HHS Public Access

Author manuscript

Cell Rep. Author manuscript; available in PMC 2026 June 08.

Published in final edited form as:

Cell Rep. ; 45(4): 117235. doi:10.1016/j.celrep.2026.117235.

Integration of aged brain multi-omics reveals cross-system mechanisms underlying Alzheimer's disease heterogeneity

Lucas P. Scheidemantel[1], **Katia de Paiva Lopes**[1,2], **Chris Gaiteri**[2,3], **Vilas Menon**[4], **Philip L. De Jager**[4], **Julie A. Schneider**[2], **Aron S. Buchman**[2], **Yanling Wang**[2], **Shinya Tasaki**[2], **Roberto T. Raittz**[1], **David A. Bennett**[2], **Ricardo A. Vialle**[1,2,5,*]

[1]Laboratory of Artificial Intelligence Applied to Bioinformatics, Professional and Technological Education Sector (SEPT), Federal University of Parana, Curitiba 81520-260, Brazil

[2]Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL 60612, USA

[3]Department of Psychiatry, SUNY Upstate Medical University, Syracuse, NY 13210, USA

[4]Center for Translational and Computational Neuroimmunology, Department of Neurology, Columbia University Irving Medical Center, New York, NY 10032, USA

[5]Lead contact

SUMMARY

The molecular correlates of Alzheimer's disease (AD) are increasingly being defined by multi-omics. However, findings from different data types are often difficult to reconcile. Here, we apply a data-driven multi-omics framework integrating seven omics layers from up to 1,358 aged human brain samples from the Religious Orders Study and Rush Memory and Aging Project. We demonstrate sprawling cross-omics biological factors relating to AD phenotypes. The strongest AD-associated factor (factor 8) is characterized by elevated immune activity at the epigenetic level, decreased heat shock gene expression in the transcriptome, and disrupted energy metabolism and cytoskeletal dynamics in the proteome. Unsupervised clustering reveals 11 molecular subtypes, including three AD-associated clusters displaying distinct molecular signatures and phenotypic characteristics. Our findings provide a comprehensive map of molecular mechanisms underlying AD heterogeneity, highlighting neuroinflammatory processes and yielding potential biomarkers and therapeutic targets for precision medicine approaches.

*Correspondence: ricardo_a_vialle@rush.edu.

AUTHOR CONTRIBUTIONS

R.A.V. designed the project. L.P.S. performed the main analysis. K.d.P.L. performed the clustering analysis. L.P.S. and R.A.V. drafted the manuscript. K.d.P.L., S.T., C.G., A.S.B., R.T.R., and D.A.B. edited the manuscript and assisted in writing. P.L.D.J. and V.M. generated the snRNA-seq atlas. Y.W. sequenced the bulk RNA-seq data. C.G., J.A.S., P.L.D.J., and D.A.B. funded the project. All authors read and approved the final manuscript.

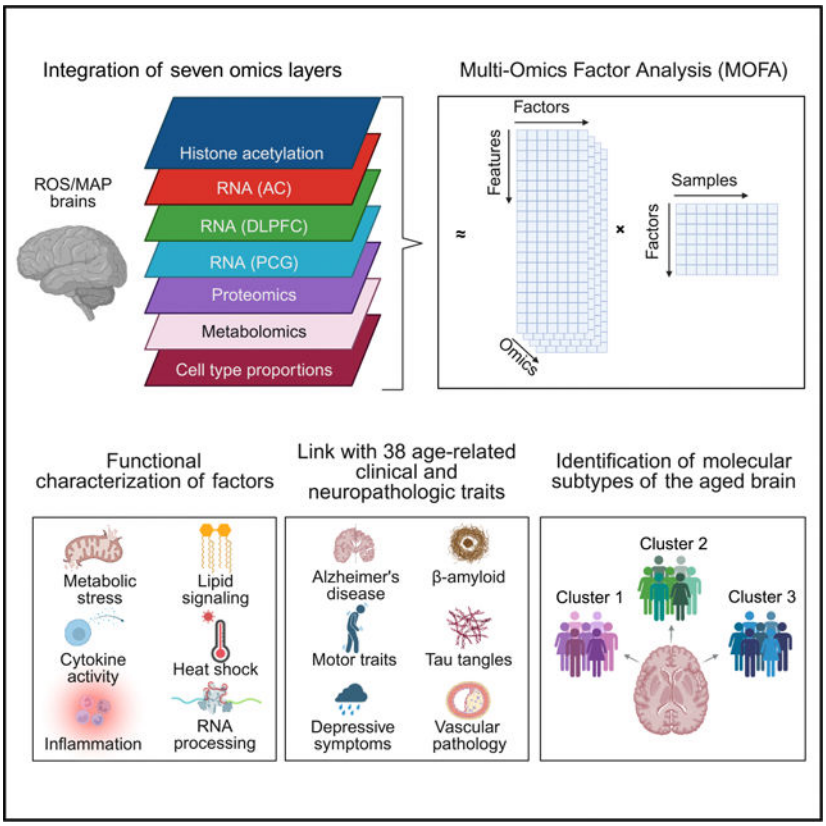
DECLARATION OF INTERESTS

The authors declare no competing interests.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work, the authors used ChatGPT to suggest improvements in grammar, clarity, precision, and impact of the writing. After using this tool or service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Graphical Abstract



In brief

Scheidemantel et al. address the complexity of Alzheimer's disease by integrating comprehensive multi-omics data from over 1,300 aged individuals, revealing coordinated molecular mechanisms across brain systems. The findings provide crucial insights into age-related traits and disease pathways, paving the way for potential therapeutic strategies.

INTRODUCTION

Alzheimer's disease (AD) is a complex neurodegenerative disorder characterized by progressive cognitive decline, accumulation of β-amyloid (Aβ) plaques and neurofibrillary tangles (NFTs), neuroinflammation, and synaptic loss, significantly impacting older populations. Multifactorial conditions are known to influence the risk of the sporadic late-onset form (LOAD) of the disease, including age, sex, genetic predisposition (mainly driven by APOE ε4 haplotype), and various lifestyle and health determinants. Despite extensive research efforts, effective treatments remain limited, in part due to the therapeutic focus predominantly on β-amyloid pathology.[1] Numerous alternative molecular hypotheses for AD have been suggested, extending beyond β-amyloid and NFTs. Cholinergic neuron damage, oxidative stress, immune dysfunction, and neuroinflammation are some examples that emerge as potential targets for treatment.[2–4] Pathophysiological changes in lipid and glucose metabolism, among others, also play key roles in the onset and development

of the disease.[5–8] Still, these mechanisms are not fully elucidated, and a unified theory remains elusive. Generating comprehensive omics data from brain tissues can improve our understanding of these biological processes and may lead to early diagnosis and the development of targeted therapies.

In recent years, a vast amount of omics data has been generated to study molecular changes in AD biology, often focusing on specific molecular levels, such as transcriptomics or proteomics.[9–12] These efforts were pivotal in revealing and validating potential risk mechanisms not only in AD but also in other neurodegenerative diseases and led to the nomination of hundreds of candidate gene targets.[13–22] For example, analysis of co-expression networks from RNA sequencing (RNA-seq) data led to the prioritization of module 109, a set of genes associated with AD traits and strongly linked to cognitive decline.[23] Similarly, networks derived from proteomics data revealed the matrisome module (M42) and the MAPK signaling and metabolism module (M7), both associated with AD but not observed in RNA networks.[24] While these studies have provided valuable insights, they are inherently limited in scope. Single-omics methods focus on isolated molecular layers and may fail to capture the interconnected complexity of the molecular framework. Studies integrating multi-omics have been proposed to overcome this limitation and have been applied in multiple contexts, including identifying molecular signatures associated with AD and suggesting better pre-symptomatic-stage treatments,[25] identifying relevant genes,[26] suggesting molecular biomarkers,[27] providing potential drug and therapy targets,[28] identifying possible disease subtypes,[29] detecting distinct molecularly differentiable disease trajectories,[30] and translating these molecular trajectory signatures from *in vivo* biomarkers.[31] However, most of these studies rely on feature preselection or explicit modeling of known phenotypes or disease states, which can bias interpretation and limit the discovery of unexpected mechanisms.

Here, we performed an unbiased, data-driven computational approach to model multilayered omics data from postmortem aged human brain tissues. Specifically, we applied multi-omics factor analysis (MOFA), an unsupervised latent factor model, designed to capture shared and modality-specific sources of variation across heterogeneous multi-omics datasets.[32,33] MOFA differs fundamentally from traditional dimensionality reduction techniques, such as principal-component analysis (PCA), in that it is generative, is probabilistic, and explicitly models missing data, which is a common challenge in real-world multi-omics studies. Moreover, it supports the interpretation of learned factors in terms of both contributing molecular features and sample-level variation, making it particularly well suited to uncover biologically meaningful patterns that span molecular modalities. In previous studies, MOFA has been used to discover molecular mechanisms and altered pathways associated with AD pathology,[8] detect sex-specific metabolic alterations in the AD mouse models,[34] and provide potential drug and therapy targets.[29] Building on this, we applied MOFA to integrate epigenomics, transcriptomics, proteomics, metabolomics, and cell-type subpopulation data from two cohort studies of aging and dementia: the Religious Orders Study (ROS) and the Rush Memory and Aging Project (MAP). We then performed integrative pathway analyses to characterize the molecular mechanisms involved in these derived factors. Finally, we performed sample clustering based on factor scores to identify molecular subtypes of AD.

RESULTS

Characteristics of the participants

We utilized data from 1,358 participants in the ROS/MAP cohorts. After enrollment, participants from both cohorts were annually evaluated for physiological and cognitive function followed by brain donation at death.[35] Of the 1,358 participants, 436 (32.1%) were male, the mean age at death was 89.4 years, and the average years of education was 16.2. A total of 600 (44.2%) participants had Alzheimer's dementia, while 318 (23.4%) had mild cognitive impairment (MCI). TDP-43 pathology extending beyond the amygdala was observed in 32.1% of participants, while Lewy bodies in the nigra, amygdala, and/or cortex regions were present in 23.2% of cases. Cerebrovascular diseases, including macroscopic infarcts and microinfarcts, were identified in 35.9% and 29.3% of cases, respectively. Additionally, moderate to severe amyloid angiopathy, atherosclerosis, and arteriolosclerosis were observed in about a third of the brains. More than a quarter of participants carried at least one APOE ε4 allele and, similarly, at least one long allele of TOMM40′523 (Table S1).

MOFA of the aging human brain

We implemented MOFA integrating seven different omics datasets (hereafter referred to as "views") sourced from a partially overlapping set of brain samples, totaling 1,358 ROS/MAP participants (Figure S1). These included H3K9ac histone acetylation markers measuring 26,384 peaks (n = 632); mRNA expression of 17,309 transcripts from three brain regions, anterior caudate (AC; n = 721), dorsolateral prefrontal cortex (DLPFC; n = 1,210), and posterior cingulate gyrus (PCG; n = 656); proteomics comprising 8,425 proteins (n = 596); metabolomics covering 667 metabolites (n = 500); and proportions of 96 cell-type-specific subpopulations from the cortex (n = 424) (Figure 1A). Similar demographic distribution was observed across views (Figure 1B; Table S1). To reduce the impact of technical confounders in the model, all input data were normalized and adjusted prior to running MOFA (Figures S2A–S2H). Moreover, to provide an unbiased characterization of brain molecular signatures, no additional preselection step was applied to the input features (i.e., genes, proteins, epigenetic markers, metabolites, or cell proportions) other than the standard cutoffs applied during the quality control (QC) stage to remove genes/proteins with very low expression/abundance (e.g., a minimum of 10 reads in at least 50% of samples in RNA-seq). MOFA was set to decompose the data into 50 largely orthogonal latent factors (Figure S3) representing the underlying sources of variation across samples. Importantly, technical and biological confounders, such as postmortem interval (PMI), RNA integrity number (RIN), omics-specific metrics, and genetic principal components, were not strongly correlated with the resulting factors (Figure S2I). We also systematically evaluated the relationship between the average abundance of each feature and its respective MOFA factor weight. We did not observe a clear bias toward low-abundance features in the driving factors (Figures S4 and S5). The most significant portion of total data variance was captured by mRNA (over 40% of variance explained), followed by proteomics (over 35%) (Figure 1C). In addition, different contribution patterns were observed from each view, for example, factors 3, 5, and 6 were mostly driven by transcriptomics from DLPFC, AC, and PCG, respectively, while other views had their variance explained by at least two or more omics,

indicating possible more complex interplay between different molecular levels or tissues (Figures 1E and S6).

