Aged brain multi-omic integration captures immunometabolic and sex variation

Justin P. Whalley*[1], Holly C. Hunsberger[2], David A. Bennett[3], Melissa Lamar[3], Daniel A. Peterson[2,4] and Grace E. Stutzmann[2,4]

[1]Center for Cancer Cell Biology, Immunology, and Infection, The Chicago Medical School, Rosalind Franklin University of Medicine and Science, North Chicago, IL, USA
[2]Center for Neurodegenerative Disease and Therapeutics, The Chicago Medical School, Rosalind Franklin University of Medicine and Science, North Chicago, IL, USA
[3]Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, USA
[4]Center for Stem Cell and Regenerative Medicine, Rosalind Franklin University of Medicine and Science, North Chicago, IL, USA
*Correspondance: justin.whalley@rosalindfranklin.edu

SUMMARY

Resolving the molecular basis for heterogeneous aging in the human brain requires integrating its diverse molecular layers, including the transcriptional, epigenetic, and proteomic states. To provide such a view, we applied a tensor decomposition framework to jointly analyze single-nucleus RNA-seq, DNA methylation, histone acetylation, and proteomic data from 276 postmortem human dorsolateral prefrontal cortex samples from the ROSMAP aging cohort. Our analysis revealed two dominant themes defined by components with strong multi-omic coherence. First, we identified a robust immunometabolic axis characterized by microglial activation, suppression of PI3K-Akt-mTOR signaling, and epigenetic signatures suggesting the repurposing of developmental transcription factors. Second our analysis resolved the effect of sex into a mosaic of distinct, cell-type-specific molecular programs within glia. These included male-biased mitochondrial programs in microglia, a shift toward precursor-like states in the female oligodendrocyte lineage, and bidirectional epigenetic remodeling in astrocytes. This work provides a high-resolution atlas of glial aging, demonstrating that broad risk factors manifest as complex, cell-specific vulnerabilities; a critical insight

for developing targeted therapeutic strategies.

INTRODUCTION

Aging of the human brain, and the development of neurodegenerative diseases such as Alzheimer's disease (AD), is accompanied by intertwined alterations in gene expression, epigenetic regulation, and protein homeostasis that unfold across diverse cell types (Iturria-Medina et al., 2022). Recent large-scale initiatives have generated richly annotated, multi-omic resources from postmortem human cortex, most notably the Religious Orders Study and Memory and Aging Project (ROSMAP) (Bennett et al., 2018). These datasets create an unprecedented opportunity to integrate transcriptomic, epigenomic and proteomic layers and obtain a systems-level view of cognitive decline and dementia progression.

However these omic layers are often studied in isolation, leaving the complex interplay between them poorly understood. A key challenge is to move beyond correlational studies and develop a more causal understanding of how changes in the epigenome, for example, propagate to alter transcription and protein function within specific cell types (Gabitto et al., 2024). To meet this challenge, computational frameworks capable of jointly modeling heterogeneous data are required. Tensor decomposition methods are particularly well-suited for this task, as they can deconstruct multi-modal datasets into a set of coherent, underlying biological components or programs. We and others have successfully used this approach to uncover systems-level patterns in other complex human diseases, demonstrating its power to find novel biological insights amidst high-dimensional data (COvid-19 Multi-omics Blood ATlas (COMBAT) Consortium, 2022; Mi et al., 2024).

This integrative approach is critically needed to dissect the roles of glia in brain aging. Microglia and astrocytes, once seen as passive support cells, are now recognized as central drivers of neuroinflammation and metabolic reprogramming in AD (De Strooper & Karran, 2016). Epidemiologically, women are at higher risk of developing AD, yet the cellular and molecular basis for this vulnerability remains poorly defined (Buckley et al., 2023). Understanding how sex shapes glial states at the transcriptional and epigenetic level is a major gap

in the field (Casaletto et al., 2022). Underlying these glial states are fundamental homeostatic processes, including the maintenance of myelin by oligodendrocytes (Behrendt et al., 2013; Perlman et al., 2020; Adams et al., 2021) and the precise regulation of intracellular calcium signaling (Stutzmann, 2007; Mustaly-Kalimi et al., 2022, 2025), both of which are known to be dysregulated in AD.

Here, using integrative analysis of transcriptomic, epigenomic and proteomic data, we reveal two dominant features of the aged brain. First, we characterize a major immunometabolic axis driven by microglial activation, the suppression of PI3K-Akt-mTOR signaling, and the apparent repurposing of developmental epigenetic programs. Second, we resolve the broad effect of sex into a mosaic of distinct, cell-type-specific molecular programs within glia, including male-biased mitochondrial programs in microglia and a shift toward precursor-like states in the female oligodendrocyte lineage. This work provides a high-resolution map of how risk factors like age and sex manifest as complex, cell-specific vulnerabilities.

RESULTS

We aimed to explain variation in the aging brain, through integration of multi-omic data from the Religious Orders Study and Memory and Aging Project (ROSMAP) (Bennett et al., 2018). Our primary dataset comprised single-nucleus RNA sequencing (snRNA-seq) from 276 prefrontal cortex samples, ensuring a minimum of 100 cells per subtype across astrocytes, oligodendrocytes, inhibitory neurons, oligodendrocyte precursor cells (OPCs), excitatory neurons, and microglia (Mathys et al., 2023). This transcriptomic data was complemented by three additional omic layers: DNA methylation at 51,798 CpG sites from (216 dorsolateral prefrontal cortex samples) (De Jager et al., 2014), chromatin immunoprecipitation sequencing (ChIP-seq) of H3K9ac histone marks at 26,384 peaks (193 prefrontal cortex samples) (Klein et al., 2019), and mass-spectrometry proteomics quantifying 5,230 proteins (105 dorsolateral prefrontal cortex samples) (Ping et al., 2018). The integration was anchored on the 276 samples with snRNA-seq data, as this cohort had the largest sample size and provided the cellular resolution necessary for our analysis. This approach maximized sample overlap

across all four omic types.

To unify these heterogeneous molecular layers, we applied Sparse Decomposition of Arrays (SDA), a Bayesian tensor factorization method that we adapted for this research (Hore et al., 2016). We chose SDA for its ability to reduce dimensionality across all four modalities and handle missing data, which is critical for complex biological systems. Our decomposition of this integrated dataset resulted in 546 latent components, each capturing a coherent axis of biological variance across omic types (Figure 1). This approach allowed us to resolve both cell-specific programs and cross-modal regulatory circuits.

