Human peripheral monocytes capture elements of the state of microglial activation in the brain

Daniel Felsky

Centre for Addiction and Mental Health   https://orcid.org/0000-0003-1831-9848

Hans-Ulrich Klein

Department of Biochemistry, Emory University School of Medicine, Atlanta, GA 30322
https://orcid.org/0000-0002-6382-9428

Vilas Menon

Columbia University Medical Center   https://orcid.org/0000-0002-4096-8601

Yiyi Ma

Columbia University Irving Medical Center   https://orcid.org/0000-0002-3609-8877

Yanling Wang

Rush University Medicine Center

Milos Milic

CAMH

Peter Zhukovsky

Centre for Addiction and Mental Health

Julie Schneider

David Bennett

Rush University Medical Center

Philip De Jager (✉ pld2115@cumc.columbia.edu)

Columbia University Medical Center   https://orcid.org/0000-0002-8057-2505

Article

Abstract

Despite a growing focus on neuroimmune mechanisms of Alzheimer's disease (AD), the role of peripheral monocytes remains largely unknown. Circulating monocytes communicate with the brain's resident myeloid cells, microglia, via chemical signaling and can directly infiltrate the brain parenchyma.[1] Thus, molecular signatures of monocytes may serve as indicators of neuropathological events unfolding in the CNS.[2-5] However, no studies have yet directly tested the association of monocyte gene expression on longitudinal cognitive decline or postmortem neuropathology and brain gene expression in aging. Here we present a resource of RNA sequencing of purified CD14+ human monocytes - including an eQTL map - from over 200 elderly individuals, most with accompanying bulk brain RNA sequencing profiles, longitudinal cognitive assessments, and detailed postmortem neuropathological examinations. We tested the direct correlation of gene expression between monocytes and bulk brain tissue, finding very few significant signals driven largely by genetic variation. However, we did identify sets of monocyte-expressed genes that were highly predictive of postmortem microglial activation, diffuse amyloid plaque deposition, and cerebrovascular disease. Our findings prioritize potential blood-based molecular biomarkers for AD; they also reveal the previously unknown architecture of shared gene expression between the CNS and peripheral immune system in aging.

Introduction

Over the past decade, the role of myeloid cells in susceptibility to Alzheimer's disease (AD) has become progressively more established given the growing number of susceptibility loci with evidence of having functional consequences in myeloid cells and the fact that such variants make up a large proportion of known AD susceptibility variants[6,7]. While microglia have received most of the attention to date since they are the resident immune cells of the brain, many other myeloid cells are found in the central nervous system and could have a role in AD since (1) they share many of the transcriptional programs of microglia and (2) enrichment analyses using genetic data are enriched for all myeloid cells. Further, monocytes have been a useful substitute in human functional genetic studies[8,9] and can be differentiated into microglia-like cells that complement induced pluripotent stem cells (iPSC)-derived microglia-like cells as a model system for microglia[10]. Both model systems share a similar extent of gene expression with primary human microglia.

Monocytes, dendritic cells, and microglia-like cells are found in the cerebrospinal fluid[11], and macrophages are found in atherosclerotic lesions, around blood vessels, and in the meninges[12,13]. Further, infiltrating monocytes differentiate into activated parenchymal macrophages that are difficult to distinguish in tissue with current tools from activated microglia. In aging, blood-brain barrier integrity is compromised and monocytes may be more likely to directly infiltrate brain tissue where they can differentiate into microglia-like mononuclear macrophages and further contribute to the central immune response to age-related pathology.[14] Overall, the genetic evidence in AD to date does not exclude any of these cell populations from having a role in AD. Myeloid cells are part of the innate immune system and

have many different roles, including acute and chronic inflammatory responses to pathogens. One example is the association of gingivitis with the risk of AD, which could lead to a chronic inflammatory response that affects the CNS through circulating immune cells and/or mediators of systemic inflammation.[15,16] Alternatively, peripheral cells could be responding to the same mediators of inflammation that affect all other organs, including the CNS.

Given the roles of monocytes in responding to events unfolding within the CNS, their molecular phenotype may be sensitive to neuropathological changes associated with aging even if they are not causally involved, making them a substrate for biomarker discovery in the diagnosis and prognosis of neurodegenerative illness.[17] Here, we set out to examine the role of peripheral blood monocytes – which are mostly recent bone marrow emigrants – and AD by measuring the transcriptome of this cell type purified from blood collected from living participants in a longitudinal study of aging with prospective blood and brain collection. Specifically, we retrieved frozen peripheral blood mononuclear cells from an archive of cryopreserved material from participants who had undergone autopsy and had transcriptomic data generated from the dorsolateral prefrontal cortex. We compare monocyte RNA expression levels directly with those from bulk postmortem brain tissue in the same individuals, and we identify genes that are predictive of cognitive decline and certain aging-related neuropathologies. These associations are validated against results from over a dozen human and preclinical experiments of immune cell transcriptomics in inflammatory and AD-related models. Finally, we explore the moderating effects of *APOE* genotype, clinical AD stage, and biological sex on identified associations.

Results

Cryopreserved peripheral blood mononuclear cells (PBMCs) were thawed and fluorescence activated cell sorting (FACS) was used to isolate the CD14+ cells from each sample into RNA lysis buffer. 2000 sorted CD14+ cells were collected per participant, and a transcriptome was generated from each sample using the SmartSeq2 protocol. Data were pre-processed as outlined in the Methods section and yielded expression values for 9,129 genes in 218 participants. Demographic details of the participants used in downstream analyses are presented in **Supplementary Table 1**.

Monocyte-brain gene expression correlation

We first asked the relatively naïve question of whether genes in the monocyte transcriptome were strongly correlated with the bulk cortical transcriptome in the same individuals; we assessed all 8,757 genes with detectable expression in both tissues. Despite a small but significant positive difference in mean correlation compared to null (1000 permutations of all gene-wise correlations yielded *p*=0, with mean observed rho=0.016), most of the transcriptome has no individual correlation between monocytes and cortex (Figure 1A). A small number of genes did exhibit strong FDR-corrected correlations in their level of RNA expression (170 positively correlated, 6 negatively correlated); negatively correlated genes had rho values ranging from -0.26 to -0.24 (*RPS16, PDE4B, RABGAP1, FOXP1, TSPY26P, RPS28*), whereas the top

10 most positively correlated genes had considerably larger effect sizes, with rho between 0.68-0.9 (*HLA-DRB5, HLA-DRB1, ERAP2, RPS26, U2AF1L5, PI4KAP1, RPL9P9, XRRA1, RPSAP58, SNHG5*).

To understand which biologically relevant pathways are overrepresented in this pool of correlated genes, we performed gene ontology (GO) enrichment analyses of the four gene pools (positive and negatively correlated at nominal *p*<0.05 or corrected FDR<0.05 thresholds) using gprofiler2 ordered queries (Figure 1B-C).[18] To improve interpretability of the results, we semantically collapsed significantly enriched biological processes (see **Methods**), which revealed enrichment of antigen processing and presentation of endogenous peptide antigen in the positively correlated pool of 170 genes (a highly similar enrichment pattern was observed in the larger mutually inclusive pool of 587 nominally correlated genes). Intriguingly, enrichment for pathogenesis of SARS-COV2 mediated by nsp9-nsp10 complex was also observed in this pool, suggesting a connection to systemic inflammation as all brain and blood samples were collected pre-pandemic. Among negatively correlated genes (219 in the nominally correlated pool), enrichment was principally observed for mRNA catabolic process and SRP-dependent co-translational protein targeting to membrane, which are composed primarily of ribosomal protein subunit (RPS) and signal recognition particle (SRP) genes (all significant enrichment results in **Supplementary Table 2**)

Next, to assess whether these correlations represent true proxies in blood for brain expression of certain genes, we generated a map of transcriptome-wide expression quantitative trait locus results for cis effects (cis-eQTL) in monocytes. This eQTL map is accessible online via the AD Knowledge Portal (*will be uploaded prior to publication*. We also accessed brain cis-eQTL data from our prior DLPFC xQTL-serve platform[19] (which was updated in June 2021) as well as GTEx (v8) frontal cortex data (Brodmann area (BA) 9). As illustrated in Figure 1D, there is strong overlap in eQTL between frontal cortex and monocytes: 275/411 egenes in monocytes have significant eQTLs in at least one other dataset, with the majority (189) shared with both (fold-enrichment=5.87, *p*=1.3x10^{-99}). In addition, genes with stronger monocyte:cortex correlations have significantly stronger eQTLs in both tissues (Figure 1E). Thus, while the monocyte RNA expression of these genes could offer a proxy for brain expression, they largely reflect the effect of genetic variants in those genes, and the added utility of measuring this largely constitutive genetic effect on monocyte gene expression is probably limited.