Identification of AD-associated factors

From the 50 decomposed factors, we selected nine that were associated (false discovery rate [FDR] < 0.05) either with pathological diagnosis of AD (NIA-Reagan score) or with clinical diagnosis of Alzheimer's dementia for downstream analysis (Figures 1D and S7). These factors accounted for a range of total variance, from 20.5% (factor 1) to 1.7% (factor 42) (Figure S8). To better characterize their phenotypic impact, we further performed association tests against 38 common age-related clinical and neuropathologic traits, including indices for cognition, pathology, motor function, disabilities, depression, and genetic risk variants for AD (Figure 1F; Tables S2 and S3). These analyses revealed multiple associations (FDR < 0.05) between factors and phenotypes, primarily with cognitive and AD pathological traits but also with motor function and disabilities, most of which persisted after adjusting for APOE ε4 (Figure S9). However, disentangling associations with specific AD endophenotypes is inherently challenging in aging cohorts such as ROS/MAP, where multiple neuropathologies and age-related diseases frequently co-occur. We also repeated the association analyses with additional adjustment for AD status, considering pathological AD (e.g., NIA-Reagan score) and clinical AD diagnosis in separate models. These analyses reinforced the strong interconnectivity of the phenotypic information. For example, associations with motor phenotypes remained significant after adjustment for pathological AD but were no longer significant after adjustment for clinical diagnosis, suggesting that these associations are more closely linked to clinical AD diagnosis than to AD pathology alone (Figure S10).

Among the main findings, factor 8 showed the strongest associations (i.e., lowest p value) with cognitive decline and was linked to 30 of the 38 tested phenotypes, followed by factors 4 and 14, associated with 27 traits each. Although many factors shared correlations across the same phenotypes (as expected due to the intrinsic correlation between phenotypes), some specific factor-trait associations emerged, suggesting that certain phenotypes might be linked to specific molecular mechanisms. For example, gross cerebral infarction was associated exclusively with factor 2, while major depressive disorder diagnosis was uniquely linked to factor 14. We also conducted a variance partition analysis to jointly assess the contribution of multiple phenotypes to each factor. This analysis revealed distinct patterns of phenotypic influence. For instance, while variance in factors 8 and 14 was mostly affected by pathologies, factor 8 had an additional influence of age at death, whereas factor 14 showed substantial effects from depression-related phenotypes (Figure 1G). By analyzing the biological features driving each factor, we were able to describe the molecular functions associated with each view. All nine AD-related factors were enriched with multiple functional terms involving key cellular mechanisms (Figure 1H; Tables S4–S12). These encompassed terms related to immune response (factors 2, 3, 4, 8, and 26), mitochondria (factors 1, 8, and 14), ribosomes (factors 1, 4, 8, 14, and 26), synapses (factors 12 and 42), proteostasis (factors 1, 8, and 26), cytoskeleton (factors 3 and 8), vasculature (factors 26 and 42), and RNA processing (factors 12 and 26).

MOFA robustness and reproducibility

To assess the robustness and the replication of the MOFA results, we compared feature weight correlations across different scenarios (Figure 2; Figures S11–S13). Since there is no guarantee of factor order or direction across runs with different inputs, we used a simple best-matching approach based on correlation of feature loadings. First, we evaluated whether MOFA could handle subsampling while preserving omics proportions. To do that, we randomly selected samples with decreasing sample sizes (i.e., 90%, 80%, 70%, 60%, and 50%). Each threshold was repeated 15 times with different random seeds. MOFA was then run with the same parameters as the full analysis, and the resulting matched factors were compared with the loadings from the 50 factors presented in the full model. Results showed that MOFA is robust to sample size reduction while keeping a similar omics contribution. As expected, the first factors, which explain more of the data variance, showed consistently high correlations across models (Pearson's $r > 0.8$ on average) (Figure 2A). We also tested the reproducibility by randomly splitting the dataset into two non-overlapping donor sets, training independent MOFA models on each, and correlating the feature weights across matching factors (also 15 random splits). Results were similar to those from 50% vs. the full model, suggesting that the conserved factors are not driven by individual-specific data (Figure 2B).

Next, we evaluated the impact of changing the omics input. First, we restricted the analysis to samples with increasing numbers of omics layers. We trained MOFA models on subsets comprising only individuals with at least two, three, or four omics layers and compared them with the complete model (i.e., including all individuals with at least one view). Under these conditions, there is also an effect of reducing sample sizes due to data availability. Overall, the results resembled those of reducing the sample size fractions (Figure 2C). The second scenario involved keeping the sample size fixed but completely removing one omics layer. For these, the features of the removed view were set to "NA" (not available), and the resulting loadings were compared with those of the remaining features for which data were available. Results showed distinct contributions of different views to the final factors (Figure 2D). For instance, removing the view of cell-type fractions had little effect on the 50 factors; however, removing RNA-seq from DLPFC showed the strongest disturbance and reduction in reproducibility. In addition, removing specific omics had stronger detrimental effects on factors for which most of the data variance was explained by the same modality. For example, factor 12 is mostly driven by H3K9ac data; thus, removing that data from the input reduced the correlation of factor 12 considerably. However, more importantly, the remaining factors seemed robust to the change. Overall, our tests suggest that, among the nine AD-related factors, factors 1, 2, 3, 4, 8, and 12 are robust and reproducible across most changes; factors 14 and 26 show moderate correlation; and factor 42 seems less likely to be reproduced under the tested conditions.

Finally, to assess the impact of missing specific views on the MOFA factor scores and their respective phenotypic associations, we conducted another set of experiments. First, using the full MOFA model, we divided the samples into two groups: one with proteomics data and the other without, matched for sample size ($n = 596$). For this test, we focused on factor 14, where variance was primarily driven by proteomics. We then performed regression analyses

on all phenotypes to evaluate how the presence or absence of proteomics measurements would influence the results (Figure 2E). We found that, although both groups showed significant associations with several phenotypes, individuals with proteomics data in the input exhibited considerably stronger signals and more biologically interpretable results. This suggests that proteomics missingness negatively affects the final factor scores and that the other views are insufficient to infer the missing information. Based on this observation, we conducted additional tests to determine whether the absence of specific omics data could be a major confounder in phenotype association results. We recalculated the factor associations with traits, including information on the presence or absence of each omics layer as a covariate in the regression models. We found that this had little effect on the results (Figure S14). Therefore, the absence of omics seems more related to statistical power and does not confound the observed findings. Overall, within this framework, it seems essential to include as many omics layers as possible and increase their respective sample sizes, as they contribute to greater statistical power to associate factor scores with traits of interest and can enhance the overall interpretability of a multi-omics analysis.

Factor 8 emerges as a key disease multi-omics unifier

To further understand the molecular functions that could impact the disease, we investigated the omics-specific feature weights contributing to each of the AD-related factors. Here, we highlight results from factor 8; results for other factors can be found in Figures S15–S22. Factor 8 was positively associated with higher disease risk, displaying the strongest positive association with Alzheimer's dementia (FDR = 2.01×10^{-13}) and NIA-Reagan score (FDR = 4.34×10^{-10}) and the strongest negative association with global cognitive function (FDR = 2.53×10^{-20}) and cognitive decline (FDR = 2.53×10^{-20}). Factor 8 was also positively associated with all pathological traits, except for Parkinson's disease (PD) pathology, as well as with APOE ε4 (FDR = 4.93×10^{-3}) and TOMM40′523-L (FDR = 4.10×10^{-2}) haplotypes, and was, among the AD-related factors, the only factor positively associated with cerebral atherosclerosis (FDR = 3.74×10^{-2}), frailty (FDR = 1.51×10^{-2}), and gait score (mUPDRS) (FDR = 8.40×10^{-3}) (Figure 1F).

Factor 8 explained 9.17% of the total variance of the input data, with overall similar contribution of multiple views, ranging from 0.94% (H3K9ac) to 1.76% (DLPFC mRNA) (Figure 1E). Still, many biologically relevant terms were identified within each view (Figure 3A). Histone acetylation peaks (H3K9ac) were primarily enriched for immune-related processes, including immune response (FDR = 3.86×10^{-7}) and interferon signaling (FDR = 1.17×10^{-3}). H3K9ac peaks also showed positive enrichment for gene markers of Mic.14 (FDR = 3.61×10^{-7}), a microglial subtype characterized by interferon response.[36] At the RNA level, AC and PCG transcriptomes shared enrichment for terms related to heat shock factor 1 (HSF1) activation and cellular response to heat stress. AC also showed enrichment for protein folding processes, including response to topologically incorrect proteins (FDR = 2.27×10^{-6}), while PCG was enriched for immune system and transport-related terms, such as immune receptor activity (FDR = 2.52×10^{-3}) and organic acid sodium symporter activity (FDR = 1.27×10^{-2}). Proteomics was enriched for cytoskeletal organization (FDR = 9.91×10^{-8}), mitochondrial matrix (FDR = 2.35×10^{-14}), and co-expression protein modules, such as Prot-M24 (ubiquitination-related, FDR = 4.99×10^{-19}) and Prot-M7

(MAPK metabolism-related, FDR = 3.96×10^{-9}) (Figure 3A). Finally, at the metabolomics level, factor 8 was enriched for pathways involving the urea cycle, arginine, and proline metabolism (FDR = 3.10×10^{-2}) (Table S8).

Analysis of the top features contributing to factor 8 (Figures 3B and 3C) highlights the multifaceted biological signature linked to AD pathophysiology. Genes such as *AQP6*, *GIPR*, *GLI1*, *S100A4*, and most members of the SLC family have previously been shown to be upregulated in AD brains.[37] In addition, also contributing to factor 8, the stress response cells of AST.10 (*SLC38A2*), previously associated with increased tau pathology and rapid cognitive decline,[36] showed the highest feature weight for factor 8 (relative to other cell proportions), indicating a strong contribution to this factor. Metabolically, glutarylcarnitine was associated with mitochondrial fatty acid β-oxidation, which was prominent in our analysis.[38,39] To further understand the possible mechanisms connecting the different modalities, we performed a network-based analysis using STRING to capture both physical and functional associations between proteins. Genes with absolute weights greater than 0.3 (Figure 3B) were used as input. Next, the largest connected network component was selected, representing the most comprehensive set of genes with supporting evidence of association (Figure 3D). In this subnetwork, among features with strong contributions to factor 8, we highlight *CLEC7A*, *P2RY12*, *GLUL*, and *CTGF*, with positive weights, and *HSPA1B*, *HSPB1*, *SDSL*, and *BAG3*, with negative weights. Moreover, some of these genes were strong contributors across multiple data modalities, including *BAG3* and *GLUL*, as well as other heat shock family genes (*HSP90AA1* and *HSP90AB1*) and purinergic G-protein-coupled receptor genes (*P2RY1*). Enrichment analysis of this subnetwork revealed strong enrichment of stress response, chaperone, host-virus interaction, glycoproteins, and MHC II keywords. Overall, the integration across multiple omics captured by factor 8 reveals key mechanisms, such as immune response (epigenetic), reduced cytoskeletal organization (mainly proteomics driven), decreased levels of heat shock chaperones, and increased transmembrane transporter gene expression, interconnected through key genes such as *BAG3* and *GLUL*, and highlights pathological mechanisms in AD.[40–44]

Discordant immune-related functions modulate the protective effects of factors 2 and 3

Factors 2 and 3 exhibited an overlap of several functional terms across multiple views, representing a convergence of the same genes and pathways at distinct molecular levels (Figure 1H). Both factors were strongly enriched for processes involving immune response and displayed overlapping functional terms across their views, with the key difference that these enrichments were mostly in opposite directions (Figures S16A and S17A). Among the top feature contributors to factors 2 and 3, we identified several genes related to neuroinflammatory processes and β-amyloid metabolism, such as *SERPINA3* and *CD14* (Figures S16B and S17B). Additionally, factor 2's highest positive weights for cell types included AST.5, OPC.3, and MIC.7, reactive-like cell subpopulations linked to the brain aging trajectory independent of AD progression.[36] In contrast, factor 3 included the same cell types among the highest negative weights alongside positive contributions from a microglial subpopulation involved in cell surveillance (MIC.4).

