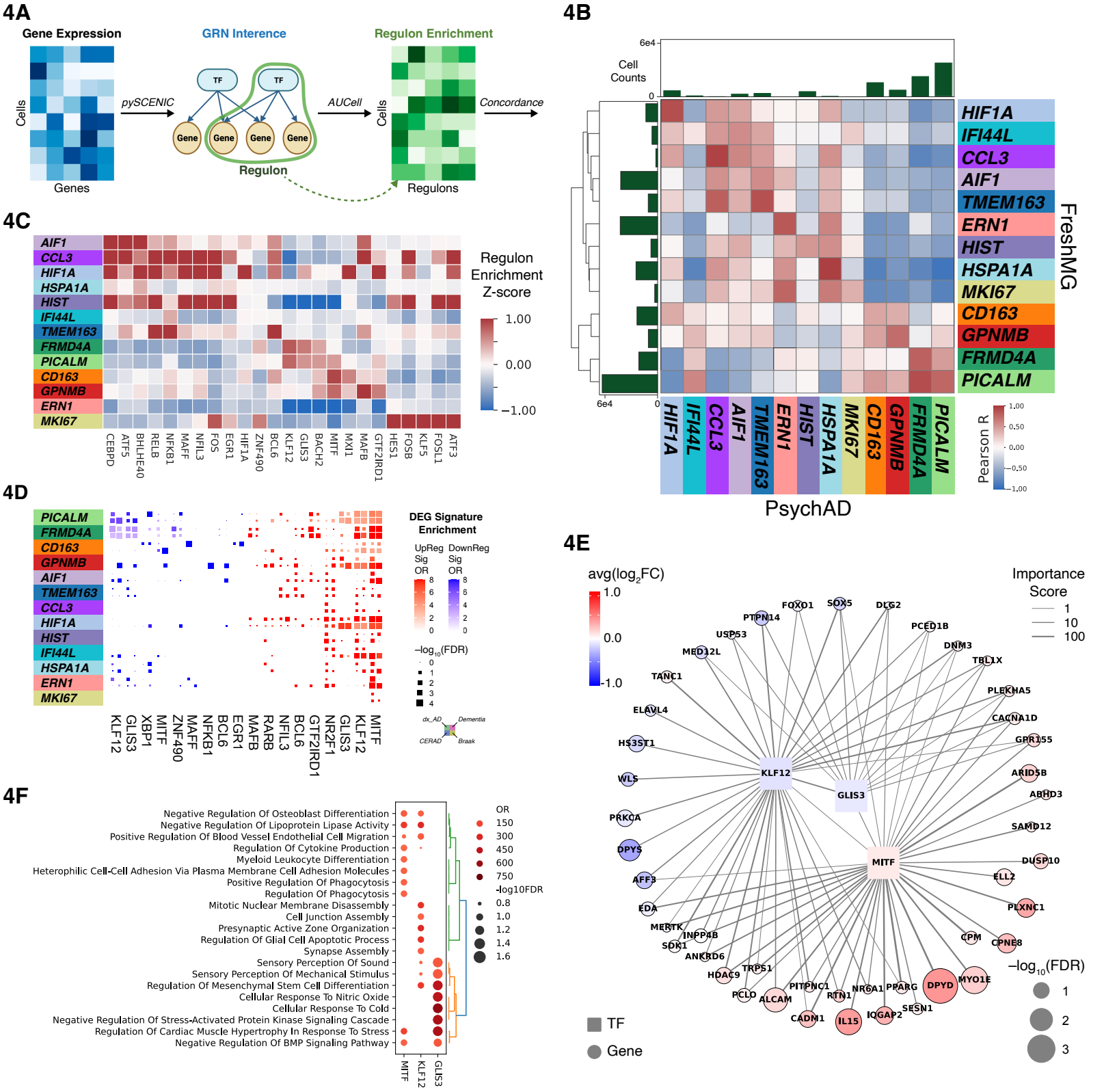
Transcriptomic and Translatomic Profiles of Purified Microglia. *eNeuro* **9**, (2022).

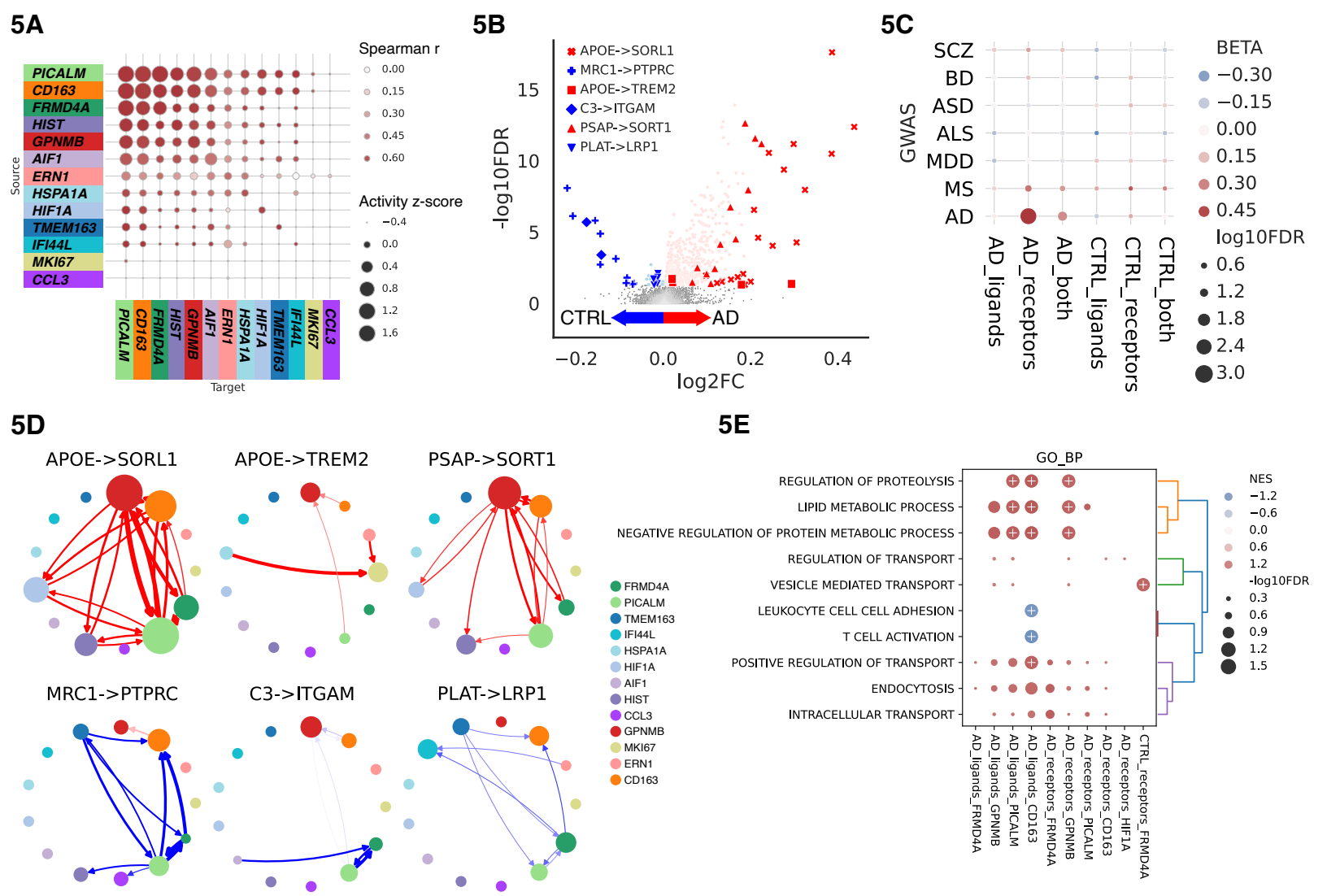
93. van den Brink, S. C. *et al.* Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. *Nat. Methods* **14**, 935–936 (2017).
94. Lee, Y. *et al.* Polo-like kinase 2 phosphorylation of amyloid precursor protein regulates activity-dependent amyloidogenic processing. *Neuropharmacology* **117**, 387–400 (2017).
95. Lee, J. S. *et al.* Inhibition of Polo-like kinase 2 ameliorates pathogenesis in Alzheimer's disease model mice. *PLoS One* **14**, e0219691 (2019).
96. Setty, M. *et al.* Characterization of cell fate probabilities in single-cell data with Palantir. *Nat. Biotechnol.* **37**, 451–460 (2019).
97. Street, K. *et al.* Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477 (2018).
98. Schiebinger, G. *et al.* Optimal-Transport Analysis of Single-Cell Gene Expression Identifies Developmental Trajectories in Reprogramming. *Cell* **176**, 928–943.e22 (2019).