Having established that the monocyte transcriptome offers little in terms of a proxy for measuring the cortical transcriptome, we turned to evaluating the relation of individual gene expression levels to a variety of ante-mortem cognition-related traits, post-mortem neuropathologic indices, and diagnostic categories.

Association of monocyte genes with cognitive and neuropathologic traits

Full monocyte transcriptome differential expression results are visualized in Figure 2A and compared to results obtained from the cortical transcriptome (Figure 2B-C) for the same traits, using the same analytic models.

Focusing first on cognition-related traits, we found few significant associations of monocyte gene expression with measures of cognitive performance in different domains that were collected at the time of blood sampling. The measure of global cognitive performance, which summarizes all of the tests, only had 10 genes meeting FDR<0.05. Most of those genes overlapped with results for the component cognitive measures, notably episodic memory, which is profoundly affected in AD; Figure 2D. By contrast, consistent with prior analyses of these data, the DLPFC transcriptome is broadly associated with cognitive performance. At a more liberal FDR correction threshold of FDR < 0.2, we did find significant overlap between genes associated with semantic memory (p=0.001) and episodic memory (p=0.002) in both tissues.

In evaluating neuropathologic indices, we first focus on the AD-related traits. Looking at a pathologic diagnosis of AD (pathoAD) using the Reagan criteria[20], we again see only a small number of associations, compared to the large changes seen in DLPFC, where measurable atrophy happens late in the disease. Pathologic AD is defined by the severity and extent of neuritic plaques and neurofibrillary tangles, the two defining pathologic features of AD. There are multiple different ways to measure these pathologies, and all strongly associated with a large fraction of the DLPFC transcriptome. In monocytes, we found no associations of gene expression with measures of tau proteinopathy (neurofibrillary tangles (NFT) measured by silver stain or paired helical filament tau (PHF) tau using immunohistochemistry for the AT8 antibody). However, we did observe some associations with measures amyloid β proteinopathy, including an immunofluorescence-based measure of amyloid β (Total Aβ), neuritic plaques measured by silver stain, and diffuse plaques (DP). In particular, the DP trait had more associations than the neuritic plaques or Aβ (n_{genes}=51); though all four genes associated with neuritic plaques were also associated with DP (Figure 2E). We also found associations with severity of cerebral amyloid angiopathy (n_{genes}=281), which is a measurement of Aβ deposition in meningeal and parenchymal blood vessels. This suggests that the monocyte transcriptome may be influenced by or have an effect on some of the same factors that influence the accumulation of amyloid proteinopathy. This narrative connects with the observation that activated microglia surround amyloid plaques and that some AD variants may influence phagocytosis of Aβ.[21]

There were also associations with other neuropathologies; in particular, there were strong associations with cerebral arteriolosclerosis (n_{genes}=635). However, the most prominent associations of monocyte gene expression are found in relation to the proportion of activated microglia (PAM), a trait based on counting the number of microglia with an activated, stage III morphology.[22] The two measures of cortical PAM (from the midfrontal (MF) cortex and inferior temporal (IT) cortex) showed the most associations, despite limited power from the reduced sample size of individuals who have monocyte RNA expression and PAM measures ($n_{subjects}$=51). All measures of PAM had significant overlap in the gene sets they were associated with (Figure 2F), though the strongest overlap was observed between cortical measures ($n_{overlap}$=154), and 7 genes were significantly associated with all four measures ($p_{hypergeometric}$=7.1x10^{-9}; *EXOSC5, MLST8, NUP58, PVT1, RAF1, RNASEH2B, SERINC1*). In Figure 3, we present an alternative visualization of the transcriptome-wide results, emphasizing global cognition, diffuse amyloid,

cerebrovascular, and microglial pathologies. Full differential expression results and scatterplots for top individual gene effects for all phenotypes are available for download online (*will be uploaded to AD Knowledge Portal prior to publication*).

Overall, while there are hundreds of individual genes significantly associated with arteriosclerosis, cerebral amyloid angiopathy, and microglial activation, the monocyte transcriptome does not appear to be strongly associated with the canonical AD-related traits of Total Aβ, neuritic amyloid plaques, PHF tau, or NFTs.

Comparison with Published Myeloid Cell Differential Expression Signatures

In the absence of comparable monocyte datasets with postmortem neuropathologic phenotype data for examination, we extended our findings by accessing summary statistics from the myeloid landscape 2 (ML2) repository.[23,24] This collection of differential gene expression results is assembled from 35 experiments in human and preclinical brain tissues and myeloid cells of the CNS and periphery, each providing effect size and significance estimates for individual gene associations with Alzheimer's disease-related and neuroinflammatory traits, perturbations, preclinical models, and cell populations. We evaluated the overlap of each set of differentially expressed genes from our monocyte analyses for all neuropathologic and cognitive traits with differentially expressed genes from each ML2 experiment in a pairwise manner (on the background of genes common in both analyses; results in **Supplementary Figure 1 and Supplementary Table 3**). Overall, there were many overlapping gene sets (hypergeometric $p<0.05$), though not all showed strong consistency in effect direction concordance. Among the top results were overlaps between genes differentially expressed in cortical CD11b+ microglia (GSE75246[25] & GSE67858[26]) from lipopolysaccharide-injected mice monocyte genes associated with cerebral arteriolosclerosis ($p=6.8\times10^{-8}$), midfrontal PAM ($p=0.027$), and diffuse amyloid plaques ($p=0.011$). Several other overlaps of interest were identified e.g., between monocyte genes associated with posterior putamen PAM and genes differentially expressed in CD11b cells of a tau-P301S mutant mouse model ($p=2.8\times10^{-3}$).

Functional enrichment for biological processes

GO overrepresentation analyses for biological processes were performed separately for up- and down-regulated gene sets for each trait (**Supplementary Table 4**). Given the low number of genes associated with some traits, we also performed rank-based gene set enrichment analyses using the full set of summary statistics for each differential expression analysis, ranked by moderated t-statistic (**top results shown in Supplementary Figure 2; full results Supplementary Table 5**). Many fewer enrichments were observed for overrepresentation in discrete sets of genes than for rank-based tests. GSEA analyses revealed the most significant enrichment for ribosomal genes, translational initiation, and mRNA metabolism across several phenotypes including subcortical PAM measures, cerebral atherosclerosis, TDP-43 proteinopathy, and cognition. The most significant enrichment for diffuse plaque-associated genes were for cellular response to decreased oxygen levels (NES=-2.4, $p_{FDR}=4.5\times10^{-8}$). Notably, beta-

amyloid clearance was significantly enriched in genes negatively associated with midfrontal PAM (NES=-2.0, p_{FDR}=0.01), suggesting that part of the monocyte signature for cortical microglial activation may reflect the loss of activated microglia's ability to clear amyloid effectively.

Analyzing modules of co-expressed genes

To complement the single gene analyses and maximize power to discover broad patterns in gene expression given our moderate sample size, we created modules of co-expressed genes using the WCGNA method[27] (Figure 4). The monocyte transcriptome (Figure 4A) exhibited a somewhat bifurcated pattern of diffuse correlations (two major sets of genes represented in large part by the blue (n_{genes}=1,096) and turquoise (n_{genes}=2,327) modules), with several smaller well-defined clusters of strongly co-expressed genes (a total of 14 clustered modules with 2,595 remaining unclustered genes). In comparison with the clustered DLPFC transcriptome (Figure 4B), there were fewer identified modules (14 vs. 38 modules), consistent with the lack of cell type heterogeneity in our purified monocyte samples and the largely immature nature of monocytes. In line with our single-gene cross-tissue analyses, the purple monocyte module (significantly enriched for ribosomal genes (Figure 4C)) had the strongest shared identity when comparing monocyte and DLPFC module membership (84 shared genes between purple[monocyte][142] and darkolivegreen[dlpfc][128]; hypergeometric p=5.6x10^{-137}). Full results of pairwise comparisons are shown in **Supplementary Figure 3**.