Although these factors shared overall functional enrichments in opposite directions, both were negatively associated with AD and neuropathologies (Figure 1F). We then investigated the feature weights contributing to all seven omics views. We found that, for most omics, there was a significant negative correlation, ranging from Pearson's r = −0.55 in H3K9ac to −0.12 in AC RNA-seq, explaining the observed rank-based enrichment results (Figure 4A). However, specifically in the DLPFC transcriptomics, we identified two clusters of seemingly contradictory gene contributions observed between factors 2 and 3 (Figure 4B). Cluster 1 shows a strong negative association (r = −0.32) and cluster 2 shows a strong positive correlation (r = 0.80). Further functional enrichment analysis of these two clusters revealed a concordant gene signature of increased synaptic function and decreased vasculature-related gene expression (cluster 2) and the previously described discordant immune-related function in the discordant set of genes (cluster 1), suggesting potential protective immune mechanisms operating through different cellular pathways but converging on similar neuronal outcomes (Figure 4C). These results observed in factors 2 and 3 suggest that the timing and context of immune activation may be more critical than the specific immune pathways involved in determining whether neuroinflammatory responses are protective or detrimental in AD pathogenesis. Further studies are needed to clarify the temporal dynamics and cellular specificity of these immune-related processes in AD progression.

Additional factors reveal complex molecular mechanisms involved in AD

The remaining AD-related factors also provide important aspects of aging and AD biology. Factor 1 (Figure S15), which explains the largest proportion of total data variation, was negatively enriched in mitochondrial and ribosome-related terms and positively enriched with MAPK and sugar metabolism, RNA splicing, and ubiquitination terms. Most of these enrichments were primarily driven by proteomics, with additional contributions from AC transcriptomics linking heat shock stress response genes. Factor 1 was also positively associated with AD in our tests, aligning with previous findings linking these functions to high AD pathology and worse cognitive outcomes at the proteomics level.[24] Factor 14 (Figure S20) demonstrated the second strongest association with AD pathology after factor 8, uniquely correlating with both PD pathology (FDR = 4.87×10^{-6}) and major depressive disorder (FDR = 2.78×10^{-2}). This factor was primarily driven by proteomics (2.83% variance) and metabolomics (1.19% variance), with enrichments pointing to the downregulation of respirasome- and oxidoreductase-related terms and increased protein co-expression modules linked to cell-extracellular matrix interactions and MAPK metabolism. Factor 14 was also linked to metabolomic alterations in energy metabolism pathways, including glycolysis, gluconeogenesis, and pyruvate metabolism, having uridine diphosphate (UDP)-glucose and FG dipeptide (F and G amino acids) as top metabolite markers. Factor 42 (Figure S22), despite explaining only 1.7% of total variance and being less robust to changes in the input (Figure 2), captured some aspects of vascular biology relevant to AD pathogenesis. Top contributors to factor 42 were PCG and AC transcriptomes, showing significant enrichments for marker genes of subpopulations of pericytes (Peri.2), smooth muscle cells (SMC.2), and arteriole, along with *APOLD1* as top positive weights in PCG, a gene involved in endothelial cell function and hemostasis.

Our analysis also identified significant protective signatures against AD pathology. Factor 26 (Figure S21) exhibited the strongest protective effect, with enrichment for RNA processing mechanisms, vascular-related pathways, and cytoskeletal organization that may preserve cognitive function despite pathological burden. Factor 4 (Figure S18) emerged as the second strongest protective factor, negatively associated with AD dementia (FDR = 2.34×10^{-6}) and NIA-Reagan score (FDR = 5.78×10^{-4}), while positively correlating with global cognitive function (FDR = 1.74×10^{-7}). Notably, factor 4 was uniquely negatively associated with Lewy body disease (FDR = 1.71×10^{-2}), PD pathology (FDR = 8.27×10^{-5}), and depressive symptoms (FDR = 9.09×10^{-3}). Its protective mechanisms involve high expression of genes such as *VGF* and *CX3CR1*, enrichment for microglial subtype MIC.14 characterized by interferon response, and metabolomics signatures involving lysophospholipids and long-chain polyunsaturated fatty acids. The features with the highest positive weights included several genes considered protective against AD, such as *CX3CR1*, which modulates microglial activation, and metabolites such as trigonelline, which has shown protective functions in mice by ameliorating axonal and dendritic atrophy in cortical neurons treated with β-amyloid.[45] Finally, factor 12 (Figure S19) showed protective effects through increased cell activation, cell signaling, and extracellular matrix organization, strongly driven by H3K9ac epigenetic markers. These markers are also enriched for RNA co-expression modules M7, involved in transcription regulation and tau tangles, and M114, involved in immunity and associated with β-amyloid.

Identification of multi-omics molecular subtypes of AD

In order to represent the molecular heterogeneity of the aging brain and its relation with Alzheimer's dementia, we applied the SpeakEasy[46] consensus clustering algorithm using all 50 MOFA latent factors to group correlated participants into distinct clusters (Figure 5A). These clusters are expected to provide an unbiased classification of participants based on molecular signatures from multi-omics in the postmortem brain, which is important because patient stratification may enable more precise understanding of disease mechanisms and tailored therapeutic approaches. We found three major clusters broadly differentiating individuals, from those in a non-disease state, with lower AD and related dementias (ADRD) pathologies (major cluster 1), to those in a disease-like state, with elevated pathology and lower cognition scores (major cluster 3) (Figure 5B). Further refinement of each major level cluster resulted in 11 subclusters with a more detailed stratification of protective and pathological subtypes of AD.

Among the clusters related to non-/low-disease individuals (major cluster 1), three were identified (M1, M2, and M3). Participants in cluster M1 displayed phenotypic traits that were not significantly different from the average. Participants in cluster M2 presented a negative correlation with Alzheimer's clinical diagnosis and NIA-Reagan scores, as well as with most pathological traits and all disabilities; however, this cluster was also associated with a younger age at death, possibly confounding observation in the other traits. Cluster M3, on the other hand, exhibited a negative correlation with Alzheimer's clinical diagnosis and NIA-Reagan scores and did not show a significant correlation with age at death (Figure 5B). Clusters M2 and M3 also presented higher than average levels of factors 3 and 4, two

factors negatively associated with Alzheimer's clinical diagnosis and most AD-related traits (Figure 1F).

Among clusters positively associated with AD (major cluster 3), four subclusters were identified (M6, M7, M8, and M9). While participants in cluster M8 did not show differences compared with the average, clusters M6, M7, and M9 shared similar clinical and pathological characteristics. These clusters were positively associated with Alzheimer's dementia, MCI, and NIA-Reagan scores, as well as with most pathological traits, and negatively correlated with cognitive function and cognitive decline (Figures 5C–5E). Despite these similarities, each cluster showed distinct phenotypic differences. Cluster M6, for example, had higher numbers of females and higher age at death compared with the average. It was also the only cluster with higher than average indices of instrumental activities of daily living (IADL). Cluster M7 had the strongest associations with motor traits, namely bradykinesia score (mUPDRS), gait score (mUPDRS), and hand strength, compared with any other cluster, but it also had higher frequencies of AD genetic risk factors (APOE/TOMM40). Meanwhile, participants in cluster M9 showed the strongest correlations with NFTs, hippocampal sclerosis, and gross cerebral infarctions compared with any other cluster. These clusters also differed in their molecular signatures. While cluster M9 was mostly influenced by factor 8, cluster M7 was primarily associated with factor 42 and, to a lesser extent, factor 12. Conversely, M6 correlated with factor 14 and, to a smaller degree, with factors 1 and 8. Overall, multi-omics stratification reveals discrete AD subtypes with unique molecular and phenotypic profiles, which may help refine prognosis and guide the development of tailored therapeutic approaches.

DISCUSSION

Here, we employed MOFA, an unsupervised method for identifying the principal sources of variation across multimodal data, to analyze multi-omics data from the ROS/MAP cohorts. The data included H3K9ac histone acetylation, mRNA expression from three brain regions (AC, PCG, and DLPFC), proteomics, metabolomics, and proportions of cell-type-specific subpopulations. MOFA inferred a set of 50 factors that captured biological and technical sources of variability across multiple omics, with 9 of these factors showing a meaningful association with AD. Then, we performed gene set enrichment analysis (GSEA) on the selected factors to determine their molecular profiles and identified genes of interest. Finally, we clustered the ROS/MAP samples and identified several groups of people associated with various AD traits and factors.

We identified multiple factors linked to pathophysiological mechanisms associated with AD. Several factors captured distinct immune system processes, including interleukin signaling, inflammatory responses, and cytokine production, all of which are known to be dysregulated in AD.[47–49] In addition, MOFA identified other key mechanisms, including cytoskeletal organization, proteostasis, and synaptic activity, each of which has also been previously implicated in AD pathology.[50–52] Moreover, some factors showed strong associations with co-morbid conditions commonly linked to AD, including depression and vascular pathologies such as arteriosclerosis and cerebral amyloid angiopathy, both of which are known to be risk factors for AD.[53,54] Finally, our clustering approach was able to

classify participants into distinct molecular subtypes reflecting the recognized molecular heterogeneity of AD.[55] One key feature that distinguishes this analysis from prior work is the description of how these AD-implicated systems span multiple omics. This approach enables disease investigations to proceed with greater confidence, focusing on coordinated signals supported across data types, which have independent sources of biological and technical variation.