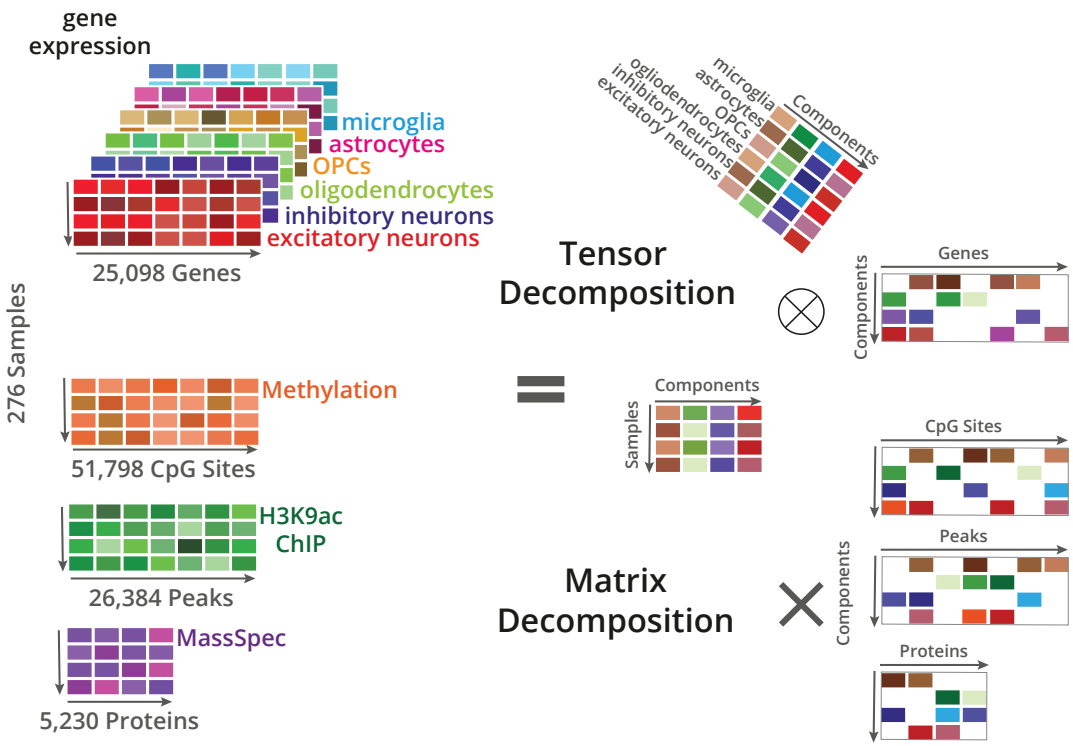

Figure 1. Schematic of work: tensor and matrix decomposition applied to the ROSMAP datasets. The left panel shows the input data: a three-dimensional gene expression tensor (samples × genes × cell types) and three two-dimensional matrices for DNA methylation (CpG sites), histone acetylation (ChIP-seq peaks), and proteomics (proteins), all aligned by sample. The right panel depicts the output of integrative decomposition using Sparse Decomposition of Arrays (SDA), which yields a shared set of latent components. Each component is represented by a vector of sample loadings, combined via tensor multiplication with cell type and gene loadings for the RNA-seq data, and via matrix multiplication with sparse feature loadings for the methylation, ChIP-seq, and proteomic data.

As an initial validation of the method's ability to uncover novel biological signals within our data, SDA revealed a previously unreported eQTL regulating a cluster of olfactory receptor genes, with the signal being driven predominantly by neuronal and oligodendroglial cells (component 21 as shown in Figure S1). This finding not only validates our analytical approach

but also highlights the complex regulatory landscape of ectopic gene expression in the brain, as olfactory receptors have been implicated in diverse cellular processes beyond olfaction (Maßberg & Hatt, 2018).

From the 546 components, we first generated a comprehensive annotation, summarized in Supplementary Table 1, that details each component's primary cell-type drivers, top multi-omic features, pathway enrichments, and significant associations with clinical phenotypes (such as cognitive decline and biological sex). Guided by this overview, we then prioritized a subset of components for detailed characterization in the main text based on their strong multi-omic coherence and robust cross-cohort reproducibility in the SEA-AD snRNA-seq dataset from 82 dorsolateral prefrontal cortex samples (Gabitto et al., 2024). The following sections are organized around two key themes that emerged from this filtered set: the dissection of complex immunometabolic axes within glial cells, and the resolution of sex differences into multiple, distinct, and cell-type-specific signatures. All component loading scores and the code used for this analysis are publicly available on Synapse and GitHub, respectively.

Component 351: A microglial immunometabolic axis in the aged brain

Integrative tensor decomposition of our multi-omic dataset revealed an immunometabolic program captured by Component 351. Figure 2A displays the loading scores for each major cell type in the snRNA-seq data. Notably, microglial transcripts associated with immune-inflammatory responses, such as *IGKC*, *IGHG1*, *CD68*, and *HLA-DRB1*, show strong contributions to this component (Figure 2B), suggesting robust microglial activation in the aged brain (Figure S2A).

Protein-level analysis (Figure 2C) reinforces the same pro-inflammatory direction, featuring multiple transporters and antioxidant proteins (e.g., *CUTA*, *SERINC3*, *SEP15*) that may reflect metabolic reprogramming under immune stress. In contrast, the subset of proteins exhibiting negative loadings (Figure 2D) is significantly enriched for PI3K-Akt-mTOR pathway members (e.g., *AKT3*, *PTEN*, *PDPK1*, *MAP2K1*), implying that microglial activation could coincide with suppressed growth signaling or cell-type loss within these aged samples.

Methylation profiling (Figure 2E) adds a further dimension, highlighting CpG loci near *AKT1*

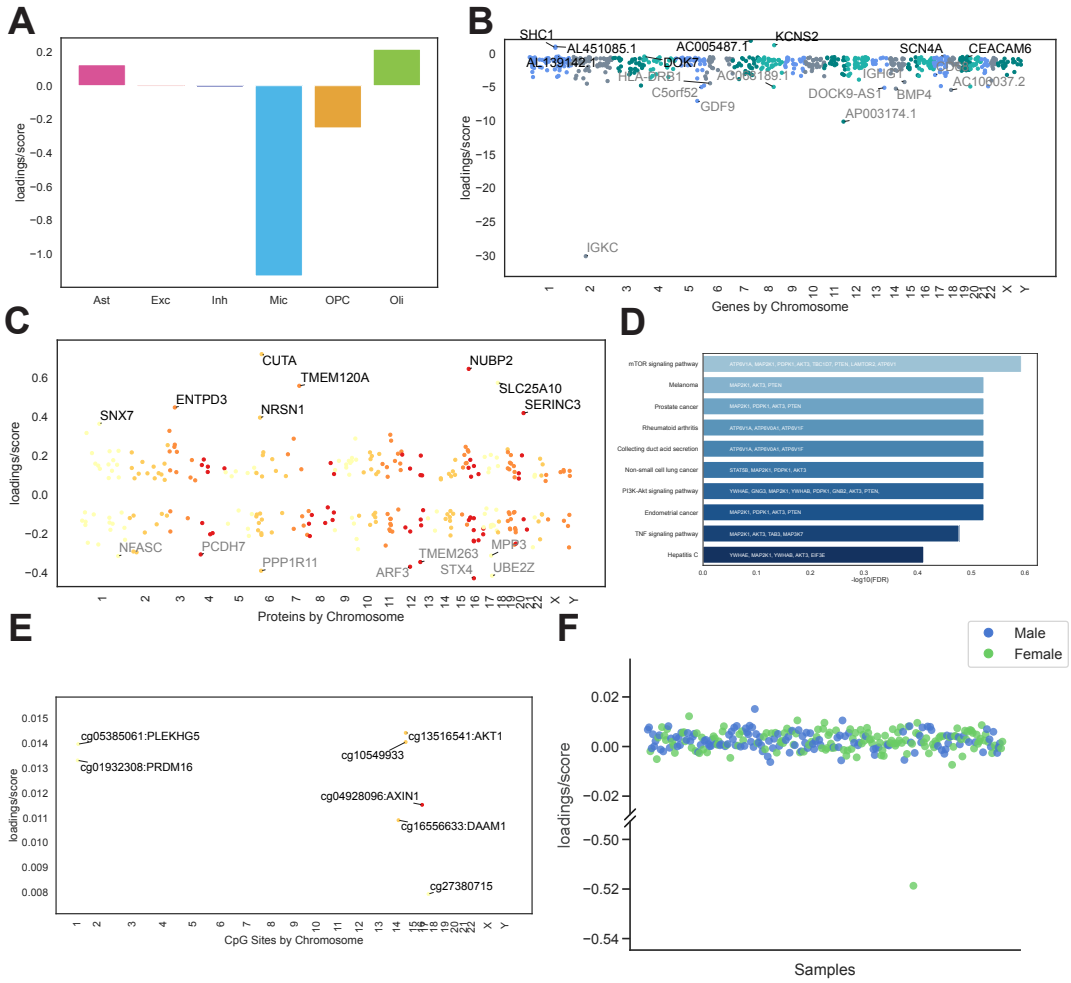

Figure 2. A microglial immunometabolic program in the aged brain (Component 351). **(A)** Cell-type loadings highlight strong microglial contributions to this component. **(B)** Top gene loadings from snRNA-seq show elevated expression of immunoglobulin and antigen-presentation transcripts (e.g., *IGKC*, *IGHG1*, *CD68*, *HLA-DRB1*), consistent with immune activation. **(C)** Positively loaded proteins include antioxidant and stress response factors (e.g., CUTA, SERINC3, SEP15), while negatively loaded proteins reflect downregulation of PI3K-Akt-mTOR signaling (e.g., AKT3, PTEN). **(D)** KEGG pathway enrichment for negatively loaded proteins shows suppression of metabolic and growth pathways, including mTOR, PI3K-Akt, and TNF signaling. **(E)** CpG sites near immune and metabolic genes (e.g., *AKT1*, *DAAM1*, *PLEKHG5*) show consistent loading directions, supporting epigenetic regulation of this axis. **(F)** Sample-level scores across individuals, with one outlier showing markedly high loading score, potentially indicating a localized or atypical inflammatory process.