In module-trait association analyses, we observed association of multiple monocyte gene modules with cerebrovascular traits as well as with Lewy body (α synucleinopathy) pathology (Figure 4D), though no associations survived FDR correction for multiple testing. This result was somewhat surprising, given the large number of individual genes associated with microglial activation and cerebrovascular phenotypes in the single-gene analyses. We investigated this further by plotting eigengene loadings for genes in each module with their standardized effect estimates from single-gene differential expression analyses (**Supplementary Figure 4**). This analysis revealed that for modules enriched with significantly-associated genes (e.g. The brown module has 87 genes significantly associated with midfrontal microglial activation), there were substantial numbers of genes with discordant directions of effect but concordant loadings. This demonstrates a lack co-expression among monocyte genes that are associated with brain pathology in the same direction – a result that is supported by the minimal biological process enrichment found in overrepresentation analyses of the PAM phenotype differentially-expressed genes (**Supplementary Table 4**). Interestingly, the yellow monocyte module, which was nominally associated with lower cerebral amyloid angiopathy (beta=-2.2, p_{raw}=0.041) and Lewy body stage (beta=-2.8, p_{raw}=0.035), had the hub gene *ARID4B* (AT-rich interactive domain-containing protein 4B); expression of family member *ARID5B* in human monocytes has been recently associated with carotid arteriosclerosis.[28]

Interaction effects with *APOE* genotype, clinical AD diagnosis, and sex

Finally, we explored whether *APOE* genotype, clinical disease stage, and biological sex moderated the observed relationships between monocyte gene expression and brain pathology and cognition. To test

this, we performed principal components analysis on expression residuals of significantly differentially expressed genes for each phenotype. We then associated the top PC (essentially an eigengene representing cohesive expression of only trait-associated genes; Figure 5A) with their respective outcomes in separate linear or logistic regression models, including an interaction term for: 1) *APOE* ε4 allele carrier status (present vs. absent), 2) clinical AD diagnosis at time of death (no cognitive impairment, mild cognitive impairment, probable AD), and 3) biological sex (male vs. female) (Figure 5B). After FDR correction across all models (18 outcomes x 3 interacting factors = 54 models), one interaction remained significant: that between monocyte gene expression and clinical AD stage on MMSE scores (interaction p=2.5x10^{-4}, p_{FDR}=0.014) suggesting that the relationship between monocyte gene expression and MMSE is only present in those with an eventual consensus diagnosis of dementia likely due to AD (Figure 5C). We interpret this result with caution, as the MMSE is often used to screen for dementia, and homoscedasticity is not maintained across diagnostic categories. Nonetheless, MMSE scores do not determine the consensus AD diagnosis, and the effect is pronounced, so we examined PC1 and found that it loaded most strongly onto *IL6ST* (0.68) and *AMPD3* (-0.72), in opposite directions. This indicates that anticorrelated expression of these genes in monocytes are strongly linked to MMSE score in those with substantial cognitive deficits, where the variation in performance is greatest. The *IL6ST* result aligns well with the association of higher circulating IL6 levels with AD, as IL6ST is one of the subunits of the IL6 receptor. In addition, we found near-significant interactions of monocyte gene expression with *APOE* ε4 status on global cognitive performance (interaction p=2.8x10^{-3}, p_{FDR}=0.15), whereby the association of expression and cognition was stronger in ε4 carriers than in non-carriers (Figure 5D). Expression PC1 for global cognition genes loaded most strongly, in concordant directions, onto *SARS1* (0.47), *ARAP1* (0.44), and *PQLC2* (0.41). No notable interactions were observed for biological sex.

Discussion

This report establishes a new resource of peripheral monocyte gene expression with matched data from human postmortem brain tissue, with all data and results available through the AD Knowledge Portal to facilitate repurposing. A recent review summarized gene expression studies in human monocytes[29], and our data are unique given the availability of rich ante- and post-mortem traits from each participant. The relationship between brain and peripheral gene expression has previously been explored in whole blood, which has limitations given the heterogeneity of component cells and smaller samples sizes of these studies.[30,31] Here, we focused on a cell type robustly implicated in AD from human genetic analyses, and we have demonstrated that the monocyte transcriptome, captured *ex vivo,* has little correlation with the postmortem brain transcriptome measured in the prefrontal cortex. There are some exceptions, with a subset of strongly correlated immune and ribosomal genes; however, these correlations are driven by strong eQTLs that are active in both tissues. We have also shown that the peripheral monocyte transcriptome has little association with the majority of AD-related traits: overall, peripheral monocytes do not appear to have any robust relationship with canonical AD-related pathologies.

However, these monocyte transcriptomes from older individuals were associated with traditionally non-AD pathologies such as cerebral atherosclerosis, consistent with existing work[32]. More interesting are the associations with microglial activation state in both cortical and subcortical brain regions. This result aligns with the intuition that central and peripheral immune cells may respond similarly to systemic cues. However, these signals appear to be diverse and may be driven by a variety of factors, as groups of individually associated genes do not yield evidence of association when collapsed within co-expressed gene modules. Further, while the proportion of activated microglia is associated with AD traits[22], these associations do not appear to be strongly linked to monocyte gene expression. Thus, peripheral monocytes deserve further attention given their relation to microglial states, but it is unlikely that this inter-connection of peripheral and central immunity is strongly related to AD. They may be more relevant for other aging-related pathologic processes.

In more detailed modelling, we found little evidence for the influence of AD disease stage or *APOE* genotype on gene-pathology associations. One exception was an effect of IL6ST on MMSE score that was only observed in individuals with a diagnosis of dementia (likely due to AD), suggesting that the later stages of AD may accompany perturbations of the IL6 pathway and other measures of systemic inflammation that are not part of an early risk profile. Thus, while monocytes may be less likely to have a causal role in AD, they may reflect some of the effects of the disease downstream of cognitive dysfunction, which is different from microglia implicated in the accumulation of amyloid and tau proteinopathy[22,33,34]. Another conditional effect (albeit not significant after correction for multiple testing) was observed for cognition-associated genes, represented most strongly by *SARS1*, *ARAP1*, and *PQLC2*, whereby their association with global cognition was stronger in *APOE* ε4 carriers. Some of these genes have known links to AD; brain expression of *ARAP1* has recently been associated with beta-amyloid load in African-Americans[35] and PQLC2 is a lysosomal lysine/arginine transporter important for the unfolded protein response[36], which is protective against neurodegeneration[37]. It is possible that *APOE* ε4 carriers, which have increased pro-inflammatory response to lipopolysaccharide injection[38], are particularly susceptible to inflammatory alterations in peripheral monocytes; our cognition-associated monocyte genes were significantly enriched for response to interleukin-1.

This study has several limitations. First, the sample size was limited, so we can only exclude the possibility of large effects in interpreting the results of our largely negative results for the AD-related traits. Second, the monocyte RNA sequencing data were produced in two batches; this technical heterogeneity was mitigated with cross-batch normalization and the use of voomWithQualityWeights. However, such methods could over-correct or attenuate certain results. Third, we performed a large number of comparisons and the FDR did not account for testing of 26 phenotypes in addition to all genes in both tissues. This is a trade-off based on the limited sample size; while not strictly conservative, this

approach allows for prioritizing genes for further evaluations. Since rank in the results does not change, our GSEA-based enrichment results and their interpretation are not altered.

In conclusion, we have identified monocyte gene expression signatures for microglial activation, neurovascular pathology, and cognitive performance in late life. Given the roles of monocytes in monitoring and responding to events unfolding within the CNS, their molecular phenotype may be sensitive to some neuropathological changes associated with aging and place them as suitable biomarkers for vascular, inflammatory, and possibly cognitive consequences of neurodegenerative illness in aging.[17]

Methods

Study Participants

All participants in this study were part of the Religious Orders Study or Rush Memory and Aging Project (ROS/MAP)[39]. All subjects were recruited free of known dementia (mean age at entry = 78 ± 7.8 (SD) years), agreed to annual clinical and neurocognitive evaluation, and signed an Anatomical Gift Act allowing for brain autopsy at time of death. Written informed consent was obtained from all participants and study protocols were approved by an Institutional Review Board of Rush University Medical Center. All participants signed a repository consent that allowed for resource sharing. Data can be requested at https://www.radc.rush.edu.

Assessment of Cognitive Performance

All subjects were administered 17 cognitive tests annually spanning five cognitive domains. Raw scores for tests within each domain were z-scored (using the mean and standard deviation of the entire ROS/MAP cohort at baseline) and averaged to form the composite measures used as outcomes in differential expression analysis. The list of individual cognitive tasks and their corresponding domains has been published[22]. Further, the Mini Mental Status Examination (MMSE), a widely used 30 item measure of global cognition and dementia severity[40], was also administered.