Factor 8 emerged as the signal most strongly associated with Alzheimer's dementia and was linked to nearly 80% of the tested phenotypes. Top features included genes related to AD pathophysiology and cell subpopulations positively associated with AD. These findings may serve as a resource for candidate genes of interest for potential therapies. For example, the positively weighted gene *S100A4* has been implicated in multiple AD-related pathways, including neuroinflammation and oxidative stress. *In vitro*, *S100A4* knockdown in amyloid-β oligomer (AβO)-treated BV2 microglia improved cell viability, reduced lactate dehydrogenase release, and partially alleviated apoptosis. *S100A4* inhibition attenuated pro-inflammatory responses and promoted an anti-inflammatory state. Furthermore, *S100A4* knockdown mitigated oxidative stress, restoring mitochondrial function and decreasing reactive oxygen species levels.[56] The *BAG3* (BCL-2-associated athanogene 3) gene exhibited high negative weights across multiple views, suggesting its implication in several pathways linked to factor 8, including immune response, Prot-M12 (related to the cytoskeleton), and protein folding. *BAG3* is a multifunctional protein involved in many biological processes, with evidence suggesting a protective role against AD by facilitating phospho-tau degradation[57] and autophagy of diseased mitochondria.[43,58] *BAG3* also regulates mitochondrial function by supporting calcium (Ca^{2+}) transport in and out of mitochondria,[58,59] drives the cytoskeleton dynamics by regulating actin folding,[43,60,61] inhibits caspase activation,[58,62] and interacts with HSF1[63] and MAPK molecules (e.g., *MAP3K5*).[64,65] In addition, recent studies in induced pluripotent stem cell (iPSC) models revealed that *BAG3* coordinates astrocytic proteostasis of AD-linked proteins via proteasome, autophagy, and retromer complex interactions.[66] We hypothesize that factor 8 captures genes acting in the regulation of immune response at the level of H3K9ac acetylation, protein folding and HSF1 activation in RNA, and cytoskeleton dynamics, MAPK metabolism, and mitochondrial matrix in proteomics, reflecting a coordinated multi-omics program that together may contribute to AD pathogenesis.

Our analysis revealed an intriguing correlation in factors 2 and 3. While both demonstrated protective effects against AD pathology, their overall enrichments pointed to discordant immune signatures. However, in DLPFC RNA, we found concordant upregulation of genes related to synapse and neuronal-related terms, along with another set of genes pointing to opposite immune cell activation patterns. This observation initially challenges the conventional view of neuroinflammation in AD as uniformly detrimental. However, our finding suggests that the timing, context, and specific nature of immune activation may determine whether neuroinflammatory responses are protective or harmful. The positive weights for reactive-like glial subpopulations in factor 2 and surveillance microglial cells in factor 3 indicate that different glial activation states may confer protection through distinct but complementary mechanisms, potentially by promoting neuronal resilience or facilitating clearance of pathological proteins. These observations are, however, based on postmortem

samples, and further studies in living human brains will be important for comparing these dynamics across disease stages.

Beyond these primary factors, our analysis identified additional significant contributors to AD heterogeneity. Factor 1, explaining the highest proportion of total variance, captured fundamental biological processes, including mitochondrial function, ribosomal activity, and proteostasis pathways, suggesting that these core cellular mechanisms constitute a major axis of variation in the aging brain. Factor 14 showed the second strongest association with AD pathology and was also associated with PD pathology and major depressive disorder, highlighting potential shared mechanisms between these disorders. Factors 26 and 4 exhibited protective signatures, with factor 26 showing the strongest protective effect through enrichment for RNA processing mechanisms and vascular-related pathways, while factor 4 was uniquely negatively associated with Lewy body disease and PD pathology, suggesting broader neuroprotective mechanisms. Factor 4 also exhibited high positive weights for genes and metabolites known to be protective against AD. Multifunctional genes, such as *FKBP5* and *VGF*, were found with high absolute weights across multiple views and direct relevance to AD. The immunophilin FK-506-binding protein 51 (FKBP51) may contribute to AD development through its interaction with the heat shock protein Hsp90, which leads to tau phosphorylation, inhibits tau degradation, and results in neurotoxic tau accumulation.[21,67,68] Conversely, the neurotrophin *VGF* (a nerve-growth-factor-inducible protein) and its derived peptides play a protective role in AD. For instance, the biopeptide TLQP-62 positively influences synaptic plasticity and memory.[69] Furthermore, increasing *VGF* expression in the 5×FAD mouse model has been shown to reduce β-amyloid levels in the brain and improve memory performance.[21,69]

In summary, our integrative multi-omics analysis provides a comprehensive map of molecular mechanisms underlying AD heterogeneity, revealing both shared pathological processes and distinct molecular subtypes. Our findings have several important implications for AD research and therapeutic development. First, they highlight the ability to identify well-known signatures from AD using unbiased dimensionality reduction of multi-omics data. Second, these findings extend beyond classical amyloid/tau pathology and are linked to a series of aging-related phenotypes. Third, by integrating multi-omics into a common framework, we were able to connect apparently disjoint disease components that are usually observed in single-omics analyses. Finally, by identifying AD-associated factors and molecular subtypes, we created a basis for patient stratification in clinical trials, potentially improving the success rate of AD drug development. Progress in the next generation of AD therapy will likely stem from “combination” approaches that reinforce proteostasis, buffer neurotransmission, and reduce chronic inflammation, tailored to patient-specific genetic and epigenetic profiles. The genes and features highlighted herein exemplify both the challenge and the opportunity in pursuing this multifactorial, precision-medicine vision for AD.

Limitations of the study

While our study provides valuable insights, it is important to acknowledge its limitations. MOFA, being a linear model, may overlook non-linear relationships between features within and across assays.[32,70] While a larger number of factors were initially specified to enable

broad exploration of latent structure, robustness analyses indicate that higher-order factors (that explain smaller fractions of variance) should be interpreted with caution, and biological conclusions are therefore primarily based on the most stable leading initial factors (e.g., first 1–15 factors). Also, MOFA is an unsupervised, data-driven method, and the degree of cross-modal consistency observed for a given factor depends on how strongly the data signal is shared across the different data types. Even with input data being adjusted by known technical factors, the differences between distinct modalities are still noticeable. Both the variance explained and the feature weights are often driven by specific omics modalities in each factor, and this should be considered when interpreting the results. Additionally, while our multi-omics approach provides unprecedented molecular resolution, it cannot fully capture the complex cellular interactions and spatial organization of the brain. There was a considerable disparity in the amount of information available for each omics type (e.g., 17,309 mRNA transcripts vs. 667 metabolites), which may limit the ability to assess the relative importance of individual features. Moreover, our epigenetic data were restricted to H3K9ac acetylation data, constraining our ability to capture a more comprehensive view of epigenetic regulation, potentially overlooking key methylation-associated mechanisms involved in the molecular framework of AD. In addition, ROS/MAP cohorts are mainly composed of individuals of European ancestry, and additional data from other human populations are needed to generalize the results. The cross-sectional nature of postmortem tissue analysis limits our ability to establish causal relationships and temporal dynamics. Also, the ROS/MAP cohorts represent an aging population in which AD frequently co-occurs with other age-related neuropathologies, leading to strong intercorrelations among pathological, cognitive, motor, and clinical measures. As a result, complete disentanglement of factor associations with specific AD endophenotypes from shared disease- or aging-related processes is inherently limited. Finally, this study employs an exploratory approach, utilizing collected omics data to generate data-driven hypotheses. These hypotheses will require further investigation through complementary, targeted experimental approaches conducted with statistical rigor.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Ricardo A. Vialle (ricardo_a_vialle@rush.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- This paper analyzes existing previously published data. Raw data used in this article are available through the AD Knowledge Portal (https://adknowledgeportal.org). Accession numbers are listed in the key resources table.
- The ROS/MAP phenotype data can be requested at the RADC Resource Sharing Hub. The URL is listed in the key resources table.

- All original code has been deposited at Zenodo and is publicly available at https://doi.org/10.5281/zenodo.18842083 as of the date of publication.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Study participants: Participants came from two ongoing longitudinal clinical-pathological cohort studies of aging and Alzheimer's disease conducted by the Rush Alzheimer's Disease Center (RADC), the Religious Orders Study (ROS) and the Rush Memory and Aging Project (MAP). ROS was established in 1994 and includes nuns and priests from across the United States[35,80,81], while MAP began in 1997 and includes individuals from retirement communities, subsidized housing, and private dwellings in Illinois[35,81,82]. Both cohorts, collectively referred to as ROS/MAP, admit older individuals without known dementia at the time of study enrollment. Participants agree to annual clinical assessments, cognitive tests, and blood collections, as well as organ donation (spinal cord, muscles, nerves and brain). Through March 2025, ROS/MAP enrolled 4,025 participants, of whom 2,127 came to autopsy. The present study focuses on 1,358 deceased individuals (Table S1) with omics data available. Of these, 436 (32.1%) were male, the mean age at death was 89.4 years, and the average years of education was 16.2. In the association tests of omics factors and phenotypes, regression were controlled for sex, age, and years of education.

Clinical and neuropathological evaluations: Here we briefly describe the harmonized, deeply phenotyped measures performed on ROS/MAP participants. These include clinical diagnoses of Alzheimer's dementia and MCI; measures of global cognitive function, cognitive decline, and resilience; neuropathologic evaluations of AD pathology including NIA-Reagan classification, global AD pathology, β-amyloid, tangles, diffuse and neuritic plaques, TDP-43, Lewy bodies, PD pathology, and hippocampal sclerosis; vascular pathologies including arteriolosclerosis, atherosclerosis, CAA, and gross and microinfarcts; frailty classification; indices of global motor and parkinsonian function; disability measures including basic activities of daily living (ADL), instrumental activities of daily living (IADL), and mobility disability; and diagnosis of depressive symptoms and major depressive disorder (MDD).

Cognitive function and Alzheimer's dementia: Cognitive function was assessed annually for up to 23 years (Mean = 7.4, SD = 4.9) using a battery of 21 performance tests. Scores from 17 tests were standardized into z-scores using the baseline mean and standard deviation and then averaged to derive a composite measure of global cognition. Lower z-scores reflect lower performance relative to baseline cohort norms[83]. For each individual, the annual rate of change in global cognitive function (cognitive decline) was estimated using a linear mixed effects model, controlling for age at baseline, sex, and years of

education[84]. To examine cognitive resilience, a similar model was used that further adjusted for neuropathologic measures including global AD pathology, β-amyloid, neurofibrillary tangles, Lewy body disease, TDP-43 pathology, infarcts, hippocampal sclerosis, and vascular pathologies[85]. Clinical diagnosis of cognitive status was made using a modified three-stage procedure based on a neuropsychologist's review of 19 test scores, followed by clinician review and diagnosis[82,86,87]. Final classification analyzed Alzheimer's dementia vs. no AD dementia (MCI + NCI) and with cognitive impairment (AD + MCI) vs. no cognitive impairment (NCI)[88].

Neuropathologic outcomes: Neuropathologic evaluations were performed blinded to clinical data and included uniform assessments of Alzheimer's disease and common co-pathologies. A modified version of the NIA-Reagan criteria was used to classify the likelihood of Alzheimer's disease pathology on a 4-point scale from no to high likelihood based on the presence of both neurofibrillary tangles (Braak) and neuritic plaques (CERAD)[89]. Quantitative measures of pathology were derived from silver-stained slides across multiple brain regions. Summary scores were computed for neuritic plaques, diffuse plaques, and neurofibrillary tangles[90]. These were standardized within each region, averaged across regions, and square root-transformed to reduce skewness. A composite global AD pathology score was then calculated by averaging these three summary scores[91,92]. Tangle density was assessed as the mean square root of tangle scores in eight regions, including the hippocampus and frontal cortex[93]. β-amyloid burden was quantified using immunohistochemistry across eight regions, and summarized as the square root of the mean regional β-amyloid score[90]. Additional pathologies assessed included TDP-43 staging (no TDP-43 pathology or in amygdala only, and TDP-43 pathology extending beyond amygdala)[94], Lewy body disease (binary classification based on presence in seven regions)[87], and hippocampal sclerosis (graded 0–5, with severe pathology defined as a score of 5 in CA1 or subiculum)[95]. In addition, a binary variable representing Parkinson's disease (PD) pathology was defined based on the presence of Lewy bodies and the severity of neuronal loss in the substantia nigra.