and other metabolic or immune genes that load in the same pro-inflammatory, low-mTOR direction. Additional evidence comes from H3K9ac ChIP-seq peaks (Figure S2B) enriched for a *BARHL1*-like motif. Although *BARHL1* is best known for its role in neurodevelopment and cerebellar formation (Lopes et al., 2006), our data imply a possible re-engagement of this transcriptional network in older brains under high inflammatory states, raising questions about BARHL1's function or its related homeobox factors in chronic neuroinflammation or neural remodeling. The re-engagement of such a developmental transcription factor in the aged brain could signify a maladaptive or incomplete attempt at neural remodeling or a cellular identity shift in response to persistent immune stress.

Parallel analysis in the independent SEA-AD dataset revealed a highly, absolute correlated transcriptional signature (Component 21) that marks a principal immune axis in aged cortex (Figure S2C–E). This correspondence validates the microglial immune program captured by ROSMAP Component 351. In both cohorts, a few outlier samples display exceptionally high *IGKC* expression (Figure 2F; Figure S2E), consistent with sporadic B-cell-driven inflammation. Importantly, although SEA-AD Component 21 appears as a single transcriptional component, our multi-omic decomposition in ROSMAP resolves it into two discrete biological programs: Component 351, an immunometabolic axis, and Component 2, an immune-regulatory axis discussed next.

Component 2: A glial immune-regulatory axis balancing inflammation and homeostasis

Component 2 is characterized transcriptionally by a strong microglial signature whose leading feature is the immunoglobulin gene *IGKC* (Figure 3A–B). This signature is also correlated with SEA-AD Component 21 (Figure S3B). Unlike Component 351, however, Component 2 is defined by unique multi-omic features and cell-type drivers: its hallmark is a pronounced inverse relationship between microglial/B-cell activation and astrocytic programs (Figure 3A). Thus, Component 2 reflects a reciprocal glial state in which immune activity rises while astrocytic pathways are selectively suppressed, and vice versa, rather than a uniform glial activation.

DNA methylation patterns add a further layer to this glial axis. Several CpG sites (e.g.,

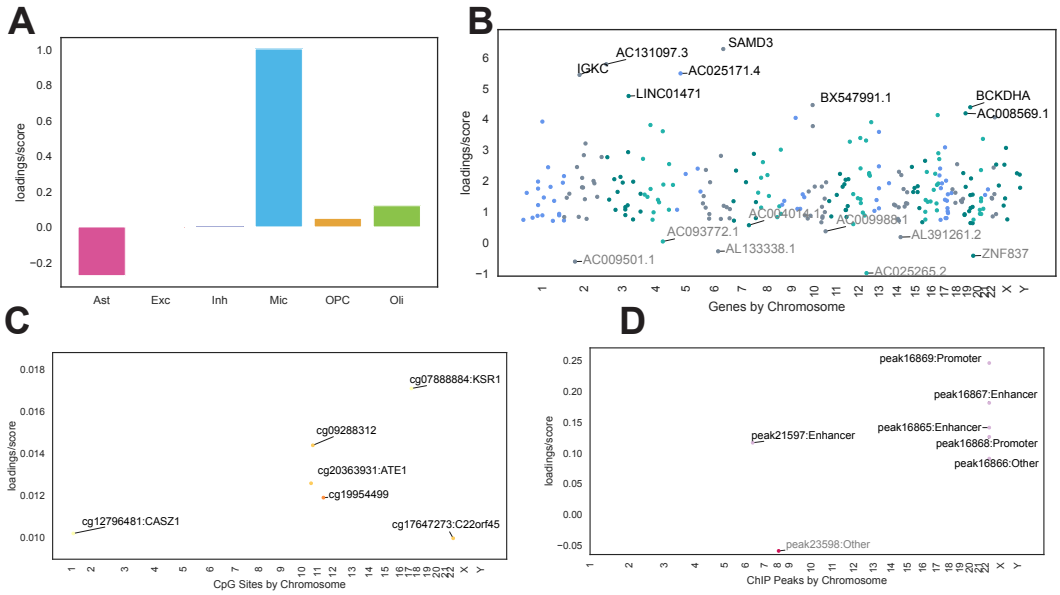

Figure 3. Multi-omic features and regulatory annotations for Component 2 in the aging brain. **(A)** Cell-type loadings from snRNA-seq data, highlighting positive microglial and negative astrocytic contributions to Component 2. **(B)** Top gene loadings from snRNA-seq, highlighting non-coding transcripts among the top loadings. **(C)** Methylation sites with high PIP values for Component 2, annotated to genes including *KSR1*, *CASZ1*, and *ADORA2A-AS1*. **(D)** H3K9ac ChIP-seq peak loadings, showing a cluster of enhancer regions on chromosome 22 overlapping the *CACNA1I* and *ENTHD1* locus. These epigenomic signals support the role of purinergic signaling in this glial axis.

cg07888884 near *KSR1*, cg12796481 near *CASZ1*, and cg17647273 in *ADORA2A-AS1*) (labelled as C22orf45) are differentially methylated in association with Component 2 loadings (Figure 3C). These loci map to signaling and metabolic regulators implicated in glial reactivity and inflammation resolution. Chromatin accessibility data (ChIP-seq) reveal a cluster of active enhancer peaks on chromosome 22 (Figure 3D), overlapping *CACNA1I*, which encodes a T-type voltage-gated calcium channel, and *ENTHD1*. Calcium signaling is a fundamental mechanism of communication for both neurons and glial cells; suggesting that immunoglobulin-associated signal is functionally coupled to alterations in the brain's core calcium signaling machinery.

Together, these features suggest that Component 2 captures a dynamic glial state characterized by immune engagement from microglia and either compensatory suppression or reactivity in astrocytes. This is further supported by the prevalence of uncharacterized or non-coding transcripts among top RNA-seq loadings, hinting at a novel, regulatory-rich axis of aging brain biology (Figure 3B). In parallel to Component 351, which reflects a coordinated

upregulation of immune and metabolic genes, Component 2 appears to mark an intrinsic regulatory balance between glial cell types, pointing to variation in how aging brains manage immune stress.

Components 74 and 177: Resolving sex differences in neurons and microglia

Our decomposition resolved multiple axes of sex-biased molecular variation. Component 74 showed expected patterns (on X- and Y-chromosome) for gene expression in the neurons and ChIP peaks (Figure S4A-E), and highly correlated with a sex specific component in SE-AD dataset (Figure S4F) and similar to sex specific components seen in our previous work in blood atlas (COvid-19 Multi-omics Blood ATlas (COMBAT) Consortium, 2022). However, unique to this analyses, are the different patterns shown in microglia and oligodendrocytes.

Component 177 reflects a distinct program with a strong microglial signature that is significantly elevated in males (Mann-Whitney U FDR-adjusted $p = 0.0253$) (Figure 4A, B). The transcriptional signature of this component is defined by a coherent and powerful set of genes related to mitochondrial bioenergetics (Figure 4C). The most strongly associated transcript is *NDUFA13*, an essential subunit of Complex I of the mitochondrial electron transport chain. This is complemented by another key hit, *COX6A1*, a subunit of Complex IV. The joint association with these two critical oxidative phosphorylation genes suggests this component captures a sex-specific difference in the core respiratory machinery of cells. This bioenergetic signature is further linked to organelle function through genes like *ARL6IP1*, which is involved in ER biology and stress responses. This pattern is similarly seen in SE-AD Component 85 (Figure 4D). At the protein level, the component's top hit is Metadherin (MTDH), a protein also implicated in cell metabolism and stress (Figure 4E).