Assessment of Postmortem Neuropathology and Consensus Diagnoses

With an average postmortem interval of 9.3 hours (SD=8.1), brains were removed in a standard fashion as previously described.[41] All brains were examined by a board-certified neuropathologist blinded to clinical data. A total of 18 disease- and age-related neuropathological and were measured brain-wide, including validated measures of Aβ peptides, neuritic and diffuse plaques, hyperphosphorylated tau protein, neurofibrillary tangles, micro and macro cerebral infarcts, cerebral atherosclerosis, degree of alpha-synucleinopathy, TDP43 proteinopathy, and hippocampal sclerosis. A binary summary diagnosis of neuropathologic Alzheimer's disease was also calculated according to NIA-REAGAN criteria (absent if likelihood was no or low and present if intermediate or high)[20]. A subset of samples were also evaluated for the presence of microglia at three stages of activation in four regions (midfrontal (MF) cortex, inferior temporal (IT) cortex, ventral medial caudate (VM), and posterior putamen (PPUT)), based on morphology: stage I (thin ramified processes), stage II (plump cytoplasm and thicker processes), and stage III (appearance of macrophages). Detailed descriptions of all neuropathological variables have been previously published.[22] In addition, at the time of death, a consensus summary diagnostic opinion was rendered by one or more neurologist(s) regarding the most likely clinical diagnosis of Parkinson's disease, blind to neuropathological data.[42]

Isolation of Peripheral Blood Monocytes

Peripheral blood mononuclear cells (PBMCs) were isolated from blood and frozen. CD14+CD16- monocytes were isolated from thawed PBMCs using the EasySep Human Monocyte Isolation Kit (Negative selection kit, Stemcell Technologies, 19359) according to manufacturer's instructions (details in **Supplementary methods**).

Sequencing of RNA from Monocytes

RNA sequencing libraries were prepared using a SMART-seq2 (Batch 1 and 2) protocol for cDNA preparation followed by Nextera XT DNA library preparation. For each sample, cDNA was prepared from 20-2000 FACS-sorted cells. Libraries were pooled and sequenced on a HiSeq 2500 (Illumina) using 2 x 101bp (Batch 1) and 2 x 76bp cycles (Batch 2). Full details are available on the ROS/MAP AMP-AD Knowledge portal (syn3219045) and have been made available in **Supplementary Methods**.

Sequencing of RNA from Bulk Brain (DLPFC)

RNA sequencing on DLPFC tissue was carried out in 13 batches within three distinct library preparation and sequencing pipelines. All samples were extracted using Qiagen's miRNeasy mini kit (cat. no. 217004) and the RNase free DNase Set (cat. no. 79254), and quantified by Nanodrop and quality was evaluated by Agilent Bioanalyzer. Full details on these methods are available on the AMP-AD knowledge portal (syn3219045). Briefly, for pipeline #1, The Broad Institutes's Genomics Platform performed RNA-Seq library preparation using the strand specific dUTP method[43] with poly-A selection[44]. Sequencing was performed on the Illumina HiSeq with 101bp paired-end reads and achieved coverage of 150M reads of the first 12 samples. The remaining samples were sequenced with coverage of 50M reads. For pipeline #2, RNA sequencing libraries were prepared using the KAPA Stranded RNA-Seq Kit with RiboErase (kapabiosystems) in accordance with the manufacturer's instructions. Sequencing was performed on the Illumina NovaSeq6000 using 2 x 100bp cycles targeting 30 million reads per sample. For pipeline #3, RNA was extracted using Chemagic RNA tissue kits (Perkin Elmer, CMG-1212) on a Chemagic 360 instrument. 500ng total RNA was used as input for sequencing library generation and rRNA was depleted with RiboGold (Illumina, 20020599). A Zephyr G3 NGS workstation (Perkin Elmer) was utilized to generate TruSeq stranded sequencing libraries (Illumina, 20020599). Sequencing was performed on a NovaSeq 6000 (Illumina) at 40-50M reads (2x150bp paired end).

RNAseq Processing and Quality Control

Alignment pipeline and quality control

For monocytes, two batches (n_{batch1}=46, n_{batch2}=201; initial n_{total}=247) were processed using the same pipeline: 1) fastq file quality control was performed using FastQC v0.11.5 (default parameters), 2) STAR v2.5.3a was used to align reads (GRCh38.91 reference), 3) RSEM v1.2.31 was used to quantify expression from aligned BAM files, and 4) multiqc v1.5 was used to aggregate quality metrics from fastqc and Picard tools v2.17.4, 5) quality reports were examined for each batch and exclusion of samples was initially carried out according to manual identification of outlying samples primarily considering low numbers of aligned reads, excess GC coverage bias, high percentage of read duplicates, and abnormal distribution of read assignments across genomic annotations (nine outliers were identified at this stage and removed; new n_{total}=238). For DLPFC, the identical pipeline was applied for all 13 batches (initial n_{total}=1,110). Expression of the *XIST* gene was evaluated at this stage to exclude subjects with contradictory reported biological sex. For monocytes, five subjects were identified (new n_{total}=233), and for DLPFC, 13 subjects were identified (new n_{total}=1,097).

Count data quality control and initial outlier removal

Expected counts, calculated by RSEM, were aggregated across both batches and used as input to limma (v3.48.3) voom in R (v4.1.1). Genes with insufficient expression (median count value was less than 15 across the combined sample) were removed. Naïve multidimensional scaling (MDS) analysis was then performed for each batch separately on the top 5000 most variable genes (using limma "plotMDS") to identify subjects with outlying expression patterns. Outliers were defined as those with values of either of the first two latent dimensions exceeding ±4 times the interquartile range (IQR) of their within-batch median value. Following this step, 224 subjects remained for monocytes, and 1,091 remained for DLPFC. To limit bias in modeling due to extreme within-subject outlier observations, we conservatively coerced extreme expression values separately per gene: any values beyond 8xIQR of the voom-transformed sample median values were coerced to the nearest maximum or minimum point of the sample distribution. In this way, we were able preserve some influence for such observations in linear modeling while limiting leverage inflation. These QC'd data were then trimmed mean of M-values (TMM) normalized.

Co-variate selection

Next, co-variates for inclusion in downstream analyses were determined by evaluating the effects of candidate variables on the top 20 principal components (PCs) of gene expression in monocytes and DLPFC separately (**Supplementary Figures 5 & 6**):

For monocytes, influential technical variables were determined to be: batch, % of usable bases, % of passed filter reads aligned, % read duplicates, median 3' bias, estimated library size, and study (ROS vs. MAP). In addition, we observed effects on expression PCs by important biological and clinical variables measured at time of blood draw that were also included in downstream association analyses of monocyte expression: biological sex, age, fasting status, hematologic medication status, blood hemoglobin, mean corpuscular volume (MCV), mean corpuscular hemoglobin concentration (MCHC), platelet count, and white blood cell count (WBC). These variables were not selected in isolation; correlations between all potential co-variates were examined prior to inclusion and variables were selected to avoid redundancy (**Supplementary Figures 7 & 8**). In downstream models of postmortem outcomes, age at death and postmortem interval (PMI) were also included as covariates. For cognitive outcomes, years of education were included.

For DLPFC, the following variables were selected using the same procedure described above: batch, study, biological sex, age at death, PMI, median coefficient of variation for coverage values of the 1000

most highly expressed genes, % of aligned bases mapping to ribosomal RNA, % coding bases, % UTR bases, log(estimated library size), log(passed filter aligned reads), median 5' to 3' bias, % of passed filter reads aligned, % read duplicates, median 3' bias, and % of intergenic bases. Again, for cognitive outcomes, years of education were also included.

Estimation of brain cell type proportions in DLPFC

Prior to calculating residual expression values, a brain cell type-corrected expression matrix was generated for DLPFC on voom-transformed expression values. This matrix was used as input for cross-tissue correlation analyses, transcriptome-wide differential expression analyses, as well as WGCNA. Cell type proportions were estimated using the Brain Cell Type Specific Gene Expression Analysis (BRETIGEA)[45] package in R. Human marker genes (n=50 per set) from Darmanis et al. (2015)[46] were used in the application of a validated singular value decomposition method[47] (the "adjustBrainCells" function).

Final QC for post-processed expression residuals

As a final step to ensure the robustness of included data, linear effects of identified technical co-variates, age at draw (age at death and PMI for DLPFC), and biological sex were removed using the "lmfit" (specifying robust Huber regression[48]), and expression values residualized. Individuals were then hierarchically clustered (agglomerative) and the resulting dendrograms (**see Supplementary Figure 9**) were manually inspected to identify any additional subject outliers escaping recognition by two-dimensional MDS (6 identified for monocytes, 18 for DLPFC). This resulted in two high quality RNAseq datasets: monocytes (218 subjects and 9,129 genes) and DLPFC bulk tissue (n_{genes}=17 465).