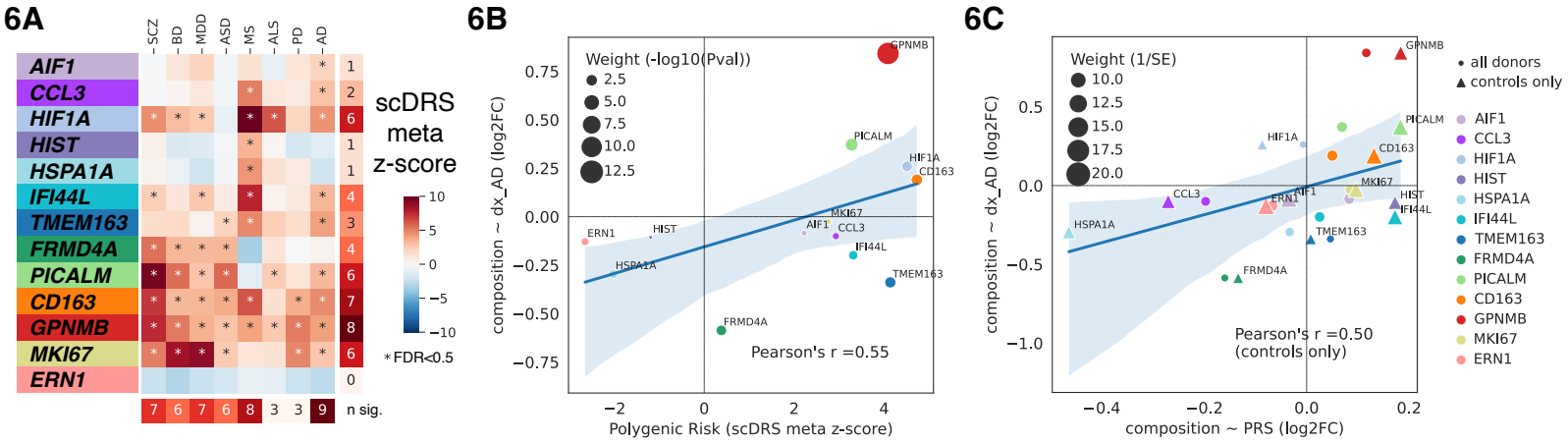

1A
FreshMG Discovery (scRNA-seq)
137 Donors
Autopsy DLPFC
CD45+ FACS Myeloid Cells
13 subtypes
FRMD4A
TMEM163
CCL3
PICALM
GPNMB
CD163
543,012 viable cells
1B
PsychAD Replication (snRNA-seq)
1470 Donors
Frozen DLPFC
In silico sorted Myeloid Nuclei
13 subtypes
CCL3
TMEM163
FRMD4A
CD163
PICALM
GPNMB
289,493 nuclei
1C
Subclass
Subtype
Homeostatic
FRMD4A
PICALM
Adaptive
TMEM163
IFI44L
HSPA1A
HIF1A
AIF1
HIST
CCL3
ADAM
GPNMB
Proliferative
MKI67
exAM
ERN1
PVM
CD163
1D
PsychAD
FreshMG
Pearson R
1E
mean z-score
FreshMG
PsychAD
1F
FreshMG (n=137)
LivinMG (n=25)
PsychAD (n=1470)
Percentage
1G
Merge
20 µm
IBA-1
CD163
IBA-1
CD163
Collagen IV
DAPI
Collagen IV (Vascular)
****
CD163+ (n=8536)
IBA1+ (n=66819)


2A
IFI44L
HIST
AIF1
TMEM163
CCL3
GPNMB
CD163
HIF1A
PICALM
FRMD4A
MKI67
ERN1
HSPA1A
Age
SexDiff
Age:Sex
−log10 FDR
Meta-analysis
FreshMG
PsychAD
36 Donors (Age ≥ 40)
257 Donors (Age ≥ 40)
Autopsy DLPFC
Frozen DLPFC
CD45+ FACS Myeloid Cells
In silico sorted Myeloid Nuclei
2B
FreshMG
PsychAD
CLR residuals + age
age
FRMD4A
PICALM
CD163
GPNMB
CCL3
2C
logFC
log10FDR
AFF3
RNASE1
BRINP2
MEIS1
ASTN1
UNC5C
CALCR
TMC7
CD36
EPB41L1
AVPI1
SCN9A
CD28
CD200R1
H2AZ1
H4C3
VIM
FAM20A
PPP1R3C
TCERG1L
CPNE8
RIMS1
MS4A4E
SPON1
KYNU
IL15
CLIC2
MGAM
FERMT2
F13A1
FGR
ALOX15B
IQGAP1
ESR1
PADI2
DDIT4
GBP2
SLC11A1
FOXP1
PSTPIP2
TMEM163
CD163
CLEC2B
MS4A6A
ZBTB16
DPYD
DOCK5
GLDN