This component's multi-omic signature is reinforced by a coordinated program of DNA methylation changes (Figure 4F). The epigenetic signature points directly to the regulation of cellular quality control, providing a link to the mitochondrial bioenergetics theme seen in the transcriptome. Specifically, we identified changes in methylation at key genes in the autophagy-lysosomal pathway, including the autophagosome initiator *ATG9B* and the lysosomal enzyme Cathepsin Z (*CTSZ*), both more likely to be hypermethylated in females. *CTSZ* is part of

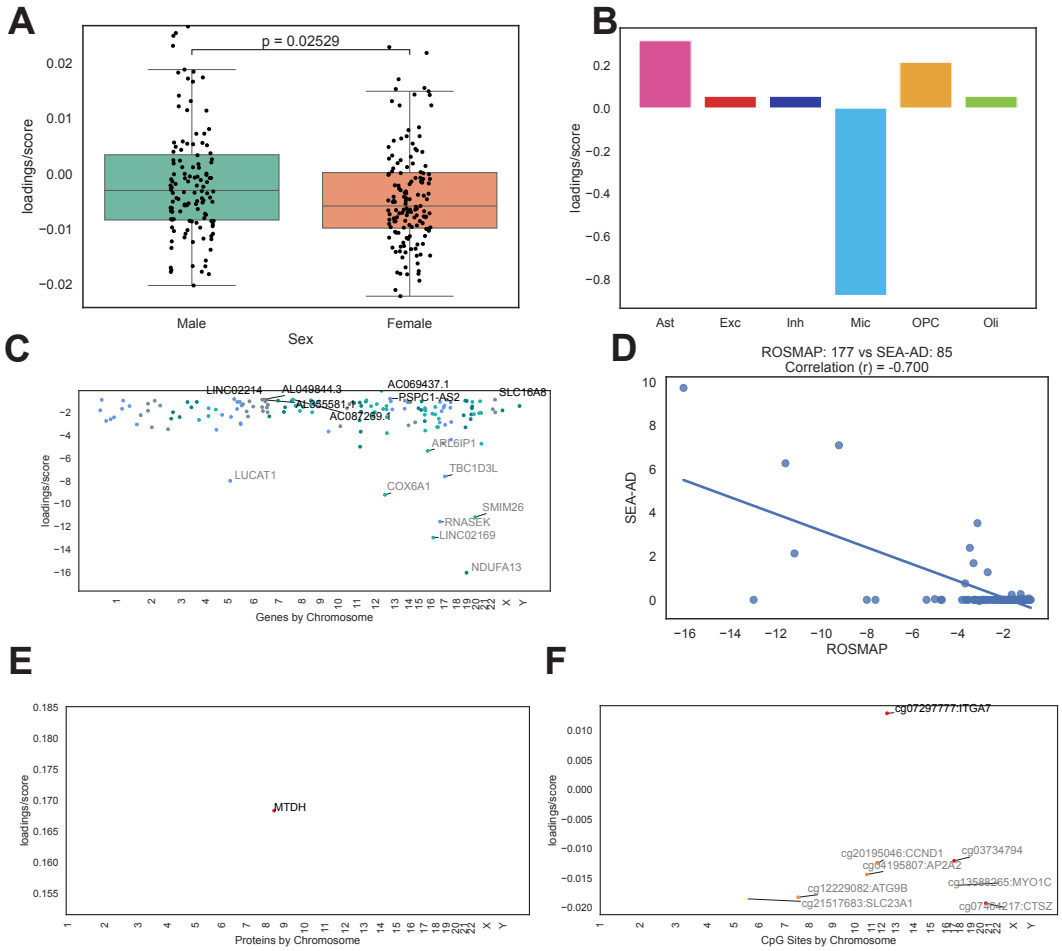

Figure 4. Microglial-specific sex-difference components capture metabolism and microglial inflammation. **(A)** Box plot showing significantly higher sample scores for Component 177 in male individuals (Mann-Whitney U test adjusted for multiple testing). **(B)** Cell type loading scores for Component 177, highlighting the difference specifically for microglia. **(C)** Gene expression loading scores for Component 177, noting mitochondrial *NDUFA13* and respiratory genes *COX6A1*. **(D)** Scatter plot showing a negative Spearman's correlation coefficient between ROSMAP Component 177 and a corresponding component 85 in the SEA-AD dataset, showing this pattern for gene expression in an independent gene expression dataset. **(E)** Protein MTDH is linked to this Component 177, with abundance generally higher in males. **(F)** Methylated CpG Sites loading scores for Component 177, highlighting key genes in the autophagy-lysosomal pathway (*ATG9B*, *CTSZ*) contrasted by *ITGA7* in a different direction.

protein clearance and so less expression may result in susceptibility to dementia.

Intriguingly, this pattern is contrasted by the methylation of the component's lead CpG site in the different direction (hypomethylated in females), located near the integrin gene *ITGA7*. This opposing epigenetic regulation is biologically insightful: it suggests a coordinated cellular reprogramming where genes for autophagy, motility (*MYO1C*), and oxidative stress management (*SLC23A1*) are activated, while genes for a specific type of cell-matrix adhesion (*ITGA7*) are simultaneously repressed, and vice versa. This may reflect a functional shift in microglia from a more stationary, surveying state to a more motile and metabolically active phenotype across our cohort.

Components 278 and 87: Sex-driven variability in oligodendrocytes and astrocytes

Component 278 captures a sex difference within the oligodendrocyte lineage. Sample scores for this component were significantly higher in females (Mann-Whitney U FDR-adjusted $p = 1.5 \times 10^{-46}$) (Figure S5A). The cell-type loadings reveal that this component does not simply represent oligodendrocyte activity, but rather a dynamic axis of opposition between oligodendrocyte precursor cells (OPCs), which contribute positively, and mature oligodendrocytes, which contribute negatively to this component (Figure S5B). This suggests that a higher score, as seen more frequently in females, reflects a cellular state that is transcriptionally more OPC-like and less mature. The biological basis of this axis is underscored by the enrichment of genes involved in canonical growth and differentiation pathways, including Ras, PI3K-Akt, and MAPK signaling (Figure S5E). Key individual genes driving this component include the oligodendrocyte differentiation factor *SOX6* and the stress-responsive gene *FKBP5* (Figure S5C). This entire sex-biased transcriptional program was found to be conserved, showing a correlation with a similar component in the independent SEA-AD cohort (Figure S5D).