Weighted Gene Co-Expression Network Analysis (WGCNA)

The WGCNA pipeline (v1.70-3)[27] was used to detect co-expressed gene modules in the monocyte and DLPFC datasets separately using the signed network approach. For network construction, gene expression residuals were used as input; for monocytes, covariates included all selected technical variables plus age at blood draw and biological sex. This set of residuals was chosen to allow for downstream module-trait analyses on both cognitive and postmortem variables, which require different analytical covariates. DLPFC expression residuals adjusted for all selected variables listed above, in addition to brain cell type proportions. Briefly, WGCNA calculates a topological overlap matrix (TOM)

based on the signed correlation between all input genes, raised to a power (which optimizes scale-free network topology and improves module detection). To improve the robustness of our TOM network structure, we used biweight midcorrelations as our pairwise similarity measure (corType="bicor"), allowing for 5% of observations for any given correlation to be considered outliers (maxPOutliers=0.05). The optimal soft threshold for calculation of adjacency was determined by visual assessment of the scale free topology index R^2 calculated from the "pickSoftPowerThreshold" function (8 for monocytes and 16 for DLPFC; see **Supplementary Figure 10** for threshold determination experiment results). The TOM is then hierarchically clustered, and distinct clusters of genes detected using a dynamic tree-cutting and merging process (non-default parameters: minClusterSize=30, deepSplit=3, detectCutHeight=0.995, minKMEtoStay=0.3, pamStage=TRUE, pamRespectsDendro=TRUE). After clustering into gene modules, eigen-decomposition was used to identify latent features capturing the most linear covariance among all genes expressed within each module and each dataset. These "eigengenes" therefore summarize the expression of all genes within a given module, weighted by the similarity of each member gene to the other members. This resulted in a set of high confidence, structurally cohesive gene module identities for further examination. Gene modules identified in monocytes and DLPFC separately were assessed for overlapping identities (gene members) using the WGCNA "overlapTable" and function.

Monocyte Expression Quantitative Trait Loci (eQTL) Analysis

Genotype data was available for 2067 ROS/MAP subjects, between two batches: n_{batch1}=1686 genotyped using the Affymetrix GeneChip 6.0 and n_{batch2}=381 genotyped using the Illumina OmniQuad Express platform. Details of raw genotype quality control have been previously described.[49] Each batch was imputed separately using the TOPMed Imputation Server (TOPMed reference r2),[50] including Eagle (v2.4) for allelic phasing and Minimac4 (v1.5.7) for imputation. Prior to submission for imputation, genotypes were preprocessed using the TOPMed Imputation Server-recommended data preparation pipeline available here: https://topmedimpute.readthedocs.io/en/latest/prepare-your-data.html. Imputed output data from the TOPMed server for each batch were filtered for imputation quality (removing SNPs with r<0.8) before merging and mapping to rsIDs (dbSNP build 155). This resulted in a final high-quality dataset of 9,329,439 bi-allelic autosomal SNPs. PLINK2 (v2.00a3) was used to perform QC on SNP data prior to eQTL analysis, including removing variants with minor allele frequency (MAF) > 0.05 and violating Hardy-Weinberg Equilibrium (HWE) with a p-value below 1×10^{-6}. BootstrapQTL (v1.0.5)[51] was used to perform cis-eQTL mapping within a distance of 1MB from the start and stop sites of each gene passing QC in our monocyte RNAseq experiment (GRCh38 coordinates). Analyses co-varied for biological sex, age at blood draw, and the top 10 genomic PCs estimated from only genotyped SNPs overlapping between genotyping platforms (n_{snps}=188,936). 1000 bootstrap iterations were performed to address the

"winner's curse" common in cis-eQTL mapping studies[51] and maximize the confidence of our results. Code used for cis-eQTL mapping is available at (www.github.com/dfelsk/monocyteRNAseq).

External Postmortem Brain Cis-eQTL Resources

For comparison of our monocyte is-eQTL map with cis-eQTL data from human frontal cortex, we accessed two existing resources: the GTEx cis-eQTL database v8 (dbGaP accession phs000424.v8.p2) and the recently updated (June 2021) xQTL-serve[19], which was built directly from DLPFC expression and genotype data from the ROS/MAP studies. Specifically, for GTEx, the single-tissue egene collection for frontal cortex (BA9) was downloaded. For both datasets, each gene was matched to its strongest eQTL (i.e. SNP) for downstream analysis.

Gene Set Enrichment Analyses

The R Bioconductor gprofiler2 (v0.2.0)[18] package was used for gene set overrepresentation analyses across a broad range of annotations in cross-tissue correlation analysis (Figure 1B-C). For biological process gene ontology (GO) overrepresentation analysis of module membership and rank-based gene set enrichment analysis (GSEA) of transcriptome-wide differential expression results, the clusterProfiler R package (v4.0.2) was used (parameters: minGSSize=20; maxGSSize=300; OrgDb=org.Hs.eg.db). For GSEA, an FDR-corrected threshold of q<0.05 was applied across all tested GO gene sets within each phenotype. Following enrichment tests, the rrvgo (v1.5.2; https://ssayols.github.io/rrvgo/) package was used to perform semantic similarity reduction of enriched GO terms of informative GO terms. Rrvgo uses semantic similarity ("Rel" threshold set at 0.7) to simplify and improve interpretability of GO analyses where many terms may be significantly enriched.

Extension of Findings in Publicly Available Myeloid Cell Experiments

To test if genes identified in our monocyte differential expression analyses had been identified previously in myeloid cells or brain tissue as associated with traits related to Alzheimer's disease of neuroinflammation, we systematically queried the myeloid landscape 2 online database.[23,24] This collection of differential expression results is assembled from 35 experiments in human and preclinical brain tissues and myeloid cells of the CNS and periphery, each providing gene-wise effect size and

significance estimates. Full descriptions of each component study and their analyses can be found online (http://research-pub.gene.com/BrainMyeloidLandscape/BrainMyeloidLandscape2/#).

Statistical Analyses

Correlation of gene expression between tissues

Using covariate-normalized gene expression residuals, Spearman rank correlations were calculated for each gene between monocytes and DLPFC, followed by correction for multiple testing using Benjamini-Hochberg FDR.[52] To assess whether the mean correlation was different from 0, we performed permutation analysis, randomly shuffling subjects in both expression datasets and re-calculating correlation coefficients for every gene 1000 times. This procedure yielded a null distribution of mean correlations from which the p-value for the null test ($H_0 : \mu_{cross\text{-}tissue\ \mathrm{rho}} \neq 0$) could be calculated: $p =$ proportion of permutations for which the absolute value of the mean correlation was greater than the mean correlation from our observed, unshuffled data.

Comparing cis-eQTL results with cross-tissue correlated genes

Values for assessing evidence for genetic regulation were calculated for each gene with at least one significant eQTL in both monocytes and DLPFC (corrected $p_{eQTL} < 0.05$) as the $-\log_{10}$ Fisher's meta p-value of its top eSNPs in both tissues (not necessarily the same lead SNPs). Cross-tabulation followed by two-sided Fisher's exact test was used to visualize and statistically test for pairwise overlap in significant ($p_{FDR} < 0.05$) eQTLs between all three datasets (monocytes, DLPFC, and GTEx frontal cortex). Hypergeometric tests were used to identify significance of overlap of significant cross-tissue correlated genes with genes possessing significant eQTLs in each dataset separately. For consistency in visualization, we also performed a Kruskall-Wallis test on eQTL significance ($-\log_{10}(p$-value$)$) between three pools of mutually-exclusive genes used in the **Figure 1A** visualization: uncorrelated ($p > 0.05$), nominally correlated ($p < 0.05$ & $p_{FDR} > 0.05$), and significantly correlated ($p_{FDR} < 0.05$). Post-hoc Wilcoxon rank sum tests were performed to identify pairwise differences between groups, with Bonferroni correction for multiple testing (three tests; $p_{threshold} = 0.0176$).

Transcriptome-wide differential expression analysis

Differential expression analysis was performed using limma/voom for associations with neuropathology and cognitive performance outcomes. For monocytes, filtered gene expression values processed with voomWithQualityWeights were used as input, co-varying for all technical and biological confounders specified above. For DLPFC, the same approach was used, except filtered and voom-transformed expression values were first corrected for brain cell type proportions, as described above. Robust linear modeling was used for differential expression, allowing up to 20,000 iterations to reach convergence. Significance of the effects for target outcomes in our multivariate models was performed using empirical Bayes moderation (eBayes function). *P*-values were adjusted using the FDR approach within each phenotype and tissue separately.