Vascular pathologies: Cerebral infarctions were documented separately as gross and microinfarcts based on standardized neuropathologic examination[96,97]. Arteriolosclerosis and cerebral atherosclerosis were graded semi-quantitatively (0–3) based on severity observed in histologic and gross pathology assessments, respectively[98,99]. Cerebral amyloid angiopathy (CAA) was quantified in four cortical regions by scoring meningeal and parenchymal vessel involvement from 0 to 4, and averaging across regions. CAA severity was also categorized semi-quantitatively using established cutpoints[100,101].

Motor and parkinsonian function: Motor function was evaluated annually using ten performance tests encompassing dexterity, gait, strength, and balance, and a derived slope representing motor decline was estimated using a linear mixed effects model, controlling for age at baseline, sex, and years of education[102]. Performance scores were standardized based on cohort baseline means and combined to generate composite measures of global motor function, gait, dexterity, and hand strength[103]. Parkinsonian signs (bradykinesia, rigidity, tremor, gait disturbance) were assessed using modified motor items from the Unified

Parkinson's Disease Rating Scale. Scores were scaled from 0 to 100, with higher scores reflecting more severe impairment[104]. Longitudinal trajectories of global parkinsonian score were derived from the average of the four domain scores modeled over time, adjusted for baseline age, sex, and years of education[104,105].

Disability and frailty: Disability was assessed using three measures: basic activities of daily living (ADL), instrumental activities of daily living (IADL), and mobility disability. ADL and IADL scores reflected the number of activities with which participants reported needing assistance[106,107], while mobility disability was assessed with the Rosow-Breslau scale[108]. Frailty measure reflects multisystem vulnerability, based on four components: grip strength, timed walk, body composition (BMI), and fatigue. The measure was constructed by converting each raw component score into a z-score using baseline mean and standard deviation values from all participants[109,110].

Psychological factors: Depressive symptoms were measured annually using a 10-item modified Center for Epidemiologic Studies Depression Scale (CES-D). Each participant's mean score across visits was used to reflect their enduring disposition toward depressive symptoms[111–113]. A clinical diagnosis of major depressive disorder was rendered by a physician based on structured interviews and DSM-III-R criteria. A dichotomous indicator of probable depression was derived by collapsing the diagnostic scale into highly probable vs. other scores[114].

Genetics: APOE genotyping was derived from DNA extracted from either brain tissue or peripheral blood mononuclear cells, followed by high-throughput sequencing targeting codon 112 (position 3937) and codon 158 (position 4075) of exon 4 of the APOE gene[115]. A dichotomized score was created where samples were categorized into those with the APOEε4 allele and those without. TOMM40'523 genotypes (rs10524523; chr19:44,899,792–44,899,826, human genome reference assembly GRCh38/hg38), a homopolymer length polymorphism (poly-T), at intron 6 of the TOMM40 gene, were determined by PCR amplification followed by Sanger Sequencing as previously described[115,116]. Allele lengths were classified according to the number of poly-T repeats: short alleles (S) have fewer than 20 repeats, long alleles (L) have 20 to 29 repeats, and very long alleles (VL) have 30 or more repeats[115]. A dichotomized score was created where samples were categorized into those with a least one long allele and those without.

H3K9ac histone acetylation (ChIP-seq): H3K9ac histone acetylation data generation has been previously described[71]. In essence, gray matter tissue from ROS/MAP subjects was dissected from the Dorsolateral Prefrontal Cortex (DLPFC), homogenized, incubated with H3K9ac antibody using the ChIP Dilution Buffer and purified with protein A Sepharose beads. The resulting DNA was used for Illumina library construction and then sequenced on the Illumina HiSeq (36 bp single-end reads). Low-quality samples were removed, and H3K9ac domains were defined by calculating all genomic regions that were detected as a peak in at least 15% of the samples. Regions neighboring within 100 bp were merged, and very small regions of less than 100 bp were removed, resulting in 26,384 H3K9ac domains

from 632 samples. The Data was normalized, log-transformed, and technical confounders removed.

Transcriptomics (RNA-Seq): RNA was extracted from gray matter tissues from 3 regions: anterior caudate (AC), posterior cingulate gyrus (PCG) and DLPFC. Extraction was performed using the Qiagen miRNeasy kit and DNase treatment, with quality assessed by Bioanalyzer and quantity by Nanodrop. Three batches of RNA-Seq libraries were prepared from different brain regions using distinct protocols: Batch 1 (DLPFC) used the dUTP method with poly-A selection; Batch 2 (DLPFC, PCC, AC) used KAPA's stranded kit with ribosomal depletion; Batch 3 (DLPFC, PCC) involved Chemagic extraction, RiboGold depletion, and a modified TruSeq protocol. Sequencing was performed using Illumina platforms: Batch 1 used HiSeq (101bp PE reads, 50–150M reads/sample), Batch 2 used NovaSeq6000 (2×100bp, 30M reads/sample), and Batch 3 used NovaSeq6000 (2×150bp, 40–50M reads/sample). A harmonized data processing was applied to all three regions as previously described[72]. Briefly, after QC, quantification was performed using Kallisto (v0.46). Gene level counts were filtered to keep only genes with at least 10 counts in >50% of samples. Counts were adjusted by GC content and gene length using CQN (conditional quantile normalization) and converted to log2-CPM (counts per million) followed by quantile normalization using the limma voom function. Finally, a linear model was applied to remove major technical confounding factors matrics as reported by Picard and Kallisto. The final resulting matrices included 17,294 mRNAs for 721 AC, 656 PCG and 1,210 DLPFC samples.

Single-nuclei RNA sequencing (snRNA-Seq): The protocol for the isolation of nuclei from frozen postmortem brain tissue, RNA sequencing and data processing was previously published[36,117]. Briefly, 50 to 100mg of DLPFC gray matter from ROS/MAP subjects were homogenized, the cells nuclei filtered and sequenced by 10X Single Cell RNA-Seq Platform allowing the creation of 128 pooled libraries. Preprocessing of each library involved alignment, normalization, cell clustering, background noise removal, de-multiplexing, classification of nuclei for cell types, removal of low-quality nuclei and detection of doublets for removal. Following major cell-type classification, a sub-clustering analysis was performed resulting in 96 cell subpopulations measured for 424 donors[36,117]. We considered the relative fractions of each subpopulation by donor in our downstream analysis.

TMT-MS-based quantitative proteomics: Methods employed data generations and processing has also been detailed in a prior publication[73]. Briefly, DLPFC tissue samples were homogenized, digested and purified. Then the resulting peptides were labeled using the TMT 10-plex kit and fractionated by high pH. All fractions were resuspended and analyzed by liquid chromatography coupled to tandem Mass Spectrometry and the resulting spectra were searched against the UniProtKB human proteome database. After removing technical variance 8,425 proteins from 596 donors were used in our downstream analysis.

Metabolomics profiling: Metabolomics data generation and processing have been previously described[39]. Briefly, tissues from DLPFC subjects were homogenized, the

proteins were removed, and the resulting extract was divided into four fractions: two for ultra-high performance liquid chromatography-tandem mass spectrometry (UPLC-MS/MS) with positive ionization, one for UPLC-MS/MS with negative ionization, and one for UPLC-MS/MS polar platform with negative ionization. Peaks were identified and quantified, and technical confounders were excluded. The process yielded 667 metabolites from 500 subjects, which were used in our analysis.

METHOD DETAILS

Multi-omics factor analysis (MOFA): MOFA is a computational framework for unsupervised multi-omics integration analysis capable of capturing both biological and technical sources of variability across multiple omics datasets. Intuitively, it can be viewed as a generalization of principal component analysis to multi-omics data. The method employs Bayesian inference and probabilistic models to derive a low-dimensional representation of multiple data modalities in terms of latent factors that capture major sources of variation within the data. For each factor, MOFA estimates feature weights (or loadings), which quantify the magnitude and direction of each feature's contribution to that factor, granting a mapping between the high and low-dimensional spaces. Furthermore, MOFA disentangles to what extent each factor is unique to a single data modality or is manifested in multiple modalities, thereby revealing shared axes of variation between the different omics layers. The model is able to efficiently handle missing values and is robust to scalability of large volumes of data[32].

MOFA implementation: The analysis was conducted utilizing the 'MOFA2' package[32], version 1.10.0, in R. Initially, to conform the data to one of MOFA's input data formats, we transformed each omics dataset into matrices, with the same number of samples (missing samples were added as NAs) arranged in the same sequence. To perform model training, we used MOFA's default data model and training settings, with the exception of two adjustments: the number of factors was set to 50 (default is 15), and the data was scaled to have the same unit variance (default is FALSE).

Factor selection: In order to retain only the factors associated with AD, we conducted a regression analysis using the 50 factors generated by MOFA and the 38 aforementioned phenotypes. The tests were controlled for sex, age at death and years of education, with P-values corrected using the False Discovery Rate (FDR) method to address multiple comparisons. For each phenotype, we employed regression techniques tailored to the corresponding data type. For continuous outcomes, linear regressions were performed in R using the 'lm' function from the 'Base' package, version 4.3.1[74]. For categorial outcomes, we employed a Bayesian logistic regression approach using the 'bayesglm' function from the 'arm' package, version 1.14.4 in R[75]. For ordinal outcomes, we used the cumulative link model implemented in R with the 'clm' function from the 'ordinal' package, version 2023.12.4[76]. We selected the factors that exhibit statistically significant association (FDR < 0.05) with either NIA-Reagan score or clinical diagnosis of AD. Nine factors were selected for downstream analysis: factors 1, 2, 3, 4, 8, 12, 14, 26, 48.

Gene Set Enrichment Analysis (GSEA): For GSEA, we used the 'fgsea' package version 1.26.0[77] in R. GSEA was performed for each factor by running the 'fgseaMultilevel' function with default settings across all omics. Gene sets for Reactome[118] and Gene Ontology[119,120] were sourced from the 'msigdbr' package version 7.5.1[78]. For proteomics, gene background was restricted to all proteins measured in the TMT experiment. The metabolomics enrichment set was adapted from[39], and the cell subpopulation type set was obtained from[36]. Additionally, we incorporated transcriptomic co-expression modules[23], proteomics co-expression modules[24], an AD genome-wide association study gene set[121], the plaque-induced genes (PIGs) co-expression network[122], and two AD-associated microglia subpopulations: DAM[123] and HAM[124]. We also used the function 'reduceSimMatrix' (threshold varying from 0.9 to 0.7) from the 'rrvgo' package version 1.12.2[79] to reduce term redundancy. P-values corrected using the False Discovery Rate (FDR) method to address multiple comparisons.

Robustness and reproducibility: To evaluate the robustness and stability of the MOFA model, we conducted a series of computational experiments on subsets of the original data. We systematically varied the number of samples and the completeness of the data to assess their impact on the inferred factors. The data was split into training and testing sets using a stratified sampling approach based on the number of omics available for each sample. We generated datasets with 90%, 80%, 70%, 60%, and 50% of the original samples. Additionally, we created non-overlapping 50/50 splits of the data to assess the reproducibility of the factors in independent datasets. For each data subset, we trained a MOFA model with 50 latent factors using a fast convergence mode. Fifteen random splits were performed in each subsampling. In addition, we evaluated the impact of changing the omics input by restricting the analysis to samples with increasing numbers of omics layers (i.e., subsets with at least 2, 3, or 4 omics layers), and by completely removing one omics layer (i.e., setting 'NA' to all features of the removed layer). To compare the factors inferred from different MOFA models, we matched factors between two models based on the Pearson correlation of their feature weights. The optimal matching of factors was determined using a best-matching approach. For each pair of matched factors, we calculated the correlation of their weights and corrected for potential sign flips. The mean correlation across all matched factors was used as a measure of the overall similarity between the two MOFA models.