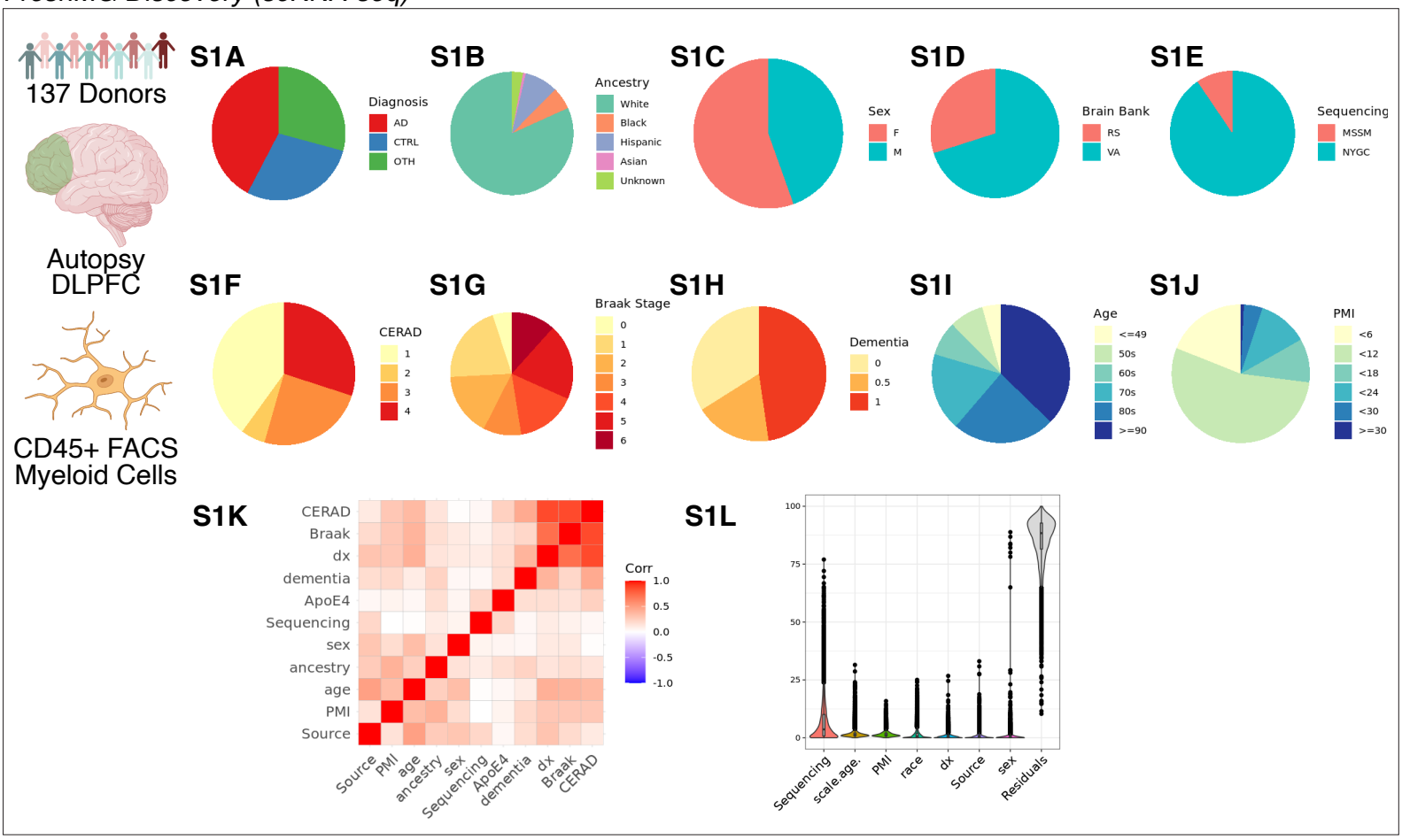

3A
IFI44L
HIST
AIF1
TMEM163
CCL3
GPNMB
CD163
HIF1A
PICALM
FRMD4A
MKI67
ERN1
HSPA1A
dx_AD
CERAD
Braak
Dementia
−log10 FDR
Meta-analysis
FreshMG
PsychAD
134 Donors (Age ≥ 40)
1314 Donors (Age ≥ 40)
Autopsy DLPFC
Frozen DLPFC
CD45+ FACS Myeloid Cells
In silico sorted Myeloid Nuclei
3B
FreshMG
PsychAD
CLR residuals + Braak
Braak
3C
Weighted Pearson's r =0.72
Weight (1/SE)
logFC_dxAD
logFC_Aging
3D
log2 (FC)
− log10 (FDR)
dx_AD
Dementia
CERAD
Braak
3E
Comparison to BulkMG
Control
AD
Pearson R
FreshMG
PsychAD
3F
Pseudotime
FRMD4A
PICALM
FreshMG
PsychAD
3G
Subtype
Braak Pair
CERAD Pair
Scaled T


4A
Gene Expression
GRN Interence
Regulon Enrichment
Cells
Genes
pySCENIC
TF
Gene
Regulon
AUCell
Regulons
Concordance
4B
Cell Counts
6e4
0
HIF1A
IFI44L
CCL3
AIF1
TMEM163
ERN1
HIST
HSPA1A
MKI67
CD163
GPNMB
FRMD4A
PICALM
FreshMG
PsychAD
Pearson R
1.00
0.50
0.00
−0.50
−1.00
4C
Regulon Enrichment Z-score
1.00
0.00
−1.00
CEBPD
ATF5
BHLHE40
RELB
NFKB1
MAFF
NFIL3
FOS
EGR1
HIF1A
ZNF490
BCL6
KLF12
GLIS3
BACH2
MITF
MXI1
MAFB
GTF2IRD1
HES1
FOSB
KLF5
FOSL1
ATF3
4D
DEG Signature Enrichment
UpReg Sig OR
DownReg Sig OR
−log10(FDR)
dx_AD
Dementia
CERAD
Braak
KLF12
GLIS3
XBP1
MITF
ZNF490
MAFF
NFKB1
BCL6
EGR1
MAFB
RARB
NFIL3
BCL6
GTF2IRD1
NR2F1
GLIS3
KLF12
MITF
4E
avg(log2FC)
1.0
0.0
-1.0
Importance Score
1
10
100
FOXD1
SOX5
DLG2
PTPN14
PCED1B
USP53
DNM3
MED12L
TBL1X
TANC1
PLEKHA5
ELAVL4
CACNA1D
GPR155
HS3ST1
KLF12
GLIS3
ARID5B
WLS
ABHD3
PRKCA
SAMD12
DPYS
MITF
DUSP10
ELL2
AFF3
PLXNC1
EDA
MERTK
INPP4B
CPM
CPNE8
SDK1
ANKRD6
HDAC9
TRPS1
MYO1E
PITPNC1
NR6A1
PPARG
DPYD
PCLO
ALCAM
RTN1
SESN1
CADM1
IL15
IQGAP2
TF
Gene
−log10(FDR)
1
2
3
4F
Negative Regulation Of Osteoblast Differentiation
Negative Regulation Of Lipoprotein Lipase Activity
Positive Regulation Of Blood Vessel Endothelial Cell Migration
Regulation Of Cytokine Production
Myeloid Leukocyte Differentiation
Heterophilic Cell-Cell Adhesion Via Plasma Membrane Cell Adhesion Molecules
Positive Regulation Of Phagocytosis
Regulation Of Phagocytosis
Mitotic Nuclear Membrane Disassembly
Cell Junction Assembly
Presynaptic Active Zone Organization
Regulation Of Glial Cell Apoptotic Process
Synapse Assembly
Sensory Perception Of Sound
Sensory Perception Of Mechanical Stimulus
Regulation Of Mesenchymal Stem Cell Differentiation
Cellular Response To Nitric Oxide
Cellular Response To Cold
Negative Regulation Of Stress-Activated Protein Kinase Signaling Cascade
Regulation Of Cardiac Muscle Hypertrophy In Response To Stress
Negative Regulation Of BMP Signaling Pathway
MITF
KLF12
GLIS3
OR
150
300
450
600
750
-log10FDR
0.8
1.0
1.2
1.4
1.6