Comparison of differential expression results between monocytes and DLPFC

For each cognition-related and neuropathologic outcome, hypergeometric tests were used to determine if the identity of associated genes significantly overlapped between monocytes and DLPFC. These tests used the background of genes commonly expressed in both tissues (n_{genes}=8,757) and were performed for three levels of FDR significance (FDR<0.05, 0.1, and 0.2). In addition, hypergeometric tests were applied to subsets of outcomes within each tissue, to determine if genes associated with similar traits also had overlapping identities. Tests were implemented using the SuperExactTest (v1.0.7)[53] package in R.

Association of gene modules with cognitive and neuropathological variables

Module eigengenes were associated with the same set of cognitive and pathological outcomes as in transcriptome-wide differential expression analyses, using the same covariates. Ordinary least squares regression was used for association. FDR-based multiple testing correction was as applied across the full set of modules and phenotypes, but separately for monocyte and DLPFC analyses.

Data Availability

All data used in the analyses described can be found on the AMP-AD Knowledge Portal (https://adknowledgeportal.synapse.org/).

References

1.   Mosser, D. M. & Edwards, J. P. Exploring the full spectrum of macrophage activation. *Nat. Rev. Immunol.* **8**, 958–969 (2008).

2.   Baufeld, C., O'Loughlin, E., Calcagno, N., Madore, C. & Butovsky, O. Differential contribution of microglia and monocytes in neurodegenerative diseases. *J. Neural Transm. Vienna Austria 1996* **125**, 809–826 (2018).

3.   Farmen, K. *et al.* Monocyte markers correlate with immune and neuronal brain changes in REM sleep behavior disorder. *Proc. Natl. Acad. Sci.* **118**, (2021).

4.   Varvel, N. H. *et al.* Infiltrating monocytes promote brain inflammation and exacerbate neuronal damage after status epilepticus. *Proc. Natl. Acad. Sci.* **113**, E5665–E5674 (2016).

5.   Zhao, M., Tuo, H., Wang, S. & Zhao, L. The Roles of Monocyte and Monocyte-Derived Macrophages in Common Brain Disorders. *BioMed Res. Int.* **2020**, e9396021 (2020).

6.   Wightman, D. P. *et al. Largest GWAS (N=1,126,563) of Alzheimer's Disease Implicates Microglia and Immune Cells*. 2020.11.20.20235275 https://www.medrxiv.org/content/10.1101/2020.11.20.20235275v1 (2020) doi:10.1101/2020.11.20.20235275.

7.   Novikova, G. *et al.* Integration of Alzheimer's disease genetics and myeloid genomics identifies disease risk regulatory elements and genes. *Nat. Commun.* **12**, 1610 (2021).

8.   Chan, G. *et al.* CD33 modulates TREM2: convergence of Alzheimer loci. *Nat. Neurosci.* **18**, 1556–1558 (2015).

9.   Huang, K. *et al.* A common haplotype lowers PU.1 expression in myeloid cells and delays onset of Alzheimer's disease. *Nat. Neurosci.* **20**, 1052–1061 (2017).

10.   Ryan, K. J. *et al.* Context-specific effects of neurodegenerative disease variants in a model of human microglia. *Sci. Transl. Med.* **9**, eaai7635 (2017).

11.   Roostaei, T. *et al. Defining the architecture of cerebrospinal fluid cellular communities in neuroinflammatory diseases*. 2021.11.01.466797 https://www.biorxiv.org/content/10.1101/2021.11.01.466797v2 (2021) doi:10.1101/2021.11.01.466797.

12.   Prinz, M., Erny, D. & Hagemeyer, N. Ontogeny and homeostasis of CNS myeloid cells. *Nat. Immunol.* **18**, 385–392 (2017).

13.   Willemsen, L. & de Winther, M. P. Macrophage subsets in atherosclerosis as defined by single-cell technologies. *J. Pathol.* **250**, 705–714 (2020).

14.   Sweeney, M. D., Sagare, A. P. & Zlokovic, B. V. Blood-brain barrier breakdown in Alzheimer disease and other neurodegenerative disorders. *Nat. Rev. Neurol.* **14**, 133–150 (2018).

15.   Noble, J. M. *et al.* Serum IgG antibody levels to periodontal microbiota are associated with incident Alzheimer disease. *PloS One* **9**, e114959 (2014).

16.   Miklossy, J. & McGeer, P. L. Common mechanisms involved in Alzheimer's disease and type 2 diabetes: a key role of chronic bacterial infection and inflammation. *Aging* **8**, 575–588 (2016).

17.   Schmitz, G., Leuthäuser-Jaschinski, K. & Orsó, E. Are circulating monocytes as microglia orthologues appropriate biomarker targets for neuronal diseases? *Cent. Nerv. Syst. Agents Med. Chem.* **9**, 307–330 (2009).

18.   Kolberg, L., Raudvere, U., Kuzmin, I., Vilo, J. & Peterson, H. gprofiler2 – an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. *F1000Research* **9**, ELIXIR-709 (2020).

19.   Ng, B. *et al.* An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome. *Nat. Neurosci.* **20**, 1418 (2017).

20.   NIA-Reagan. Consensus recommendations for the postmortem diagnosis of Alzheimer's disease. The National Institute on Aging, and Reagan Institute Working Group on Diagnostic Criteria for the Neuropathological Assessment of Alzheimer's Disease. *Neurobiol. Aging* **18**, S1-2 (1997).

21.   Hickman, S. E., Allison, E. K. & Khoury, J. E. Microglial Dysfunction and Defective β-Amyloid Clearance Pathways in Aging Alzheimer's Disease Mice. *J. Neurosci.* **28**, 8354–8360 (2008).

22.   Felsky, D. *et al.* Neuropathological correlates and genetic architecture of microglial activation in elderly human brain. *Nat. Commun.* **10**, 409 (2019).

23.   Srinivasan, K. *et al.* Alzheimer's Patient Microglia Exhibit Enhanced Aging and Unique Transcriptional Activation. *Cell Rep.* **31**, (2020).

24.   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).

25.   Srinivasan, K. *et al.* Untangling the brain's neuroinflammatory and neurodegenerative transcriptional responses. *Nat. Commun.* **7**, 11295 (2016).

26.   Erny, D. *et al.* Host microbiota constantly control maturation and function of microglia in the CNS. *Nat. Neurosci.* **18**, 965–977 (2015).

27.   Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).

28.   Liu, Y. *et al.* Blood monocyte transcriptome and epigenome analyses reveal loci associated with human atherosclerosis. *Nat. Commun.* **8**, 393 (2017).

29.   Cormican, S. & Griffin, M. D. Human Monocyte Subset Distinctions and Function: Insights From Gene Expression Analysis. *Front. Immunol.* **11**, (2020).

30.   Sullivan, P. F., Fan, C. & Perou, C. M. Evaluating the comparability of gene expression in blood and brain. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet. Off. Publ. Int. Soc. Psychiatr. Genet.* **141B**, 261–268 (2006).

31.   Tylee, D. S., Kawaguchi, D. M. & Glatt, S. J. On the outside, looking in: a review and evaluation of the comparability of blood and brain '-omes'. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet. Off. Publ. Int. Soc. Psychiatr. Genet.* **162B**, 595–603 (2013).

32.   Gupta, R. M., Lee-Kim, V. S. & Libby, P. The March of Monocytes in Atherosclerosis. *Circ. Res.* **126**, 1324–1326 (2020).

33.   Patrick, E. *et al.* A cortical immune network map identifies distinct microglial transcriptional programs associated with β-amyloid and Tau pathologies. *Transl. Psychiatry* **11**, 1–17 (2021).

34.   Bradshaw, E. M. *et al.* CD33 Alzheimer's disease locus: Altered monocyte function and amyloid biology. *Nat. Neurosci.* **16**, 848–850 (2013).

35.   Kunkle, B. W. *et al.* Novel Alzheimer Disease Risk Loci and Pathways in African American Individuals Using the African Genome Resources Panel: A Meta-analysis. *JAMA Neurol.* **78**, 102–113 (2021).

36.   Liu, B., Du, H., Rutkowski, R., Gartner, A. & Wang, X. LAAT-1 Is the Lysosomal Lysine/Arginine Transporter That Maintains Amino Acid Homeostasis. *Science* **337**, 351–354 (2012).