Cluster analysis with speakeasy: To categorize the ROS/MAP participants into AD subtypes, we used the 50 factors generated by MOFA as input for the Speakeasy algorithm[46]. This method is well suited for the analysis, as it has been previously tested with the Lancichinetti-Fortunato-Radicchi (LFR) benchmarks, a widely used test for evaluating overlapping and non-overlapping clustering methods across different types of data[125]. Speakeasy has been used before with RNA-Seq data[23] and proteomics data[126] and has been extensively compared against other clustering methods, such as normalized mutual information and weighted gene co-expression networks. Speakeasy does not require extensive parameterization; however, we used the default settings of 100 permutations, Pearson correlation, and a minimum of 30 individuals per cluster. Finally, we analyzed the

results at three levels of clustering, with larger clusters at level 1 and more refined clusters containing fewer individuals at levels 2 and 3.

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical analyses and software used in the study are detailed in the “Method details” and “Key resources table” sections. Specific statistical tests used for each experiment, the exact value of n, and what n represents are described in the corresponding figure legends and “Results” section. Adjustments for multiple comparisons were applied where appropriate.

ADDITIONAL RESOURCES

Source code and derived data generated by this study are available on the GitHub repository: https://github.com/RushAlz/ROSMAP_MOFA_public_release.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We thank the participants in the ROS/MAP cohorts for their essential contributions and gift to these projects, as well as the investigators and staff at the Rush Alzheimer’s Disease Center. This work has been supported by the following National Institute on Aging (NIA) grants: P30AG10161 (D.A.B.), P30AG72975 (J.A.S.), R01AG15819 (D.A.B.), R01AG17917 (D.A.B.), U01AG079847 (C.G.), U01AG46152 (P.L.D.J. and D.A.B.), and U01AG61356 (P.L.D.J. and D.A.B.). L.P.S. received a fellowship from the Brazilian agency CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior). The data available in the AD Knowledge Portal would not be possible without the participation of research volunteers and the contribution of data by collaborating researchers. The results published here are in whole or in part based on data obtained from the AD Knowledge Portal (https://adknowledgeportal.org). Study data were provided by the Rush Alzheimer’s Disease Center, Rush University Medical Center, Chicago, IL, USA. Data collection was supported through funding by NIA grants P30AG10161 (ROS), R01AG15819 (ROS/MAP; genomics and RNA-seq), R01AG17917 (MAP), RC2AG036547 (H3K9Ac), R01AG36836 (RNA-seq), RF1AG57473 (single-nucleus RNA-seq), U01AG32984 (genomic sequencing), U01AG46161 (TMT proteomics), U01AG61356 (whole-genome sequencing, ROS/MAP AMP-AD), P30AG072975, the Illinois Department of Public Health (ROS/MAP), and the Translational Genomics Research Institute (genomics). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The ROS and MAP studies were approved by an Institutional Review Board (IRB) of Rush University Medical Center, Chicago, IL, USA. All participants agreed to an annual clinical evaluation and signed an informed consent and an Anatomic Gift Act agreeing to postmortem brain donation. All procedures performed in studies involving human participants were in accordance with the ethical standards of the Institutional Review Board of Rush University Medical Center and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. Each participant signed an informed consent form to participate in the study.

REFERENCES

1. Selkoe DJ, and Hardy J (2016). The amyloid hypothesis of Alzheimer’s disease at 25 years. EMBO Mol. Med. 8, 595–608.
2. Zhang J, Zhang Y, Wang J, Xia Y, Zhang J, and Chen L (2024). Recent advances in Alzheimer’s disease: Mechanisms, clinical trials and new drug development strategies. Signal Transduct. Target. Ther. 9, 211. [PubMed: 39174535]
3. Du X, Wang X, and Geng M (2018). Alzheimer’s disease hypothesis and related therapies. Transl. Neurodegener. 7, 2–7. [PubMed: 29423193]
4. Murdock MH, and Tsai L-H (2023). Insights into Alzheimer’s disease from single-cell genomic approaches. Nat. Neurosci. 26, 181–195. [PubMed: 36593328]

5. Wong MW, Braidy N, Poljak A, Pickford R, Thambisetty M, and Sachdev PS (2017). Dysregulation of lipids in Alzheimer's disease and their role as potential biomarkers. Alzheimer's Dement. 13, 810–827. [PubMed: 28242299]
6. Leng F, and Edison P (2021). Neuroinflammation and microglial activation in Alzheimer disease: where do we go from here? Nat. Rev. Neurol. 17, 157–172. [PubMed: 33318676]
7. Yan X, Hu Y, Wang B, Wang S, and Zhang X (2020). Metabolic Dysregulation Contributes to the Progression of Alzheimer's Disease. Front. Neurosci. 14, 530219. [PubMed: 33250703]
8. Clark C, Dayon L, Masoodi M, Bowman GL, and Popp J (2021). An integrative multi-omics approach reveals new central nervous system pathway alterations in Alzheimer's disease. Alzheimers Res. Ther. 13, 71. [PubMed: 33794997]
9. St John-Williams L, Blach C, Toledo JB, Rotroff DM, Kim S, Klavins K, Baillie R, Han X, Mahmoudiandehkordi S, Jack J, et al. (2017). Targeted metabolomics and medication classification data from participants in the ADNI1 cohort. Sci. Data 4, 170140. [PubMed: 29039849]
10. Wang M, Beckmann ND, Roussos P, Wang E, Zhou X, Wang Q, Ming C, Neff R, Ma W, Fullard JF, et al. (2018). The Mount Sinai cohort of large-scale genomic, transcriptomic and proteomic data in Alzheimer's disease. Sci. Data 5, 180185. [PubMed: 30204156]
11. De Jager PL, Ma Y, McCabe C, Xu J, Vardarajan BN, Felsky D, Klein H-U, White CC, Peters MA, Lodgson B, et al. (2018). A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research. Sci. Data 5, 180142. [PubMed: 30084846]
12. Greenwood AK, Montgomery KS, Kauer N, Woo KH, Leanza ZJ, Poehlman WL, Gockley J, Sieberts SK, Bradic L, Logsdon BA, et al. (2020). The AD Knowledge Portal: A Repository for Multi-Omic Data on Alzheimer's Disease and Aging. Curr. Protoc. Hum. Genet. 108, e105. [PubMed: 33085189]
13. De Jager PL, Srivastava G, Lunnon K, Burgess J, Schalkwyk LC, Yu L, Eaton ML, Keenan BT, Ernst J, McCabe C, et al. (2014). Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. Nat. Neurosci. 17, 1156–1163. [PubMed: 25129075]
14. Allen M, Burgess JD, Ballard T, Serie D, Wang X, Younkin CS, Sun Z, Kouri N, Baheti S, Wang C, et al. (2016). Gene expression, methylation and neuropathology correlations at progressive supranuclear palsy risk loci. Acta Neuropathol. 132, 197–211. [PubMed: 27115769]
15. Carrasquillo MM, Allen M, Burgess JD, Wang X, Strickland SL, Aryal S, Siuda J, Kachadoorian ML, Medway C, Younkin CS, et al. (2017). A candidate regulatory variant at the TREM gene cluster associates with decreased Alzheimer's disease risk and increased TREML1 and TREM2 brain gene expression. Alzheimer's Dement. 13, 663–673. [PubMed: 27939925]
16. Allen M, Wang X, Burgess JD, Watzlawik J, Serie DJ, Younkin CS, Nguyen T, Malphrus KG, Lincoln S, Carrasquillo MM, et al. (2018). Conserved brain myelination networks are altered in Alzheimer's and other neurodegenerative diseases. Alzheimer's Dement. 14, 352–366. [PubMed: 29107053]
17. Allen M, Wang X, Serie DJ, Strickland SL, Burgess JD, Koga S, Younkin CS, Nguyen TT, Malphrus KG, Lincoln SJ, et al. (2018). Divergent brain gene expression patterns associate with distinct cell-specific tau neuropathology traits in progressive supranuclear palsy. Acta Neuropathol. 136, 709–727. [PubMed: 30136084]
18. Nho K, Nudelman K, Allen M, Hodges A, Kim S, Risacher SL, Apostolova LG, Lin K, Lunnon K, Wang X, et al. (2020). Genome-wide transcriptome analysis identifies novel dysregulated genes implicated in Alzheimer's pathology. Alzheimer's Dement. 16, 1213–1223. [PubMed: 32755048]
19. Ma Y, Dammer EB, Felsky D, Duong DM, Klein H-U, White CC, Zhou M, Logsdon BA, McCabe C, Xu J, et al. (2021). Atlas of RNA editing events affecting protein expression in aged and Alzheimer's disease human brain tissue. Nat. Commun. 12, 7035. [PubMed: 34857756]
20. McKenzie AT, Moyon S, Wang M, Katsyv I, Song W-M, Zhou X, Dammer EB, Duong DM, Aaker J, Zhao Y, et al. (2017). Multiscale network modeling of oligodendrocytes reveals molecular components of myelin dysregulation in Alzheimer's disease. Mol. Neurodegener. 12, 82. [PubMed: 29110684]
21. Beckmann ND, Lin W-J, Wang M, Cohain AT, Charney AW, Wang P, Ma W, Wang Y-C, Jiang C, Audrain M, et al. (2020). Multiscale causal networks identify VGF as a key regulator of Alzheimer's disease. Nat. Commun. 11, 3942. [PubMed: 32770063]