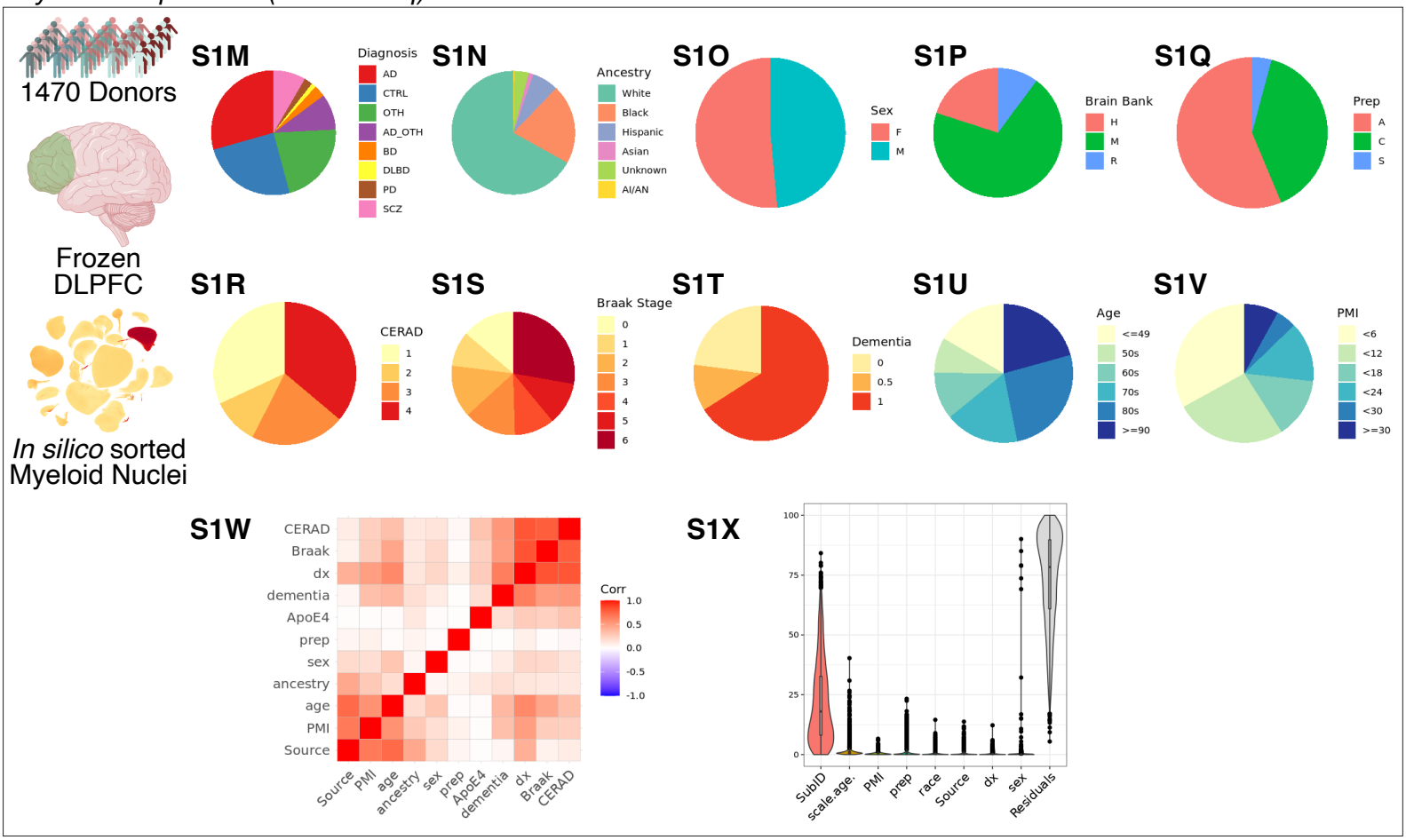

5A
Spearman r
0.00
0.15
0.30
0.45
0.60
Activity z-score
−0.4
0.0
0.4
0.8
1.2
1.6
Source
Target
PICALM
CD163
FRMD4A
HIST
GPNMB
AIF1
ERN1
HSPA1A
HIF1A
TMEM163
IFI44L
MKI67
CCL3
5B
APOE->SORL1
MRC1->PTPRC
APOE->TREM2
C3->ITGAM
PSAP->SORT1
PLAT->LRP1
-log10FDR
log2FC
CTRL
AD
5C
SCZ
BD
ASD
ALS
MDD
MS
AD
GWAS
AD_ligands
AD_receptors
AD_both
CTRL_ligands
CTRL_receptors
CTRL_both
BETA
−0.30
−0.15
0.00
0.15
0.30
0.45
log10FDR
0.6
1.2
1.8
2.4
3.0
5D
APOE->SORL1
APOE->TREM2
PSAP->SORT1
MRC1->PTPRC
C3->ITGAM
PLAT->LRP1
FRMD4A
PICALM
TMEM163
IFI44L
HSPA1A
HIF1A
AIF1
HIST
CCL3
GPNMB
MKI67
ERN1
CD163
5E
GO_BP
REGULATION OF PROTEOLYSIS
LIPID METABOLIC PROCESS
NEGATIVE REGULATION OF PROTEIN METABOLIC PROCESS
REGULATION OF TRANSPORT
VESICLE MEDIATED TRANSPORT
LEUKOCYTE CELL CELL ADHESION
T CELL ACTIVATION
POSITIVE REGULATION OF TRANSPORT
ENDOCYTOSIS
INTRACELLULAR TRANSPORT
AD_ligands_FRMD4A
AD_ligands_GPNMB
AD_ligands_PICALM
AD_ligands_CD163
AD_receptors_FRMD4A
AD_receptors_GPNMB
AD_receptors_PICALM
AD_receptors_CD163
AD_receptors_HIF1A
CTRL_receptors_FRMD4A
NES
−1.2
−0.6
0.0
0.6
1.2
-log10FDR
0.3
0.6
0.9
1.2
1.5