37.   Hetz, C., Zhang, K. & Kaufman, R. J. Mechanisms, regulation and functions of the unfolded protein response. *Nat. Rev. Mol. Cell Biol.* **21**, 421–438 (2020).

38.   Gale, S. C. *et al.* APOε4 is associated with enhanced in vivo innate immune responses in human subjects. *J. Allergy Clin. Immunol.* (2014) doi:10.1016/j.jaci.2014.01.032.

39.   Bennett, D. A. *et al.* Religious Orders Study and Rush Memory and Aging Project. *J. Alzheimers Dis. JAD* **64**, S161–S189 (2018).

40.   Folstein, M. F., Folstein, S. E. & McHugh, P. R. 'Mini-mental state'. A practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* **12**, 189–198 (1975).

41.   Bennett, D. A. *et al.* Neuropathology of older persons without cognitive impairment from two community-based studies. *Neurology* **66**, 1837–1844 (2006).

42.   Schneider, J. A., Arvanitakis, Z., Bang, W. & Bennett, D. A. Mixed brain pathologies account for most dementia cases in community-dwelling older persons. *Neurology* **69**, 2197–2204 (2007).

43.   Levin, J. Z. *et al.* Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nat. Methods* **7**, 709–715 (2010).

44.   Adiconis, X. *et al.* Comprehensive comparative analysis of RNA sequencing methods for degraded or low input samples. *Nat. Methods* **10**, 623–629 (2013).

45.   McKenzie, A. T. *et al.* Brain Cell Type Specific Gene Expression and Co-expression Network Architectures. *Sci. Rep.* **8**, 1–19 (2018).

46.   Darmanis, S. *et al.* A survey of human brain transcriptome diversity at the single cell level. *Proc. Natl. Acad. Sci.* **112**, 7285–7290 (2015).

47.   Chikina, M., Zaslavsky, E. & Sealfon, S. C. CellCODE: a robust latent variable approach to differential expression analysis for heterogeneous cell populations. *Bioinforma. Oxf. Engl.* **31**, 1584–1591 (2015).

48.   Huber, P. J. *Robust statistics*. (1981).

49.   Jager, P. L. D. *et al.* A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research. *Sci. Data* **5**, 180142 (2018).

50.   Taliun, D. *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290–299 (2021).

51.   Huang, Q. Q., Ritchie, S. C., Brozynska, M. & Inouye, M. Power, false discovery rate and Winner's Curse in eQTL studies. *Nucleic Acids Res.* **46**, e133–e133 (2018).

52.   Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300 (1995).

53.   Wang, M., Zhao, Y. & Zhang, B. Efficient Test and Visualization of Multi-Set Intersections. *Sci. Rep.* **5**, 16923 (2015).

Declarations

Author Contributions

D.F. was responsible for study design, data management and pre-processing, ROS/MAP statistical analyses, and writing of the manuscript. T.R. and E.M.B. contributed to the study design, ROS/MAP statistical analyses, and editing of the final manuscript. S.L.R. and K.N. were responsible for processing IMAS PET imaging data and performing all imaging analyses. V.P. was responsible for the SRM proteomic methodology and analysis, and assisted in editing of the final manuscript. J.A.S. was responsible for overseeing neuropathological data acquisition, ensuring quality control of the data, and editing the final manuscript. A.S. was responsible for IMAS study design and acquisition of IMAS data, and manuscript editing. D.A.B. oversees the ROS/MAP studies, contributed to study design, and assisted in manuscript editing. P.L.D.J. contributed to the study design, evaluation of results, and writing of the manuscript. All authors read and approved the final manuscript.

Competing Interests

The authors declare no competing interests.

Materials and Correspondence

Please direct correspondence to Drs. Daniel Felsky or Philip L. De Jager.

Figures

Figure 1

Cross-Tissue Correlation Analysis. A) Histogram of Spearman rho correlation coefficients for 8,757 genes detected in both monocytes and DLPFC tissue (n=194). The yellow dashed line indicates the mean correlation (μ_{rho}=0.016), and colors correspond to non-overlapping categories of significance following multiple testing correction (p-values are two-sided). Panel A also shows the histogram of 1000 permuted correlation means, demonstrating p_{perm}<0.001 for test of mean of transcriptomic correlations being equal to 0. Functional enrichment result summaries are shown for B) positively correlated genes, and C) negatively correlated genes. Highlighted top terms were chosen to emphasize key consolidated processes and are described below dot plots for added detail. D) Venn diagram showing the overlap in significant

cis-eGenes (genes with significant cis-eQTLs after correction) in monocytes, DLPFC (xQTL serve; ROS/MAP), and frontal cortex (BA9; GTEx v8). Fold-enrichment (FE) and p-value shown from hypergeometric test of overlap of all three tissue eQTLs. E) Beeswarm plots showing median cis-eQTL strength (y-axis) for all eGenes in each tissue as a function of cross-tissue correlation significance group (x-axis; 0=uncorrelated, 1=nominally correlated at $p_{uncorrected}<0.05$ and $p_{FDR}>0.05$; 2=significantly correlated at $p_{FDR}<0.05$). Omnibus Kruskall-Wallis tests revealed significant effects of correlation significance group on eQTL strength in all tissues (all $p<4.7\times10^{-9}$), with all post-hoc Wilcoxon rank-sum tests showing $p<0.001$ after Bonferroni correction except for the contrast between uncorrelated and nominally correlated genes in monocytes ($p_{Bonf}=0.29$). All p-values are two-sided. BP=biological process; CC=cellular component; DLPFC=dorsolateral prefrontal cortex; eQTL=expression quantitative trait locus; GO=gene ontology; MF=molecular function; WP=Wikipathways.
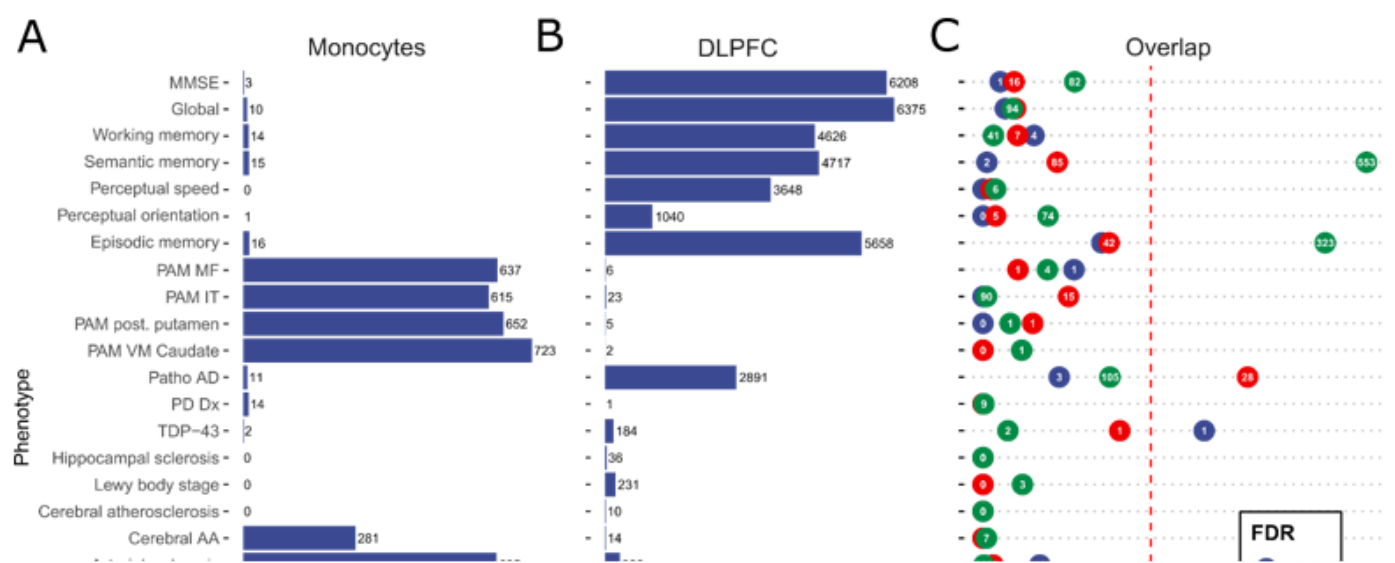