22. Seyfried NT, Dammer EB, Swarup V, Nandakumar D, Duong DM, Yin L, Deng Q, Nguyen T, Hales CM, Wingo T, et al. (2017). A Multi-network Approach Identifies Protein-Specific Co-expression in Asymptomatic and Symptomatic Alzheimer's Disease. Cell Syst. 4, 60–72.e4. [PubMed: 27989508]
23. Mostafavi S, Gaiteri C, Sullivan SE, White CC, Tasaki S, Xu J, Taga M, Klein H-U, Patrick E, Komashko V, et al. (2018). A molecular network of the aging human brain provides insights into the pathology and cognitive decline of Alzheimer's disease. Nat. Neurosci. 21, 811–819. [PubMed: 29802388]
24. Johnson ECB, Carter EK, Dammer EB, Duong DM, Gerasimov ES, Liu Y, Liu J, Betarbet R, Ping L, Yin L, et al. (2022). Large-scale deep multi-layer analysis of Alzheimer's disease brain reveals strong proteomic disease-related changes not observed at the RNA level. Nat. Neurosci. 25, 213–225. [PubMed: 35115731]
25. Kodam P, Sai Swaroop R, Pradhan SS, Sivaramakrishnan V, and Vadrevu R (2023). Integrated multi-omics analysis of Alzheimer's disease shows molecular signatures associated with disease progression and potential therapeutic targets. Sci. Rep. 13, 3695. [PubMed: 36879094]
26. Zhang J, Sun X, Jia X, Sun B, Xu S, Zhang W, and Liu Z (2024). Integrative multi-omics analysis reveals the critical role of the PBXIP1 gene in Alzheimer's disease. Aging Cell 23, e14044. [PubMed: 37984333]
27. Liu N, Wu Y, Cheng W, Wu Y, Wang L, and Zhuang L (2021). Identification of novel prognostic biomarkers by integrating multi-omics data in gastric cancer. BMC Cancer 21, 460. [PubMed: 33902514]
28. Jin X, Mei Y, Yang P, Huang R, Zhang H, Wu Y, Wang M, He X, Jiang Z, Zhu W, and Wang L (2024). Prioritization of therapeutic targets for cancers using integrative multi-omics analysis. Hum. Genomics 18, 42. [PubMed: 38659038]
29. Park J-C, Barahona-Torres N, Jang S-Y, Mok KY, Kim HJ, Han S-H, Cho K-H, Zhou X, Fu AKY, Ip NY, et al. (2022). Multi-Omics-Based Autophagy-Related Untypical Subtypes in Patients with Cerebral Amyloid Pathology. Adv. Sci. 9, e2201212.
30. Iturria-Medina Y, Adewale Q, Khan AF, Ducharme S, Rosa-Neto P, O'Donnell K, Petyuk VA, Gauthier S, De Jager PL, Breitner J, and Bennett DA (2022). Unified epigenomic, transcriptomic, proteomic, and metabolomic taxonomy of Alzheimer's disease progression and heterogeneity. Sci. Adv. 8, eabo6764. [PubMed: 36399579]
31. Iturria-Medina Y, Poole VN, Zammit AR, Yu L, Tasaki S, Hong JH, Lopes K.d.P, Batalha C., Ridwan AR, Vialle RA, et al. (2025). Translating the Post-Mortem Brain Multi-Omics Molecular Taxonomy of Alzheimer's Dementia to Living Humans. Preprint at bioRxiv, 2025. 03.20.644323. 10.1101/2025.03.20.644323.
32. Argelaguet R, Velten B, Arnol D, Dietrich S, Zenz T, Marioni JC, Buettner F, Huber W, and Stegle O (2018). Multi-Omics Factor Analysis-a framework for unsupervised integration of multi-omics data sets. Mol. Syst. Biol. 14, e8124. [PubMed: 29925568]
33. Argelaguet R, Arnol D, Bredikhin D, Deloro Y, Velten B, Marioni JC, and Stegle O (2020). MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. Genome Biol. 21, 111. [PubMed: 32393329]
34. Strefeler A, Jan M, Quadroni M, Teav T, Rosenberg N, Chatton J-Y, Guex N, Gallart-Ayala H, and Ivanisevic J (2023). Molecular insights into sex-specific metabolic alterations in Alzheimer's mouse brain using multi-omics approach. Alzheimers Res. Ther. 15, 8. [PubMed: 36624525]
35. Bennett DA, Buchman AS, Boyle PA, Barnes LL, Wilson RS, and Schneider JA (2018). Religious orders study and rush memory and aging project. J. Alzheimers Dis. 64, S161–S189. [PubMed: 29865057]
36. Green GS, Fujita M, Yang H-S, Taga M, Cain A, McCabe C, Comandante-Lou N, White CC, Schmidtner AK, Zeng L, et al. (2024). Cellular communities reveal trajectories of brain ageing and Alzheimer's disease. Nature 633, 634–645. [PubMed: 39198642]
37. Bionetworks S (2023). Agora AD Knowledge Portal (Sage Bionetworks). 10.57718/AGORA-ADKNOWLEDGEPORTAL.

38. Houten SM, Violante S, Ventura FV, and Wanders RJA (2016). The Biochemistry and Physiology of Mitochondrial Fatty Acid β-Oxidation and Its Genetic Disorders. Annu. Rev. Physiol. 78, 23–44. [PubMed: 26474213]
39. Batra R, Arnold M, Wörheide MA, Allen M, Wang X, Blach C, Levey AI, Seyfried NT, Ertekin-Taner N, Bennett DA, et al. (2022). The landscape of metabolic brain alterations in Alzheimer's disease. Alzheimer's Dement. 10.1002/alz.12714.
40. Aaes TL, Burgoa Cardás J, and Ravichandran KS (2023). Defining solute carrier transporter signatures of murine immune cell subsets. Front. Immunol. 14, 1276196. [PubMed: 38077407]
41. Chen R, and Chen L (2022). Solute carrier transporters: emerging central players in tumour immunotherapy. Trends Cell Biol. 32, 186–201. [PubMed: 34511324]
42. Iwasaki M, Homma S, Hishiya A, Dolezal SJ, Reed JC, and Takayama S (2007). BAG3 regulates motility and adhesion of epithelial cancer cells. Cancer Res. 67, 10252–10259. [PubMed: 17974966]
43. Marzullo L, Turco MC, and De Marco M (2020). The multiple activities of BAG3 protein: Mechanisms. Biochim. Biophys. Acta. Gen. Subj. 1864, 129628. [PubMed: 32360144]
44. Stürner E, and Behl C (2017). The Role of the Multifunctional BAG3 Protein in Cellular Protein Quality Control and in Disease. Front. Mol. Neurosci. 10, 177. [PubMed: 28680391]
45. Farid MM, Yang X, Kuboyama T, and Tohda C (2020). Trigonelline recovers memory function in Alzheimer's disease model mice: evidence of brain penetration and target molecule. Sci. Rep. 10, 16424. [PubMed: 33009465]
46. Gaiteri C, Chen M, Szymanski B, Kuzmin K, Xie J, Lee C, Blanche T, Chaibub Neto E, Huang S-C, Grabowski T, et al. (2015). Identifying robust communities and multi-community nodes by combining top-down and bottom-up approaches to clustering. Sci. Rep. 5, 16361. [PubMed: 26549511]
47. Si Z-Z, Zou C-J, Mei X, Li X-F, Luo H, Shen Y, Hu J, Li X-X, Wu L, and Liu Y (2023). Targeting neuroinflammation in Alzheimer's disease: from mechanisms to clinical applications. Neural Regen. Res. 18, 708–715. [PubMed: 36204826]
48. Onyango IG, Jauregui GV, Ćarná M, Bennett JP Jr., and Stokin GB (2021). Neuroinflammation in Alzheimer's Disease. Biomedicines 9. 10.3390/biomedicines9050524.
49. Rajesh Y, and Kanneganti T-D (2022). Innate Immune Cell Death in Neuroinflammation and Alzheimer's Disease. Cells 11, 1885. 10.3390/cells11121885. [PubMed: 35741014]
50. Bamburg JR, and Bloom GS (2009). Cytoskeletal pathologies of Alzheimer disease. Cell Motil Cytoskeleton 66, 635–649. [PubMed: 19479823]
51. Barmaki H, Nourazarian A, and Khaki-Khatibi F (2023). Proteostasis and neurodegeneration: a closer look at autophagy in Alzheimer's disease. Front. Aging Neurosci. 15, 1281338. [PubMed: 38020769]
52. Scheff SW, Price DA, Schmitt FA, DeKosky ST, and Mufson EJ (2007). Synaptic alterations in CA1 in mild Alzheimer disease and mild cognitive impairment. Neurology 68, 1501–1508. [PubMed: 17470753]
53. Ownby RL, Crocco E, Acevedo A, John V, and Loewenstein D (2006). Depression and risk for Alzheimer disease: systematic review, meta-analysis, and metaregression analysis. Arch. Gen. Psychiatry 63, 530–538. [PubMed: 16651510]
54. Boyle PA, Yu L, Leurgans SE, Wilson RS, Brookmeyer R, Schneider JA, and Bennett DA (2019). Attributable risk of Alzheimer's dementia attributed to age-related neuropathologies. Ann. Neurol. 85, 114–124. [PubMed: 30421454]
55. Eteleeb AM, Novotny BC, Tarraga CS, Sohn C, Dhungel E, Brase L, Nallapu A, Buss J, Farias F, Bergmann K, et al. (2024). Brain high-throughput multi-omics data reveal molecular heterogeneity in Alzheimer's disease. PLoS Biol. 22, e3002607. [PubMed: 38687811]
56. Li M-M, Yang Y-X, Huang Y-L, Wu S-J, Huang W-L, Ye L-C, and Xu Y-Y (2025). Programmed cell death signatures-driven microglial transformation in Alzheimer's disease: single-cell transcriptomics and functional validation. Front. Immunol. 16, 1610717. [PubMed: 40787447]
57. Lei Z, Brizzee C, and Johnson GVW (2015). BAG3 facilitates the clearance of endogenous tau in primary neurons. Neurobiol. Aging 36, 241–248. [PubMed: 25212465]

58. Brenner CM, Choudhary M, McCormick MG, Cheung D, Landesberg GP, Wang J-F, Song J, Martin TG, Cheung JY, Qu H-Q, et al. (2023). BAG3: Nature's Quintessential Multi-Functional Protein Functions as a Ubiquitous Intra-Cellular Glue. Cells 12, 937. 10.3390/cells12060937. [PubMed: 36980278]
59. Wang J, Tomar D, Martin TG, Dubey S, Dubey PK, Song J, Landesberg G, McCormick MG, Myers VD, Merali S, et al. (2023). Bag3 Regulates Mitochondrial Function and the Inflammasome Through Canonical and Noncanonical Pathways in the Heart. JACC, Basic Transl. Sci. 8, 820–839. [PubMed: 37547075]
60. Fontanella B, Birolo L, Infusini G, Cirulli C, Marzullo L, Pucci P, Turco MC, and Tosco A (2010). The co-chaperone BAG3 interacts with the cytosolic chaperonin CCT: new hints for actin folding. Int. J. Biochem. Cell Biol. 42, 641–650. [PubMed: 20018251]
61. Franceschelli S, Bruno AP, Festa M, Falco A, Gionti E, d'Avenia M, De Marco M, Basile A, Iorio V, Marzullo L, et al. (2018). BAG3 Protein Is Involved in Endothelial Cell Response to Phenethyl Isothiocyanate. Oxid. Med. Cell. Longev. 2018, 5967890. [PubMed: 29955247]
62. Ying Z-M, Lv Q-K, Yao X-Y, Dong A-Q, Yang Y-P, Cao Y-L, Wang F, Gong A-P, and Liu C-F (2022). BAG3 promotes autophagy and suppresses NLRP3 inflammasome activation in Parkinson's disease. Ann. Transl. Med. 10, 1218. [PubMed: 36544667]
63. Franceschelli S, Rosati A, Lerose R, De Nicola S, Turco MC, and Pascale M (2008). Bag3 gene expression is regulated by heat shock factor 1. J. Cell. Physiol. 215, 575–577. [PubMed: 18286539]
64. Meriin AB, Narayanan A, Meng L, Alexandrov I, Varelas X, Cissé II, and Sherman MY (2018). Hsp70-Bag3 complex is a hub for proteotoxicity-induced signaling that controls protein aggregation. Proc. Natl. Acad. Sci. USA 115, E7043–E7052. [PubMed: 29987014]
65. Hiebel C, Stürner E, Hoffmeister M, Tascher G, Schwarz M, Nagel H, Behrends C, Münch C, and Behl C (2020). BAG3 Proteomic Signature under Proteostasis Stress. Cells 9, 2416. 10.3390/cells9112416. [PubMed: 33158300]
66. Augur ZM, Fogo GM, Benoit CR, Terzioglu G, Murphy ZR, Arbery MR, Comandante-Lou N, Duong DM, Seyfried NT, De Jager PL, and Young-Pearse TL (2025). BAG3 coordinates astrocytic proteostasis of Alzheimer's disease-linked proteins via proteasome, autophagy, and retromer complex interactions. Preprint at bioRxiv. 10.1101/2025.09.20.677505.
67. Jinwal UK, Koren J 3rd, Borysov SI, Schmid AB, Abisambra JF, Blair LJ, Johnson AG, Jones JR, Shults CL, O'Leary JC 3rd., et al. (2010). The Hsp90 cochaperone, FKBP51, increases Tau stability and polymerizes microtubules. J. Neurosci. 30, 591–599. [PubMed: 20071522]
68. Blair LJ, Nordhues BA, Hill SE, Scaglione KM, O'Leary JC 3rd, Fontaine SN, Breydo L, Zhang B, Li P, Wang L, et al. (2013). Accelerated neurodegeneration through chaperone-mediated oligomerization of tau. J. Clin. Investig. 123, 4158–4169. [PubMed: 23999428]
69. Bozdagi O, Rich E, Tronel S, Sadahiro M, Patterson K, Shapiro ML, Alberini CM, Huntley GW, and Salton SRJ (2008). The neurotrophin-inducible gene Vgf regulates hippocampal function and behavior through a brain-derived neurotrophic factor-dependent mechanism. J. Neurosci. 28, 9857–9869. [PubMed: 18815270]
70. Buettner F, and Theis FJ (2012). A novel approach for resolving differences in single-cell gene expression patterns from zygote to blastocyst. Bioinformatics 28, i626–i632. [PubMed: 22962491]
71. Klein H-U, McCabe C, Gjoneska E, Sullivan SE, Kaskow BJ, Tang A, Smith RV, Xu J, Pfenning AR, Bernstein BE, et al. (2019). Epigenome-wide study uncovers large-scale changes in histone acetylation driven by tau pathology in aging and Alzheimer's human brains. Nat. Neurosci. 22, 37–46. [PubMed: 30559478]
72. Tasaki S, Xu J, Avey DR, Johnson L, Petyuk VA, Dawe RJ, Bennett DA, Wang Y, and Gaiteri C (2022). Inferring protein expression changes from mRNA in Alzheimer's dementia using deep neural networks. Nat. Commun. 13, 655. [PubMed: 35115553]
73. Johnson ECB, Dammer EB, Duong DM, Ping L, Zhou M, Yin L, Higginbotham LA, Guajardo A, White B, Troncoso JC, et al. (2020). Large-scale proteomic analysis of Alzheimer's disease brain and cerebrospinal fluid reveals early changes in energy metabolism associated with microglia and astrocyte activation. Nat. Med. 26, 769–780. [PubMed: 32284590]