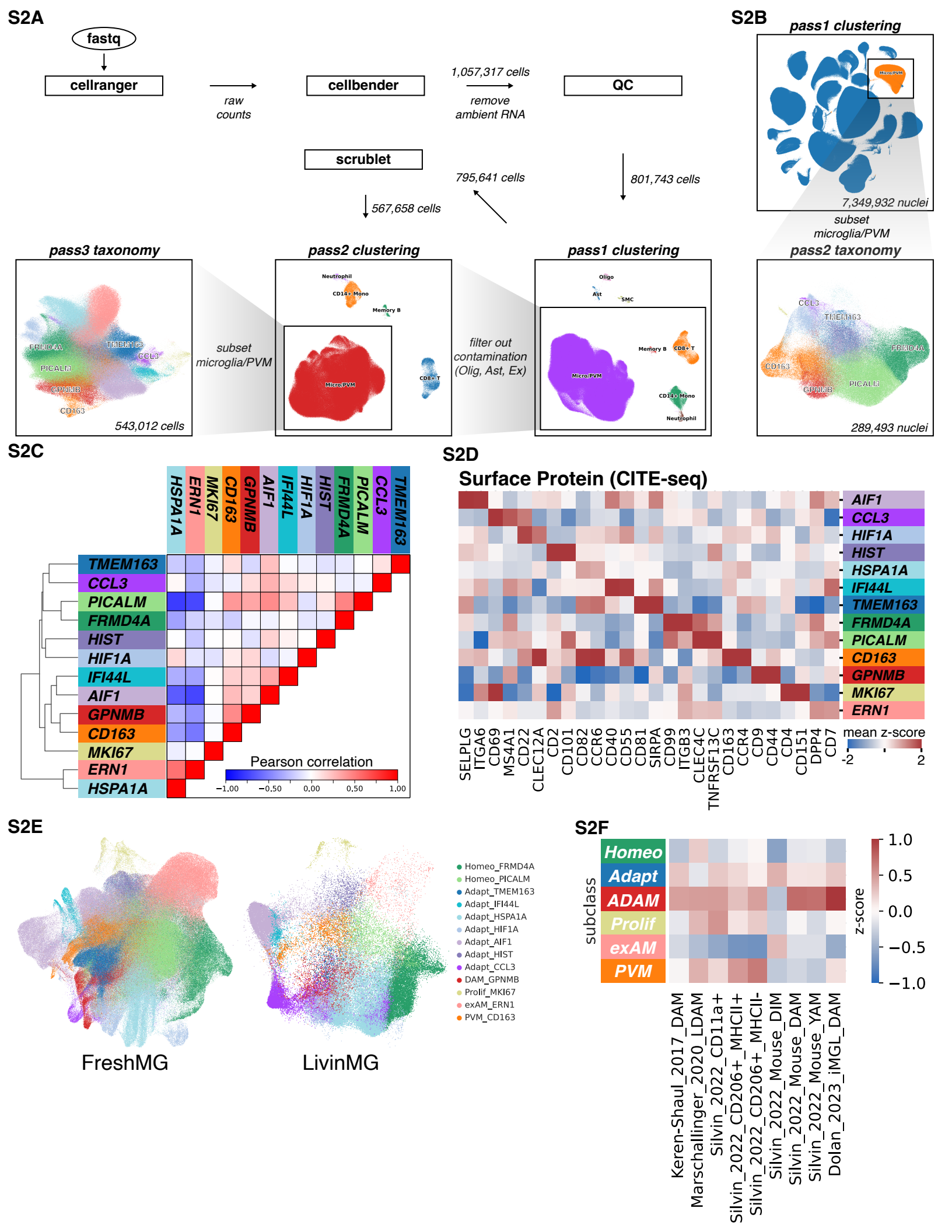

6A
SCZ
BD
MDD
ASD
MS
ALS
PD
AD
AIF1
CCL3
HIF1A
HIST
HSPA1A
IFI44L
TMEM163
FRMD4A
PICALM
CD163
GPNMB
MKI67
ERN1
scDRS meta z-score
*FDR<0.5
n sig.
6B
Weight (-log10(Pval))
2.5
5.0
7.5
10.0
12.5
Pearson's r =0.55
composition ~ dx_AD (log2FC)
Polygenic Risk (scDRS meta z-score)
6C
Weight (1/SE)
10.0
12.5
15.0
17.5
20.0
Pearson's r =0.50
(controls only)
composition ~ dx_AD (log2FC)
composition ~ PRS (log2FC)
all donors
controls only

FreshMG Discovery (scRNA-seq)

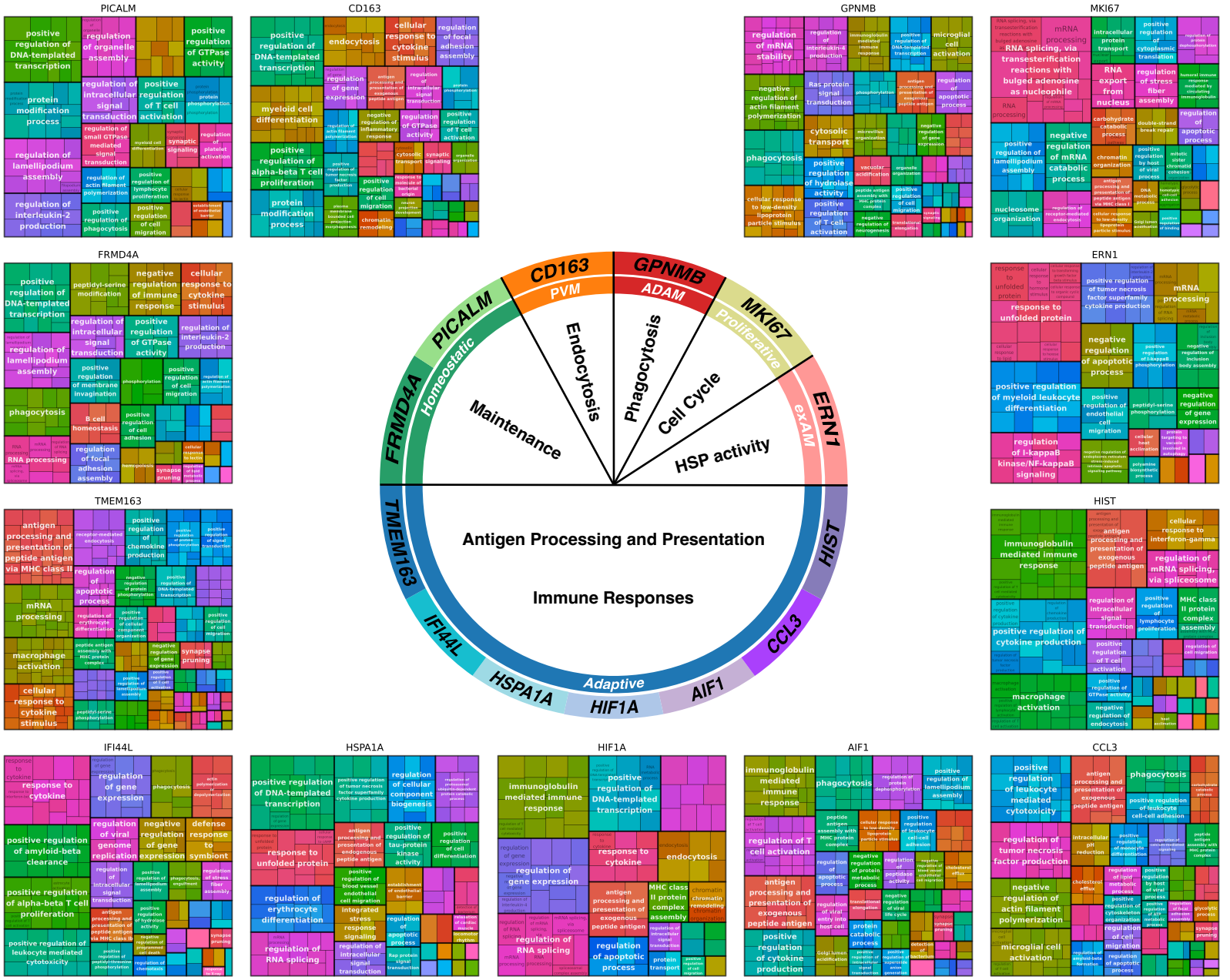


PsychAD Replication (snRNA-seq)