Component 87 captures a complex, multi-glial program with a significant sex bias, with sample scores are notably higher in males (Mann-Whitney U FDR-adjusted $p = 0.0017$) (Figure 5A). While the component is most strongly driven by astrocytes, it also reflects positive contributions from oligodendrocytes and microglia, which are contrasted by a strong negative loading from oligodendrocyte precursor cells (OPCs) (Figure 5B). This suggests a coordinated

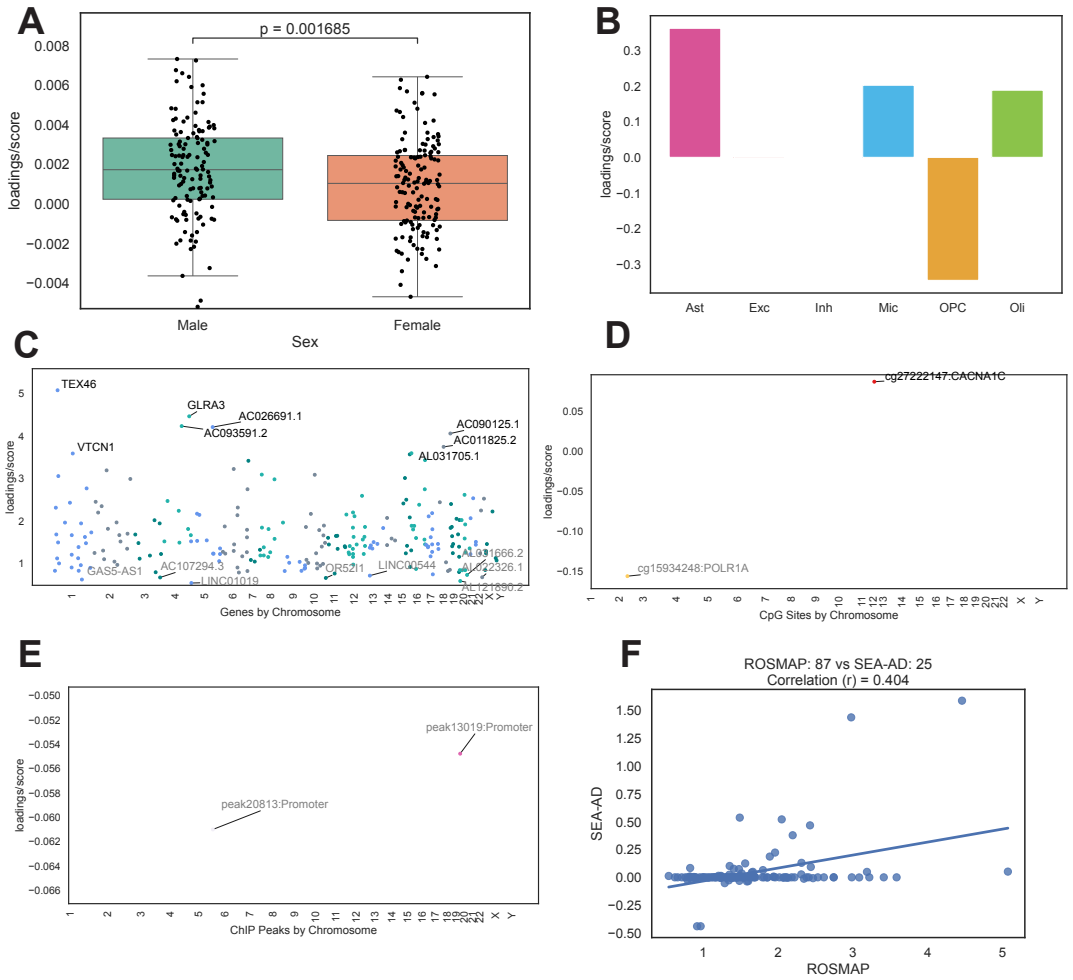

Figure 5. Astrocyte-driven component links immune modulation to coordinated epigenetic reprogramming. **(A)** Box plot showing significantly higher sample scores for Component 87 in male individuals (Mann-Whitney U test adjusted for multiple testing). **(B)** Cell-type loading scores for Component 87, highlighting a strong positive contribution from astrocytes (Ast) and an opposing negative contribution from oligodendrocyte precursor cells (OPC). **(C)** Gene expression loadings for Component 87, with top hits including the immunomodulatory gene *VTCN1*. **(D)** DNA methylation loadings for Component 87, showing opposing regulation between a CpG site in CACNA1C (positive loading) and one near POLR1A (negative loading). **(E)** ChIP-seq peak loadings for Component 87, showing decreased accessibility at specific promoter sites. **(F)** Scatter plot showing a positive Spearman's correlation coefficient between ROSMAP Component 87 and a corresponding component in the SEA-AD dataset, confirming the program's replicability.

glial state, rather than the action of a single cell type. The transcriptional signature of this male-biased state is enriched for immunomodulatory genes, with a top hit being *VTCN1* (also known as *B7-H4*), a key negative regulator of T-cell responses (Figure 5C).

The component's multi-omic signature is defined by a coordinated program of opposing epigenetic changes. In the DNA methylation data, the male-biased state is associated with hypermethylation (positive loading) of a CpG site within the neuronal calcium channel gene *CACNA1C*. This is directly contrasted by hypomethylation (negative loading) at a site near *POLR1A*, a core component of the RNA polymerase I machinery (Figure 5D). This theme of coordinated epigenetic control is strikingly echoed in the chromatin accessibility data. Here, the male-biased state is linked to decreased accessibility (negative loading) at the promoter peaks for the receptor tyrosine kinases *AXL* and *FLT4* (Figure 5E). *AXL* is a central regulator of glial phagocytosis and innate immune suppression, while *FLT4* is critical for glymphatic waste clearance. The concurrent epigenetic poising for repression of these two key glial maintenance and immunomodulatory genes provides a powerful mechanistic basis for the male-biased state. Finally, we see this transcriptomic pattern, a positive correlation, in a corresponding component in the independent SEA-AD dataset (Figure 5F).

DISCUSSION

Our integrative multi-omics analysis reveals previously unappreciated molecular programs that emerge when transcriptomic, epigenomic, and proteomic layers from the aging human cortex are examined in concert. Using Sparse Decomposition of Arrays (SDA), we distilled these layers into two overarching themes: a glial immuno-metabolic axis driven by microglial activation, and a mosaic of pervasive, cell-type-specific sex differences across glial lineages. These findings provide a high-resolution map of how risk factors like age and sex manifest as complex, cell-specific vulnerabilities that are coordinated across multiple biological layers.

A feature of the aged brain was a tightly coordinated immunometabolic program linking pro-inflammatory gene and protein expression with the suppression of PI3K-Akt-mTOR signaling. This coupling supports the concept of inflammaging, where persistent, low-grade immune

activity and metabolic dysregulation create a self-reinforcing loop (Franceschi et al., 2018). Interestingly, our finding of suppressed mTOR signaling in the human cortex presents a nuanced contrast to studies in aging female mice, which report heightened mTOR signaling as a driver of microglial glycolysis (Kang et al., 2024). This apparent discrepancy may reflect fundamental differences between species, brain regions, or, perhaps most likely, the distinction between adaptive metabolic shifts in aging versus the chronic, potentially exhausted or maladaptive state of glia in the context of human neurodegenerative pathology.

An unexpected feature of this immunometabolic axis was the enrichment of BARHL1-like motifs in H3K9ac-marked chromatin. As BARHL1 is typically restricted to neurodevelopment, its re-emergence suggests that under prolonged stress, aging glia may re-engage latent developmental programs. This convergence of immune and developmental pathways aligns with a growing understanding of cellular plasticity in disease, where cells can adopt dysfunctional or hybrid identity states (De Strooper & Karran, 2016). This raises the possibility that targeting the epigenetic machinery that governs these latent programs could be a novel strategy to reset maladaptive glial states in late life.

Beyond this shared immunometabolic axis, our analysis resolved the influence of sex into a mosaic of distinct, cell-type-specific molecular programs. In microglia, males exhibited a bias toward mitochondrial bioenergetic programs, suggesting a sex-specific vulnerability in how these crucial immune cells manage energy and oxidative stress. In contrast, the female oligodendrocyte lineage showed a shift toward a more precursor-like transcriptional state, which could have significant implications for myelin dynamics and repair capacity in the female brain (Mathys et al., 2019; Akay et al., 2021; Lopez-Lee et al., 2024). Furthermore, the male-biased astrocyte component was linked to epigenetic changes at the crucial calcium channel gene *CACNA1C*. This finding is particularly striking, as *CACNA1C* is a well-established risk gene for neuropsychiatric conditions, including mood disorders. Our observation of a male-biased epigenetic signature in the aging cortex aligns remarkably with recent findings, where chronic stress altered *Cacna1c* expression specifically in the prefrontal cortex of male mice (Bastos et al., 2024). This provides a powerful, cell-type-specific link to the long-standing

"calciumopathy" hypothesis of AD (Stutzmann, 2007) and suggests that *CACNA1C* may be a critical node where sex-specific vulnerability to both chronic stress and age-related pathology converge.