Figure 2

Summary of Results from Transcriptome-wide Differential Expression Analysis in Monocytes and DLPFC. A) and B) show the number of genes passing the FDR-corrected significance threshold of p_{FDR}<0.05 (two-sided). C) The number of genes passing the indicated FDR threshold for trait association overlapping between monocytes and DLPFC. The x-axis indicates significance of hypergeometric overlap test with the full background of genes detected in both monocytes and DLPFC (n_{genes}=8,757). As a visual aid for interpretation, the vertical red dashed line indicates an uncorrected p-value of 0.05. D-F) Sets of genes passing p_{FDR}<0.05 in monocytes in common between related cognitive and neuropathologic groups of traits. Bar height shows the number of genes in common between significant gene sets for all traits with filled-in black dots directly below. Each bar and corresponding overlap set is mutually exclusive. The bar fill color scale indicates significance of hypergeometric tests for the indicated overlap, showing substantial overlap between genes significantly associated with related traits. AA=amyloid angiopathy; AB=total amyloid-beta; Dx=diagnosis; Global=global cognitive summary score; IT=inferior temporal cortex; MF=midfrontal cortex; MMSE=mini mental status exam; NFT=neurofibrillary tangles; PAM=proportion of active microglia; PD=Parkinson's disease; TDP-43=tar-DNA binding-43 proteinopathy; VM=ventral medial.
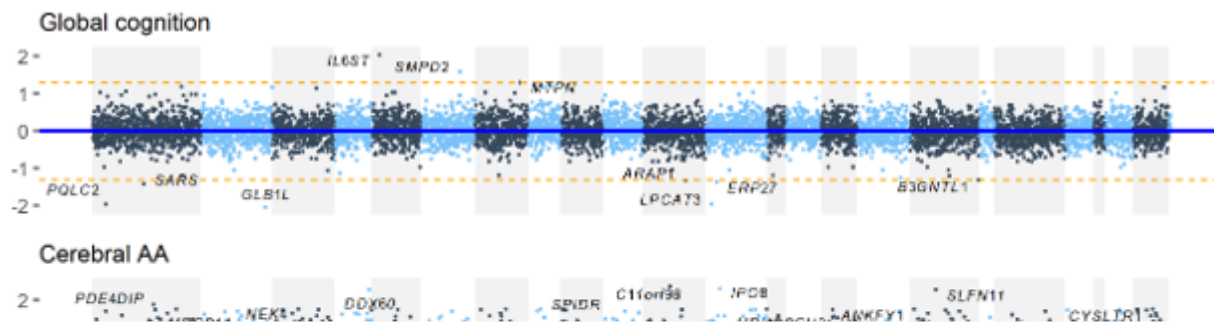
Figure 3

Chicago Plots Showing Detailed Results of Differential Expression Analysis in Monocytes for Selected Traits. Y-axes show the signed $\log_{10}(p_{FDR})$ for each gene (n_{genes}=9,129), x-axes and highlighted segments show chromosomes. The top 20 genes are labelled for reach trait, except where fewer genes meet corrected significance (p_{FDR}<0.05), in which case only those significant genes are labelled. MT=mitochondrial chromosome.

Figure 4

Monocyte Gene Co-Expression Module Analysis. A) network visualization of monocyte transcriptome colored by WGCNA modules, accompanied by histogram showing size and central hub gene of each module. B) Gene dendrograms for the monocyte and DLPFC transcriptomes processed using the same analytical pipeline. Module identities are shown in the colored bar below each dendrogram. The central alluvial plot shows overlapping identities of gene modules, with each curved line connection representing one of the 8,757 genes commonly expressed in both tissues. C) Results of gene enrichment analysis for biological processes. Only the top five significant GO terms are shown for each module, and only five modules were significantly enriched for at least one term ($p_{FDR} < 0.05$). Gene module sizes in GO enrichment analysis may differ slightly from those reported elsewhere as mapping from Ensembl gene IDs to Entrez gene IDs is imperfect. D) Heatmaps showing association statistics for each gene module in monocytes and DLPFC against each cognitive and neuropathological outcome (colors indicate strength and direction of association t-statistic, with blue indicating negative and red indicating positive values; $*p_{raw} < 0.05$, $**p_{FDR} < 0.05$). AA=amyloid angiopathy; AB=total amyloid-beta; Dx=diagnosis; Global=global cognitive summary score; IT=inferior temporal cortex; MF=midfrontal cortex; MMSE=mini mental status exam; NFT=neurofibrillary tangles; PAM=proportion of active microglia; PD=Parkinson's disease; TDP-43=tar-DNA binding-43 proteinopathy; VM=ventral medial.
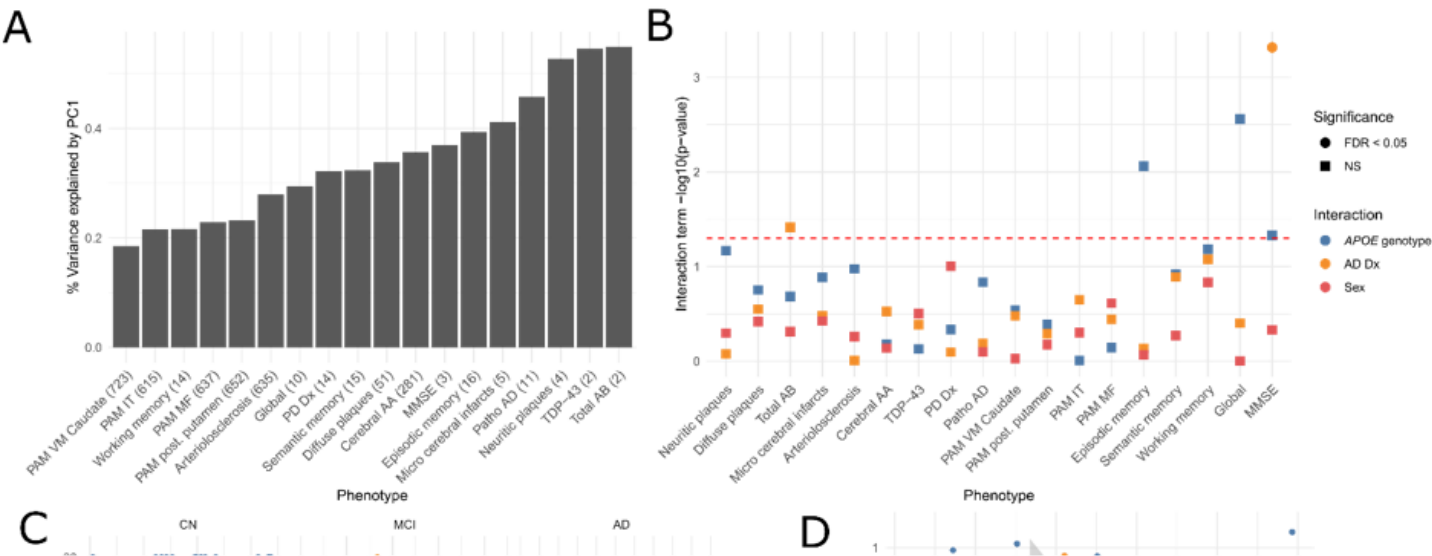
Figure 5

Interaction effects of monocyte gene expression with APOE genotype, clinical AD stage, and biological sex. A) Barplot showing the proportion of variance (y-axis) explained by the top principal component (PC1) for each set of significant differentially expressed genes for each phenotype (number of significant genes entered into PCA shown in brackets on x-axis). B) Dotplot showing the significance of the interaction term between PC1 and either APOE e4 genotype status, AD diagnosis (Dx), or biological sex, in regression models of each cognitive and pathological outcome. All p−values are two-sided. FDR correction was applied across all phenotypes and interaction models, with significance (corrected p<0.05) indicated by point shape. C) Scatterplot showing the association between PC1 of MMSE-associated monocyte gene expression residuals (x-axis) and MMSE score at time of blood draw, faceted by AD diagnosis stage (interaction p_{raw}=2.5x10^{-4}, p_{FDR}=0.01). D) Scatterplot showing the interaction between PC1 of global cognition-associated monocyte gene expression residuals and APOE genotype, in association with global cognition at time of blood draw (interaction p_{raw}= 2.8x10^{-3}, p_{FDR}=0.15). AA=amyloid angiopathy; AB=total amyloid-beta; Dx=diagnosis; Global=global cognitive summary score; IT=inferior temporal cortex; MF=midfrontal cortex; MMSE=mini mental status exam; NFT=neurofibrillary tangles; PAM=proportion of active microglia; PD=Parkinson's disease; TDP-43=tar-DNA binding-43 proteinopathy; VM=ventral medial.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementv5.docx