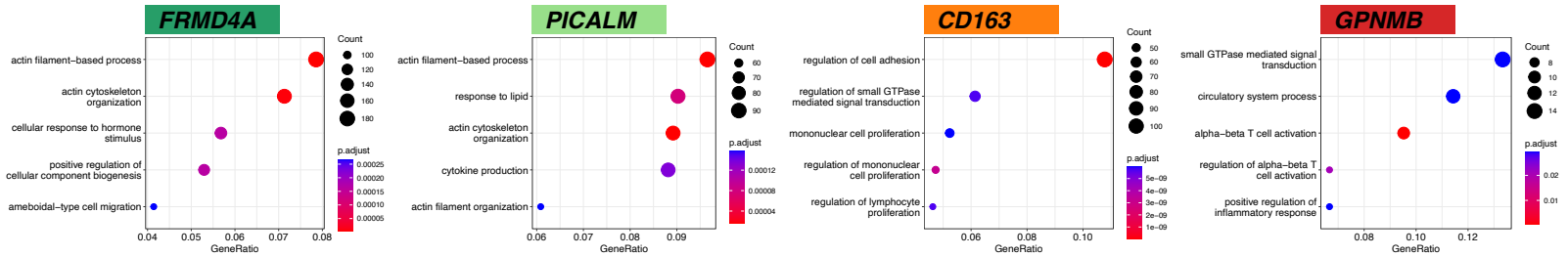

S2A
fastq
cellranger
raw counts
cellbender
1,057,317 cells
remove ambient RNA
QC
801,743 cells
scrublet
567,658 cells
795,641 cells
S2B
pass1 clustering
Micro/PVM
7,349,932 nuclei
subset microglia/PVM
pass3 taxonomy
FRMD4A
TMEM163
CCL3
PICALM
GPNMB
CD163
543,012 cells
subset microglia/PVM
pass2 clustering
Neutrophil
CD14+ Mono
Memory B
Micro/PVM
CD8+ T
filter out contamination (Olig, Ast, Ex)
pass1 clustering
Oligo
Ast
SMC
Memory B
CD8+ T
Micro/PVM
CD14+ Mono
Neutrophil
pass2 taxonomy
CCL3
TMEM163
FRMD4A
CD163
GPNMB
PICALM
289,493 nuclei
S2C
HSPA1A
ERN1
MKI67
CD163
GPNMB
AIF1
IFI44L
HIF1A
HIST
FRMD4A
PICALM
CCL3
TMEM163
Pearson correlation
−1.00
−0.50
0.00
0.50
1.00
S2D
Surface Protein (CITE-seq)
AIF1
CCL3
HIF1A
HIST
HSPA1A
IFI44L
TMEM163
FRMD4A
PICALM
CD163
GPNMB
MKI67
ERN1
SELPLG
ITGA6
CD69
MS4A1
CD22
CLEC12A
CD2
CD101
CD82
CCR6
CD40
CD55
CD81
SIRPA
CD99
ITGB3
CLEC4C
TNFRSF13C
CD163
CCR4
CD9
CD44
CD4
CD151
DPP4
CD7
mean z-score
-2
2
S2E
Homeo_FRMD4A
Homeo_PICALM
Adapt_TMEM163
Adapt_IFI44L
Adapt_HSPA1A
Adapt_HIF1A
Adapt_AIF1
Adapt_HIST
Adapt_CCL3
DAM_GPNMB
Prolif_MKI67
exAM_ERN1
PVM_CD163
FreshMG
LivinMG
S2F
subclass
Homeo
Adapt
ADAM
Prolif
exAM
PVM
z-score
1.0
0.5
0.0
−0.5
−1.0
Keren-Shaul_2017_DAM
Marschallinger_2020_LDAM
Silvin_2022_CD11a+
Silvin_2022_CD206+_MHCII+
Silvin_2022_CD206+_MHCII-
Silvin_2022_Mouse_DIM
Silvin_2022_Mouse_DAM
Silvin_2022_Mouse_YAM
Dolan_2023_iMGL_DAM