These findings have therapeutic implications. The immunometabolic axis nominates mTOR and its upstream regulators as tractable entry points for modulating chronic neuroinflammation. Crucially, the highly specific nature of the sex-dimorphic components argues that a one-size-fits-all approach to treating neurodegenerative disease is likely to be insufficient. Precision therapeutics may need to be tailored not only to a patient's disease subtype but also to their sex, targeting, for example, mitochondrial function in male microglia or myelin maintenance pathways in females. The replication of our key findings in the independent SEA-AD cohort (Gabitto et al., 2024) demonstrates that these programs are fundamental features of human brain aging, not artifacts of a single dataset. By unifying multiple molecular layers, our work moves beyond catalogues of single-omic changes to reveal the coherent biological themes that orchestrate cellular function in the aged brain, providing a critical framework for guiding the next generation of precision interventions.

Limitations of the study

Our study has several limitations that provide context for our findings. First, there is an inherent imbalance in the depth and sample size of the integrated data modalities. The single-nucleus RNA-seq data, with its cellular resolution and largest sample cohort, necessarily provides the richest source of variation. In contrast, the proteomics data had the fewest features and samples. Consequently, while our approach integrates all layers, the resulting components are naturally weighted towards the more deeply characterized transcriptomic layer.

Second, this weighting extends to our validation strategy. The independent SEA-AD cohort provided an excellent resource for replicating the transcriptional signatures of our key components, strongly supporting their generalizability. However, as this dataset lacks matched multi-omic data, it provides indirect, rather than direct, validation for the specific epigenetic and proteomic patterns we identified.

Third, our results focus predominantly on glial-driven programs. This focus arose because

the most robust multi-omic signals, those with coherent patterns across multiple molecular layers, were found in glial cell types. Our analysis did resolve neuron-specific components, many of which replicated known associations between neuronal gene expression and cognitive phenotypes. For example, we identified a purely inhibitory neuron component (Component 172) enriched for MAPK signaling (Figure S6A-D) that was associated with Alzheimer's diagnosis (Munoz & Ammit, 2010; Kim & Choi, 2010), and an excitatory neuron component (Component 277) enriched for GABAergic synapse genes that was associated with cognitive decline (Figure S6E-H) (McQuail et al., 2015; Rozycka & Liguz-Lecznar, 2017). However, as these largely recapitulated findings from single-modality studies, we prioritized the novel, integrative glial programs that represent the unique strength of this multi-omic approach.

Finally, like all studies based on postmortem tissue, our analysis provides a static snapshot of the aged brain. While we identify robust molecular states and their associations, we cannot directly capture the dynamic cellular transitions that occur over the course of aging and disease progression. Future work using longitudinal data or advanced model systems will be needed to dissect the causal relationships and temporal dynamics of the multi-omic programs identified here.

METHODS

Data sources

Data from the Religious Orders Study and Memory and Aging Project (ROSMAP) served as the primary dataset for this study (Bennett et al., 2018). ROSMAP consists of two longitudinal cohorts based at the Rush Alzheimer's Disease Center (RADC). All participants enroll without known dementia and agree to annual evaluation and organ donation. Both cohorts were approved by the Institutional Review Board of Rush University Medical Center and all participants signed informed and repository consents and an Anatomic Gift Act. Here, processed data modalities were downloaded from the AD Knowledge Portal[1]. Specifically, this included: single-nucleus RNA-sequencing (snRNA-seq) data (syn52293417) (Mathys et al., 2023); DNA methylation data (syn4896408) (De Jager et al., 2014); H3K9ac chromatin immunoprecipitation sequencing (ChIP-seq) data (syn3157275) (Klein et al., 2019); and mass-spectrometry proteomics data (syn17008935) (Ping et al., 2018). Our analysis is centered on 276 of these participants, a cohort that is well-balanced for sex (147 female, 129 male) and represents a full spectrum of cognitive function at the time of death, including 173 individuals with no cognitive impairment and 103 with a diagnosis of dementia.

For independent validation of our transcriptomic findings, we used data from the Seattle Alzheimer's Disease Brain Cell Atlas (SEA-AD) (Gabitto et al., 2024). Processed single-nucleus RNA-sequencing count matrices and associated metadata were downloaded from the Allen Brain Map portal[2] for the dorsolateral prefrontal cortex.

For all datasets, we utilized the publicly available, processed data matrices to the greatest extent possible. Feature names (e.g., gene IDs, CpG loci identifiers) were retained from the original sources to ensure consistency. Minor pre-processing and harmonization steps required to integrate the different data types are detailed in the following sections.

Data processing and harmonization for joint decomposition

All four data modalities were processed and harmonized to a final cohort of 276 individuals defined by the snRNA-seq dataset. Each matrix was aligned to have the same 276 sample

[1] https://adknowledgeportal.synapse.org
[2] https://portal.brain-map.org/explore/seattle-alzheimers-disease

columns in the same order, with missing data for any individual in any modality represented as `NaN` values, a format handled by the SDA model.

Single-Nucleus RNA-Sequencing (snRNA-seq): Processing began with the pre-calculated AnnData object from the original ROSMAP study. After excluding two low-quality samples and all computationally-predicted doublets, we filtered for samples containing at least 100 cells in each of the six major glial and neuronal cell types. For the final 276 samples, cell-type-specific pseudobulk profiles were generated by summing raw counts for each cell type. A gene filter was applied to retain genes with at least 10 counts in 10 or more pseudobulk samples. The final normalized expression data was structured as a three-dimensional tensor of (276 samples) $\times$ (6 cell types) $\times$ (25,098 genes), with no missing samples. The validation SEA-AD dataset was processed in a parallel manner, resulting in a tensor of (82 samples) $\times$ (10 cell types) $\times$ (31,648 genes).

H3K9ac ChIP-Sequencing (ChIP-seq): The raw read count matrix was harmonized to match the projids of the 276-sample snRNA-seq cohort. Data was normalized to log2-transformed Counts Per Million (log2(CPM+1)). The final input matrix had dimensions of (276 samples) $\times$ (26,384 peaks), which included 83 samples with no ChIP-seq data (represented as missing values).

DNA methylation. We began with the ROSMAP imputed β-value matrix and harmonized it to the 276 individuals profiled by snRNA-seq. Each β value was temporarily binarized as methylated (1; $\beta \geq 0.5$) or unmethylated (0; $\beta < 0.5$), preserving `NaN` for missing measurements. CpG sites that were invariant after binarization (all 0, all 1, or all missing) across the cohort were discarded, thereby removing constitutively methylated or unmethylated loci. The surviving probes retained their original continuous β values for downstream analysis, yielding a matrix of 51,798 variable CpG sites across 276 samples; with 60 missing samples.

Proteomics: We used the fully processed log2-transformed TMT proteomics data. After harmonizing sample identifiers to the 276 projids, we applied a stringent feature filter, removing any protein with a missing value in any of the present samples. The final input matrix had dimensions of (276 samples) $\times$ (5,230 proteins), with 171 missing samples.

Tensor and matrix decomposition

To integrate the four processed omic datasets, we applied Sparse Decomposition of Arrays (SDA), a Bayesian framework for tensor and matrix decomposition (Hore et al., 2016). The snRNA-seq data was structured as a three-dimensional tensor of (276 samples) × (6 cell types) × (25,098 genes). The DNA methylation (51,798 CpG sites), H3K9ac ChIP-seq (26,384 peaks), and proteomics (5,230 proteins) data were each structured as two-dimensional matrices of (276 samples) × (features). These arrays were jointly decomposed by SDA, which is robust to the missing data present across the different modalities (Figure 1).