S3A



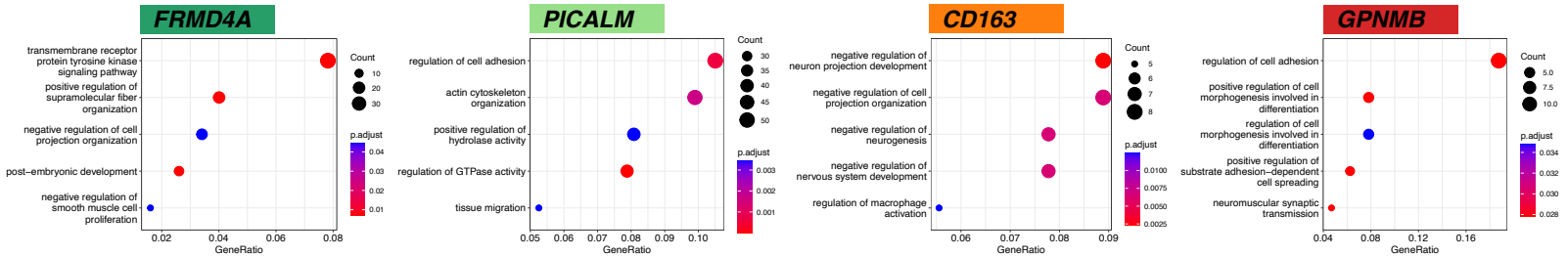
S3B

Aging

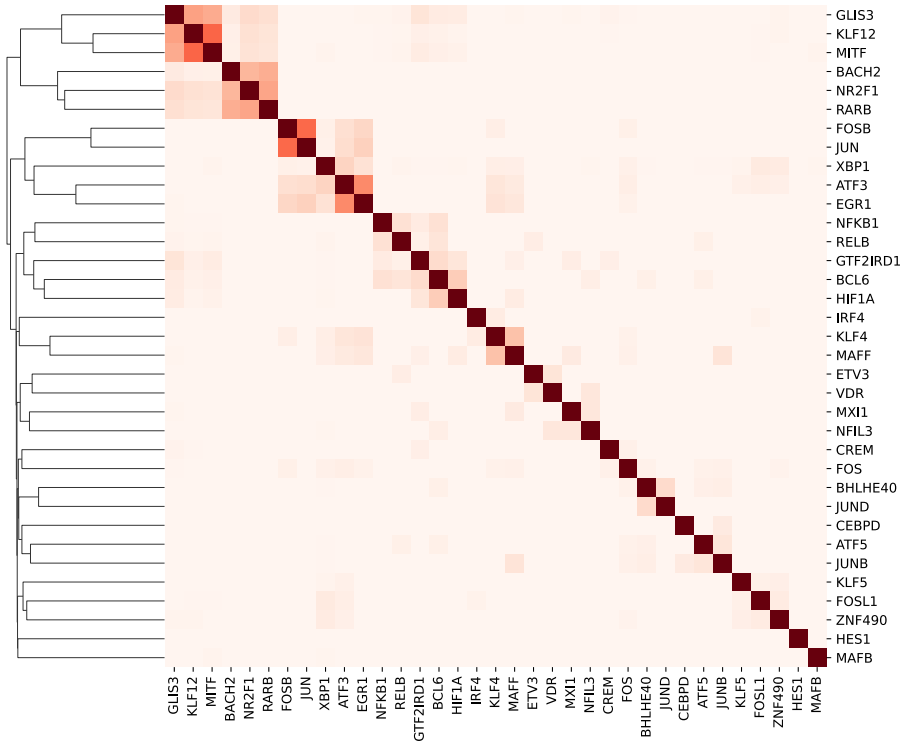


S3C

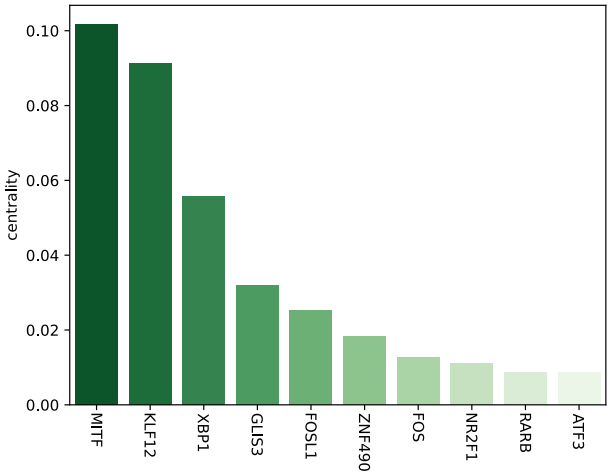
AD phenotypes



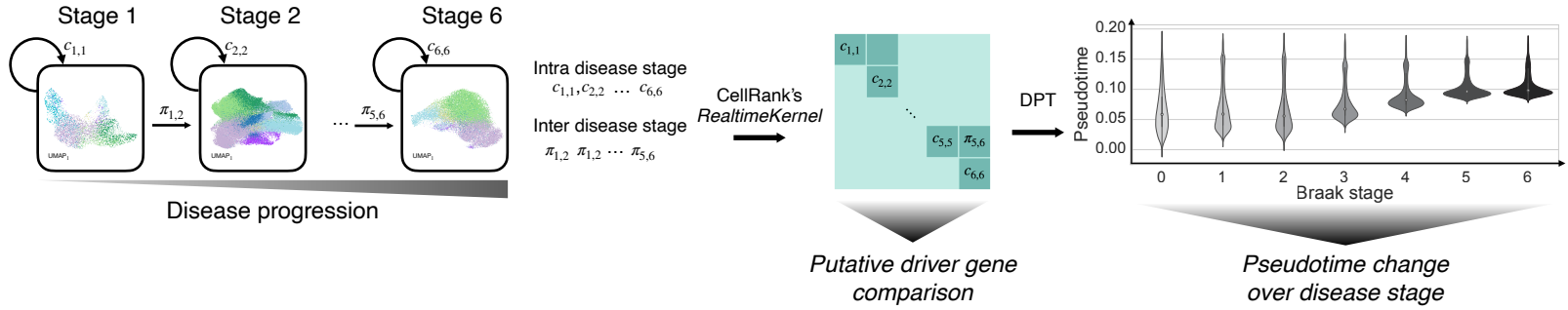
S4A


GLIS3
KLF12
MITF
BACH2
NR2F1
RARB
FOSB
JUN
XBP1
ATF3
EGR1
NFKB1
RELB
GTF2IRD1
BCL6
HIF1A
IRF4
KLF4
MAFF
ETV3
VDR
MXI1
NFIL3
CREM
FOS
BHLHE40
JUND
CEBPD
ATF5
JUNB
KLF5
FOSL1
ZNF490
HES1
MAFB

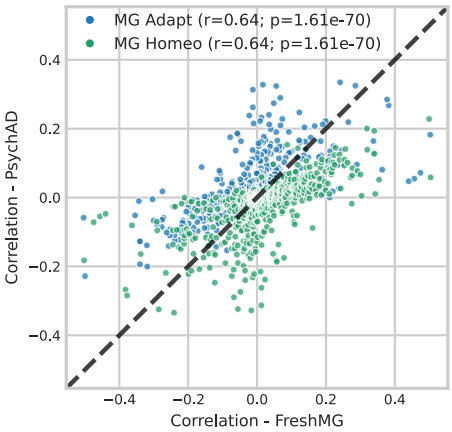
S4B


centrality
0.00
0.02
0.04
0.06
0.08
0.10
MITF
KLF12
XBP1
GLIS3
FOSL1
ZNF490
FOS
NR2F1
RARB
ATF3