To ensure the identification of robust and reproducible biological signals, we performed ten independent runs of the SDA model. Each run was initiated with a request for 2,000 components and performed for 1,500 iterations. All non-zero components from all ten runs were then pooled and subjected to a rigorous clustering and filtering procedure. Following the established protocol (Hore et al., 2016), we performed hierarchical clustering on the components based on their pairwise dissimilarity, defined as 1 minus the absolute Pearson correlation of their sample loading scores. Flat clusters were formed using a cophenetic distance threshold of 0.4. A stringent stability criterion was then applied, retaining only those clusters containing components from at least five of the ten independent runs. This process yielded a final set of 546 stable components. A final, representative loading score for each feature and sample was calculated by averaging the scores of all components within each stable cluster, after aligning their signs.

A key feature of the SDA framework is its ability to perform automatic feature selection through its use of a spike-and-slab prior. For each feature's loading score in a given component, the model calculates a Posterior Inclusion Probability (PIP), a value between 0 and 1 representing the confidence that the feature is genuinely associated with that component. For all downstream biological interpretation, we applied a stringent threshold, considering only features with a PIP > 0.5 to be robustly associated with a component.

This step is crucial for two reasons. First, it creates a sparse representation, effectively reducing the vast number of initial features to a core set of high-confidence drivers for each

biological program. Second, it means that components are not required to be active across all modalities. A significant implication of this feature selection is that many of the resulting components are driven by only one or two omic types (e.g., a purely transcriptomic signature or a coupled transcriptomic-epigenomic axis), allowing the model to capture both single-modality and truly integrative multi-omic patterns. The final loading scores reported for these selected, high-confidence features represent the expected value from the model's posterior distribution.

The interpretation of each component's effect is based on the multiplicative relationship between its loading scores. For the three-dimensional snRNA-seq data, a component's contribution to the expression of a given gene in a particular cell type for a specific sample is proportional to the product of the three corresponding loading scores (sample, cell type, and gene). Consequently, the sign of this product determines the direction of the effect: a positive product indicates that the component contributes to higher relative expression, while a negative product indicates lower relative expression. The magnitude of this contribution is, in turn, determined by the product of the magnitudes of these scores. This principle extends to the two-dimensional matrices, where a component's effect is the product of the sample and feature loading scores.

Validation in an independent cohort: To assess reproducibility, we correlated each ROSMAP component with components derived from the SEA-AD snRNA-seq dataset (dorsolateral prefrontal cortex). For a given ROSMAP component we first extracted genes with PIP > 0.5, intersected this set with the SEA-AD gene list, and assembled two loading vectors (ROSMAP vs SEA-AD) on the common genes. We then computed the Spearman correlation coefficient between every pair of vectors and recorded, for each ROSMAP component, the SEA-AD component that achieved the highest *absolute* correlation. Absolute correlation is used because the sign of SDA loadings is arbitrary: multiplying all sample and feature loadings by –1 yields an equivalent factor and leaves the reconstructed data unchanged.

Component prioritization. From the 546 ROSMAP components we selected those most likely to represent robust, integrative biology through a two-step filter.

1. Multi-omic coherence. We required high-confidence features (PIP > 0.5) in at least three

of the four modalities (gene expression, DNA methylation, H3K9ac peaks, and proteomics) thereby focusing on genuinely integrative axes.

2. Cross-cohort reproducibility. Among these coherent components we retained only those whose transcriptional signature correlated strongly with a SEA-AD component ($|r| > 0.4$ by Spearman). Because of the sign indeterminacy noted above, the magnitude of the correlation, not its direction, indicates biological concordance.

The resulting set of integrative, reproducible components forms the basis of the detailed analyses presented in the Results.

Association with phenotypes. For each component we tested whether sample loading scores differed across key clinical variables. Binary traits; biological sex (male vs. female) and neuropathological diagnosis (CERAD: AD vs. non-AD); were evaluated with two-sided Mann-Whitney U tests, restricted to components for which the smaller group comprised at least 20% of the cohort. Ordinal cognitive status (no decline, mild cognitive impairment, AD dementia) was assessed with one-way ANOVA. Resulting p-values from all 546 components were adjusted for multiple testing within each phenotype using the Benjamini-Hochberg false-discovery rate (FDR); components with FDR < 0.05 were deemed significant.

Pathway enrichment analysis: To identify enriched biological pathways for each component, we used the gseapy Python package, an interface to the Enrichr toolset (Xie et al., 2021). For a given component, we selected all features (genes, proteins, etc.) with a Posterior Inclusion Probability (PIP) > 0.5. The official gene symbols corresponding to these high-confidence features were used as the input gene list for enrichment analysis against the KEGG 2021 Human gene set. Pathways with an FDR-adjusted p-value < 0.05 were considered significantly enriched. The most significant pathway per component per omic type is reported in Supplementary Table 1.

Motif enrichment analysis. For each component, we searched for transcription-factor–binding motifs enriched in its H3K9ac peaks using the HOMER suite (v5.1). Genomic coordinates of all peaks with PIP >0.5 were exported as a BED file and passed to `findMotifsGenome.pl` (default parameters) against the GRCh37.p13 reference genome. HOMER builds a GC-

matched background set automatically. Motifs with a HOMER-reported enrichment $p < 0.05$ (binomial test) were deemed significant; the most significant motif per component is reported in Supplementary Table 1.

Plotting and visualization: All figures in this manuscript were generated in Python. Data manipulation and analysis were performed using the pandas and scipy libraries. Plots were generated using matplotlib and seaborn. The statannotations package was used to display statistical significance on box plots, and the adjustText package was used for non-overlapping label placement in scatter plots. To effectively visualize sample loading plots containing extreme outliers, we defined outliers as any data point more than 30 times the interquartile range (IQR) from the upper or lower quartile. For these specific plots, the brokenaxes package was used to create a discontinuous y-axis, allowing for the clear display of both the main data distribution and the extreme outlier values on the same figure, or not included in the boxplots.

ACKNOWLEDGMENTS

We thank Dr. Robert Marr for his insightful feedback during the preparation of this manuscript. The results published here are in whole or in part based on data obtained from the Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, and downloaded from the AD Knowledge Portal. We are deeply grateful to the participants of the Religious Orders Study and Memory and Aging Project (ROSMAP) for their invaluable contributions to this research.

This work was supported by grants from the National Institutes of Health (NIH) and the Alzheimer's Association. Specific funding for the authors includes: NIH grants NS135301 and NS135306 to D.A.P.; National Institute on Aging (NIA) grant 5R00AG059953-05 and Alzheimer's Association grant 24AARG-1188074 to H.C.H.; and NIH grants R01AG065628 and R21AG083638 to G.E.S. ROSMAP is supported by P30AG10161, P30AG72975, R01AG17917, R01AG015819, U01AG072572, and U01AG046152.

AUTHOR CONTRIBUTIONS

J.P.W. conceptualized the study, processed the data, performed the formal analysis and visualization, and wrote the original draft of the manuscript. H.C.H., D.A.B., M.L., D.A.P., and G.E.S. contributed to study design, provided critical feedback, and aided in the interpretation of results. All authors reviewed and edited the manuscript and approved the final submission.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Adams, K. L., Dahl, K. D., Gallo, V., & Macklin, W. B. (2021). Intrinsic and extrinsic regulators of oligodendrocyte progenitor proliferation and differentiation. *Seminars in Cell & Developmental Biology*, *116*, 16–24. Special Issue: Myelin edited by Gonçalo Castelo-Branco and Roman Chrast / Special issue: Aging in the nervous system edited by Mara Mather.

Akay, L. A., Effenberger, A. H., & Tsai, L.-H. (2021). Cell of all trades: oligodendrocyte precursor cells in synaptic, vascular, and immune function. *Genes & development*, *35*, 180–198.

Bastos, C. R., Bevilacqua, L. M., Mendes, L. F. B., Xavier, J., Gruhn, K., Kaster, M. P., & Ghisleni, G. (2024). Amygdala-specific changes in cacna1c, nfat5, and bdnf expression are associated with stress responsivity in mice: A possible mechanism for psychiatric disorders. *Journal of Psychiatric Research*, *175*, 259–270.

Behrendt, G., Baer, K., Buffo, A., Curtis, M. A., Faull, R. L., Rees, M. I., Götz, M., & Dimou, L. (2013). Dynamic changes in myelin aberrations and oligodendrocyte generation in chronic amyloidosis in mice and men. *Glia*, *61*, 273–286.

Bennett, D. A., Buchman, A. S., Boyle, P. A., Barnes, L. L., Wilson, R. S., & Schneider, J. A. (2018). Religious orders study and rush memory and aging project. *Journal of Alzheimer's Disease*, *64*, S161–S189. PMID: 29865057.

Buckley, R. F., Gong, J., & Woodward, M. (2023). A call to action to address sex differences in alzheimer disease clinical trials. *JAMA Neurology*, *80*, 769–770.

Casaletto, K. B., Nichols, E., Aslanyan, V., Simone, S. M., Rabin, J. S., La Joie, R., Brickman, A. M., Dams-O'Connor, K., Palta, P. et al. (2022). Sex-specific effects of microglial activation on alzheimer's disease proteinopathy in older adults. *Brain*, *145*, 3536–3545.

COvid-19 Multi-omics Blood ATlas (COMBAT) Consortium (2022). A blood atlas of covid-19 defines hallmarks of disease severity and specificity. *Cell*, *185*, 916–938.e58.

De Jager, P. L., Srivastava, G., Lunnon, K., Burgess, J., Schalkwyk, L. C., Yu, L., Eaton, M. L., Keenan, B. T., Ernst, J. et al. (2014). Alzheimer's disease: early alterations in brain dna methylation at ank1, bin1, rhbdf2 and other loci. *Nature Neuroscience*, *17*, 1156–1163.

De Strooper, B., & Karran, E. (2016). The cellular phase of alzheimer's disease. *Cell*, *164*, 603–615.

Franceschi, C., Garagnani, P., Parini, P., Giuliani, C., & Santoro, A. (2018). Inflammaging: a new immune–metabolic viewpoint for age-related diseases. *Nature Reviews Endocrinology*, *14*, 576–590.

Gabitto, M. I., Travaglini, K. J., Rachleff, V. M., Kaplan, E. S., Long, B., Ariza, J., Ding, Y., Mahoney, J. T., Dee, N. et al. (2024). Integrated multimodal cell atlas of alzheimer's disease. *Nature Neuroscience*, *27*, 2366–2383.

Hore, V., Viñuela, A., Buil, A., Knight, J., McCarthy, M. I., Small, K., & Marchini, J. (2016). Tensor decomposition for multiple-tissue gene expression experiments. *Nature Genetics*, *48*, 1094–100.

Iturria-Medina, Y., Adewale, Q., Khan, A. F., Ducharme, S., Rosa-Neto, P., O'Donnell, K., Petyuk, V. A., Gauthier, S., Jager, P. L. D. et al. (2022). Unified epigenomic, transcriptomic, proteomic, and metabolomic taxonomy of alzheimer's disease progression and heterogeneity. *Science Advances*, *8*, eabo6764.

Kang, S., Ko, E. Y., Andrews, A. E., Shin, J. E., Nance, K. J., Barman, P. K., Heeger, P. S., Freeman, W. M., Benayoun, B. A., & Goodridge, H. S. (2024). Microglia undergo sex-dimorphic transcriptional and metabolic rewiring during aging. *Journal of Neuroinflammation*, *21*, 150.

Kim, E. K., & Choi, E.-J. (2010). Pathological roles of mapk signaling pathways in human diseases. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, *1802*, 396–405.

Klein, H.-U., McCabe, C., Gjoneska, E., Sullivan, S. E., Kaskow, B. J., Tang, A., Smith, R. V., Xu, J., Pfenning, A. R. et al. (2019). Epigenome-wide study uncovers large-scale changes in histone acetylation driven by tau pathology in aging and alzheimer's human brains. *Nature Neuroscience*, *22*, 37–46.

Lopes, C., Delezoide, A.-L., Delabar, J.-M., & Rachidi, M. (2006). Barhl1 homeogene, the human ortholog of the mouse barhl1 involved in cerebellum development, shows regional and cellular specificities in restricted domains of developing human central nervous system. *Biochemical and Biophysical Research Communications*, *339*, 296–304.

Lopez-Lee, C., Kodama, L., Fan, L., Zhu, D., Zhu, J., Wong, M. Y., Ye, P., Norman, K., Foxe, N. R. et al. (2024). Tlr7 drives sex differences in age- and alzheimer's disease–related demyelination. *Science*, *386*, eadk7844.

Maßberg, D., & Hatt, H. (2018). Human olfactory receptors: Novel cellular functions outside of the nose. *Physiological Reviews*, *98*, 1739–1763. PMID: 29897292.

Mathys, H., Davila-Velderrain, J., Peng, Z., Gao, F., Mohammadi, S., Young, J. Z., Menon, M., He, L., Abdurrob, F. et al. (2019). Single-cell transcriptomic analysis of alzheimer's disease. *Nature*, *570*, 332–337.

Mathys, H., Peng, Z., Boix, C. A., Victor, M. B., Leary, N., Babu, S., Abdelhady, G., Jiang, X., Ng, A. P. et al. (2023). Single-cell atlas reveals correlates of high cognitive function, dementia, and resilience to alzheimer's disease pathology. *Cell*, *186*, 4365–4385.e27.

McQuail, J. A., Frazier, C. J., & Bizon, J. L. (2015). Molecular aspects of age-related cognitive decline: the role of gaba signaling. *Trends in Molecular Medicine*, *21*, 450–460.

Mi, Y., Burnham, K. L., Charles, P. D., Heilig, R., Vendrell, I., Whalley, J. P., Torrance, H. D., Antcliffe, D. B., May, S. M. et al. (2024). High-throughput mass spectrometry maps the sepsis plasma proteome and differences in patient response. *Science Translational Medicine*, *16*, eadh0185.

Munoz, L., & Ammit, A. J. (2010). Targeting p38 mapk pathway for the treatment of alzheimer's disease. *Neuropharmacology*, *58*, 561–568.

Mustaly-Kalimi, S., Gallegos, W., Marr, R. A., Gilman-Sachs, A., Peterson, D. A., Sekler, I., & Stutzmann, G. E. (2022). Protein mishandling and impaired lysosomal proteolysis generated through calcium dysregulation in alzheimer's disease. *Proceedings of the National Academy of Sciences*, *119*, e2211999119.

Mustaly-Kalimi, S., Gallegos, W., Steinbrenner, D., Gupta, S., Houcek, A. J., Bennett, D. A., Marr, R. A., Peterson, D. A., Sekler, I., & Stutzmann, G. E. (2025). Mitochondrial dysfunction mediated by er-calcium dysregulation in neurons derived from alzheimer's disease patients. *Acta Neuropathologica Communications*, *13*, 165.

Perlman, K., Couturier, C. P., Yaqubi, M., Tanti, A., Cui, Q.-L., Pernin, F., Stratton, J. A., Ragoussis, J., Healy, L. et al. (2020). Developmental trajectory of oligodendrocyte progenitor cells in the human brain revealed by single cell rna sequencing. *Glia*, *68*, 1291–1303.

Ping, L., Duong, D. M., Yin, L., Gearing, M., Lah, J. J., Levey, A. I., & Seyfried, N. T. (2018). Global quantitative analysis of the human brain proteome in alzheimer's and parkinson's disease. *Scientific Data*, *5*, 180036.

Rozycka, A., & Liguz-Lecznar, M. (2017). The space where aging acts: focus on the gabaergic synapse. *Aging Cell*, *16*, 634–643.

Stutzmann, G. E. (2007). The pathogenesis of alzheimers disease—is it a lifelong "calciumopathy"? *The Neuroscientist*, *13*, 546–559. PMID: 17901262.

Xie, Z., Bailey, A., Kuleshov, M. V., Clarke, D. J. B., Evangelista, J. E., Jenkins, S. L., Lachmann, A., Wojciechowicz, M. L., Kropiwnicki, E. et al. (2021). Gene set knowledge discovery with enrichr. *Current Protocols*, *1*, e90.