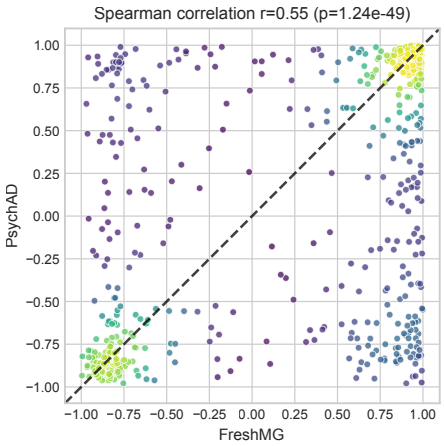
S5A



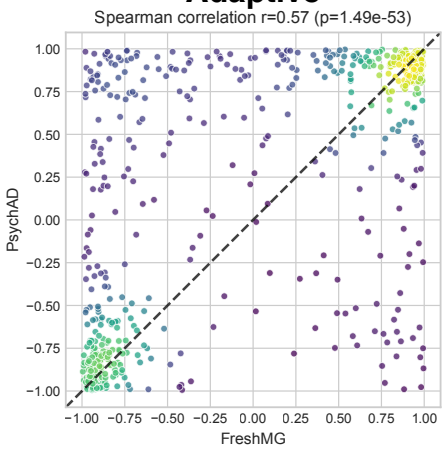
S5B



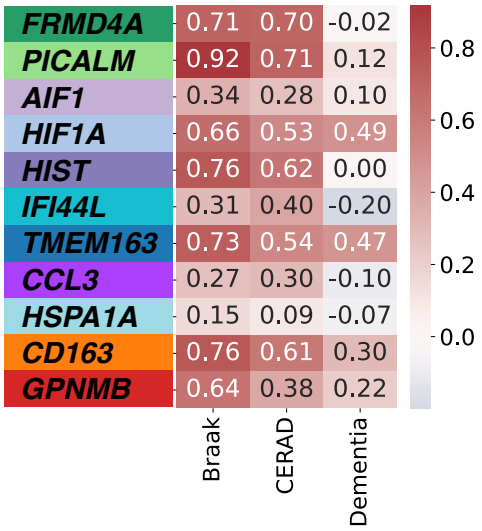
S5C

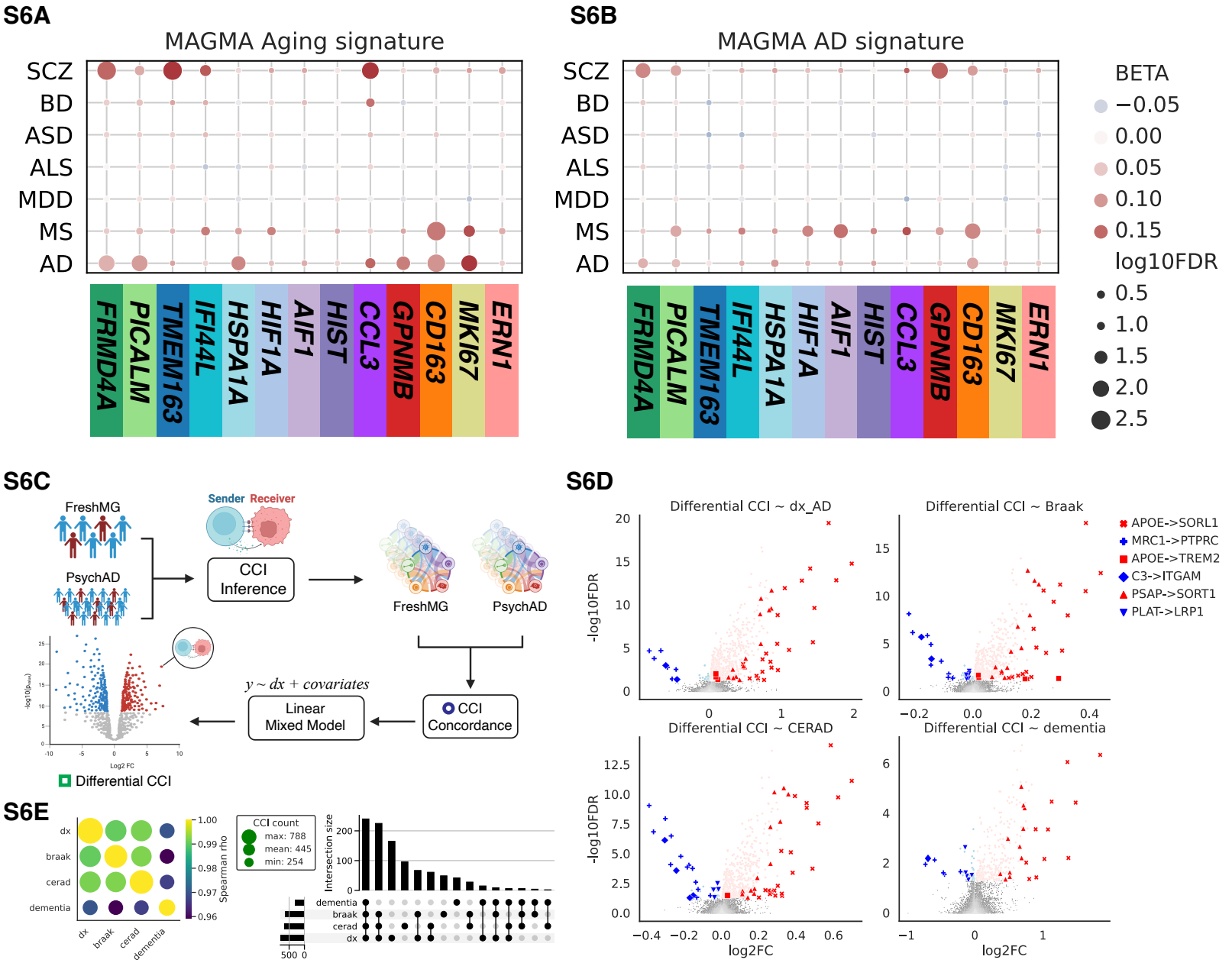


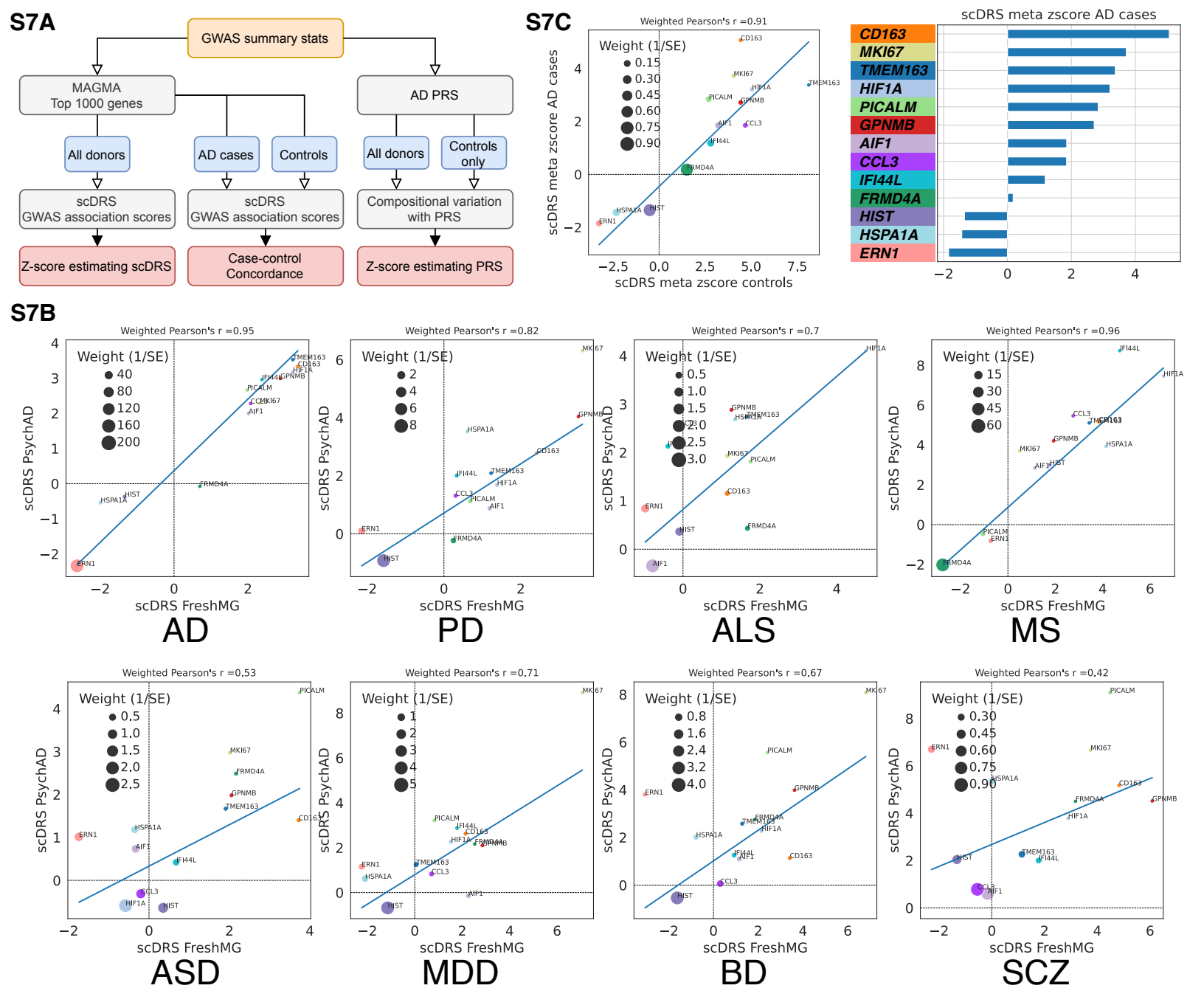
S5D



S5E




S6A
MAGMA Aging signature
SCZ
BD
ASD
ALS
MDD
MS
AD
FRMD4A
PICALM
TMEM163
IFI44L
HSPA1A
HIF1A
AIF1
HIST
CCL3
GPNMB
CD163
MKI67
ERN1
S6B
MAGMA AD signature
BETA
−0.05
0.00
0.05
0.10
0.15
log10FDR
0.5
1.0
1.5
2.0
2.5
S6C
FreshMG
PsychAD
Sender
Receiver
CCI Inference
FreshMG
PsychAD
CCI Concordance
y ~ dx + covariates
Linear Mixed Model
Log2 FC
Differential CCI
S6D
Differential CCI ~ dx_AD
Differential CCI ~ Braak
Differential CCI ~ CERAD
Differential CCI ~ dementia
-log10FDR
log2FC
APOE->SORL1
MRC1->PTPRC
APOE->TREM2
C3->ITGAM
PSAP->SORT1
PLAT->LRP1
S6E
dx
braak
cerad
dementia
Spearman rho
CCI count
max: 788
mean: 445
min: 254
Intersection size


S7A
GWAS summary stats
MAGMA
Top 1000 genes
AD PRS
All donors
AD cases
Controls
All donors
Controls only
scDRS
GWAS association scores
scDRS
GWAS association scores
Compositional variation
with PRS
Z-score estimating scDRS
Case-control
Concordance
Z-score estimating PRS
S7C
Weighted Pearson's r =0.91
Weight (1/SE)
0.15
0.30
0.45
0.60
0.75
0.90
scDRS meta zscore AD cases
scDRS meta zscore controls
scDRS meta zscore AD cases
CD163
MKI67
TMEM163
HIF1A
PICALM
GPNMB
AIF1
CCL3
IFI44L
FRMD4A
HIST
HSPA1A
ERN1
S7B
Weighted Pearson's r =0.95
Weight (1/SE)
40
80
120
160
200
scDRS PsychAD
scDRS FreshMG
AD
Weighted Pearson's r =0.82
Weight (1/SE)
2
4
6
8
PD
Weighted Pearson's r =0.7
Weight (1/SE)
0.5
1.0
1.5
2.0
2.5
3.0
ALS
Weighted Pearson's r =0.96
Weight (1/SE)
15
30
45
60
MS
Weighted Pearson's r =0.53
Weight (1/SE)
0.5
1.0
1.5
2.0
2.5
ASD
Weighted Pearson's r =0.71
Weight (1/SE)
1
2
3
4
5
MDD
Weighted Pearson's r =0.67
Weight (1/SE)
0.8
1.6
2.4
3.2
4.0
BD
Weighted Pearson's r =0.42
Weight (1/SE)
0.30
0.45
0.60
0.75
0.90
SCZ
S7D
TMEM163
AIF1
HIST
IFI44L
GPNMB
CD163
PICALM
FRMD4A
CCL3
MKI67
HSPA1A
HIF1A
ERN1
ALL DONORS
CONTROLS
−log10 adj.P.Val
Weighted Pearson's r =0.62
Weight (1/SE)
0.048
0.054
0.060
0.066
0.072
CRUMBLR PRS logFC CONTROLS
CRUMBLR PRS logFC ALL DONORS