74. R Development Core Team (2023). R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing).

75. Gelman A, and Su Y-S (2007). arm: Data Analysis Using Regression and Multilevel/Hierarchical Models. R package. 10.32614/CRAN.package.arm.

76. Christensen RHB (2010). ordinal: Regression Models for Ordinal Data. R package. 10.32614/CRAN.package.ordinal.

77. Korotkevich G, Sukhov V, Budin N, Shpak B, Artyomov MN, and Sergushichev A (2016). Fast gene set enrichment analysis. Preprint at bioRxiv. 10.1101/060012.

78. Dolgalev I (2018). msigdbr: MSigDB Gene Sets for Multiple Organisms in a Tidy Data Format. R package. 10.32614/CRAN.package.msigdbr.

79. Sayols S (2023). rrvgo: a Bioconductor package for interpreting lists of Gene Ontology terms. MicroPubl. Biol. 2023. 10.17912/micropub.biology.000811.

80. Bennett DA, Schneider JA, Bienias JL, Evans DA, and Wilson RS (2005). Mild cognitive impairment is related to Alzheimer disease pathology and cerebral infarctions. Neurology 64, 834–841. [PubMed: 15753419]

81. Bennett DA, Schneider JA, Arvanitakis Z, and Wilson RS (2012). Overview and findings from the religious orders study. Curr. Alzheimer Res. 9, 628–645. [PubMed: 22471860]

82. Bennett DA, Schneider JA, Aggarwal NT, Arvanitakis Z, Shah RC, Kelly JF, Fox JH, Cochran EJ, Arends D, Treinkman AD, and Wilson RS (2006). Decision rules guiding the clinical diagnosis of Alzheimer's disease in two community-based cohort studies compared to standard practice in a clinic-based cohort study. Neuroepidemiology 27, 169–176. [PubMed: 17035694]

83. Wilson RS, Boyle PA, Yang J, James BD, and Bennett DA (2015). Early life instruction in foreign language and music and incidence of mild cognitive impairment. Neuropsychology 29, 292–302. [PubMed: 25110933]

84. De Jager PL, Shulman JM, Chibnik LB, Keenan BT, Raj T, Wilson RS, Yu L, Leurgans SE, Tran D, Aubin C, et al. (2012). A genome-wide scan for common variants affecting the rate of age-related cognitive decline. Neurobiol. Aging 33. 1017.e1–15.

85. Boyle PA, Wang T, Yu L, Wilson RS, Dawe R, Arfanakis K, Schneider JA, and Bennett DA (2021). To what degree is late life cognitive decline driven by age-related neuropathologies? Brain 144, 2166–2175. [PubMed: 33742668]

86. Bennett DA, Wilson RS, Schneider JA, Evans DA, Beckett LA, Aggarwal NT, Barnes LL, Fox JH, and Bach J (2002). Natural history of mild cognitive impairment in older persons. Neurology 59, 198–205. [PubMed: 12136057]

87. Schneider JA, Arvanitakis Z, Bang W, and Bennett DA (2007). Mixed brain pathologies account for most dementia cases in community-dwelling older persons. Neurology 69, 2197–2204. [PubMed: 17568013]

88. Vialle RA, de Paiva Lopes K, Li Y, Ng B, Schneider JA, Buchman AS, Wang Y, Farfel JM, Barnes LL, Wingo AP, et al. (2025). Structural variants linked to Alzheimer's disease and other common age-related clinical and neuropathologic traits. Genome Med. 17, 20. [PubMed: 40038788]

89. Bennett DA, Schneider JA, Arvanitakis Z, Kelly JF, Aggarwal NT, Shah RC, and Wilson RS (2006). Neuropathology of older persons without cognitive impairment from two community-based studies. Neurology 66, 1837–1844. [PubMed: 16801647]

90. Wilson RS, Arnold SE, Schneider JA, Tang Y, and Bennett DA (2007). The relationship between cerebral Alzheimer's disease pathology and odour identification in old age. J. Neurol. Neurosurg. Psychiatry 78, 30–35. [PubMed: 17012338]

91. Bennett DA, Wilson RS, Schneider JA, Evans DA, Aggarwal NT, Arnold SE, Cochran EJ, Berry-Kravis E, and Bienias JL (2003). Apolipoprotein E epsilon4 allele, AD pathology, and the clinical expression of Alzheimer's disease. Neurology 60, 246–252. [PubMed: 12552039]

92. Bennett DA, Schneider JA, Tang Y, Arnold SE, and Wilson RS (2006). The effect of social networks on the relation between Alzheimer's disease pathology and level of cognitive function in old people: a longitudinal cohort study. Lancet Neurol. 5, 406–412. [PubMed: 16632311]

93. Kapasi A, Poirier J, Hedayat A, Scherlek A, Mondal S, Wu T, Gibbons J, Barnes LL, Bennett DA, Leurgans SE, and Schneider JA (2023). High-throughput digital quantification of Alzheimer

disease pathology and associated infrastructure in large autopsy studies. J. Neuropathol. Exp. Neurol. 82, 976–986. [PubMed: 37944065]

94. Nag S, Yu L, Boyle PA, Leurgans SE, Bennett DA, and Schneider JA (2018). TDP-43 pathology in anterior temporal pole cortex in aging and Alzheimer's disease. Acta Neuropathol. Commun. 6, 33. [PubMed: 29716643]
95. Nag S, Yu L, Capuano AW, Wilson RS, Leurgans SE, Bennett DA, and Schneider JA (2015). Hippocampal sclerosis and TDP-43 pathology in aging and Alzheimer disease. Ann. Neurol. 77, 942–952. [PubMed: 25707479]
96. Schneider JA, Bienias JL, Wilson RS, Berry-Kravis E, Evans DA, and Bennett DA (2005). The apolipoprotein E epsilon4 allele increases the odds of chronic cerebral infarction [corrected] detected at autopsy in older persons. Stroke 36, 954–959. [PubMed: 15774818]
97. Arvanitakis Z, Leurgans SE, Barnes LL, Bennett DA, and Schneider JA (2011). Microinfarct pathology, dementia, and cognitive systems. Stroke 42, 722–727. [PubMed: 21212395]
98. Arvanitakis Z, Capuano AW, Leurgans SE, Buchman AS, Bennett DA, and Schneider JA (2017). The Relationship of Cerebral Vessel Pathology to Brain Microinfarcts. Brain Pathol. 27, 77–85. [PubMed: 26844934]
99. Buchman AS, Leurgans SE, Nag S, Bennett DA, and Schneider JA (2011). Cerebrovascular disease pathology and parkinsonian signs in old age. Stroke 42, 3183–3189. [PubMed: 21885844]
100. Love S, Chalmers K, Ince P, Esiri M, Attems J, Jellinger K, Yamada M, McCarron M, Minett T, Matthews F, et al. (2014). Development, appraisal, validation and implementation of a consensus protocol for the assessment of cerebral amyloid angiopathy in post-mortem brain tissue. Am. J. Neurodegener. Dis. 3, 19–32. [PubMed: 24754000]
101. Boyle PA, Yu L, Nag S, Leurgans S, Wilson RS, Bennett DA, and Schneider JA (2015). Cerebral amyloid angiopathy and cognitive outcomes in community-based older persons. Neurology 85, 1930–1936. [PubMed: 26537052]
102. Buchman AS, Leurgans SE, VanderHorst VGJM, Nag S, Schneider JA, and Bennett DA (2019). Spinal motor neurons and motor function in older adults. J. Neurol. 266, 174–182. [PubMed: 30446967]
103. Buchman AS, Wilson RS, Leurgans SE, Bennett DA, and Barnes LL (2015). Change in motor function and adverse health outcomes in older African-Americans. Exp. Gerontol. 70, 71–77. [PubMed: 26209439]
104. Buchman AS, Shulman JM, Nag S, Leurgans SE, Arnold SE, Morris MC, Schneider JA, and Bennett DA (2012). Nigral pathology and parkinsonian signs in elders without Parkinson disease. Ann. Neurol. 71, 258–266. [PubMed: 22367997]
105. Buchman AS, Yu L, Oveisgharan S, Farfel JM, Schneider JA, and Bennett DA (2021). Person-specific contributions of brain pathologies to progressive parkinsonism in older adults. J. Gerontol. A Biol. Sci. Med. Sci. 76, 615–621. [PubMed: 32720690]
106. Katz S, and Akpom CA (1976). A measure of primary sociobiological functions. Int. J. Health Serv. 6, 493–508. [PubMed: 133997]
107. Boyle PA, Buchman AS, Wilson RS, Bienias JL, and Bennett DA (2007). Physical activity is associated with incident disability in community-based older persons. J. Am. Geriatr. Soc. 55, 195–201. [PubMed: 17302655]
108. Buchman AS, Boyle PA, Leurgans SE, Evans DA, and Bennett DA (2009). Pulmonary function, muscle strength, and incident mobility disability in elders. Proc. Am. Thorac. Soc. 6, 581–587. [PubMed: 19934353]
109. Buchman AS, Boyle PA, Wilson RS, Tang Y, and Bennett DA (2007). Frailty is associated with incident Alzheimer's disease and cognitive decline in the elderly. Psychosom. Med. 69, 483–489. [PubMed: 17556640]
110. Buchman AS, Wilson RS, Bienias JL, and Bennett DA (2009). Change in frailty and risk of death in older persons. Exp. Aging Res. 35, 61–82. [PubMed: 19173102]
111. Kohout FJ, Berkman LF, Evans DA, and Cornoni-Huntley J (1993). Two shorter forms of the CES-D (Center for Epidemiological Studies Depression) depression symptoms index. J. Aging Health 5, 179–193. [PubMed: 10125443]

112. Wilson RS, Barnes LL, Mendes de Leon CF, Aggarwal NT, Schneider JS, Bach J, Pilat J, Beckett LA, Arnold SE, Evans DA, and Bennett DA (2002). Depressive symptoms, cognitive decline, and risk of AD in older persons. Neurology 59, 364–370. [PubMed: 12177369]
113. Wilson RS, Capuano AW, Boyle PA, Hoganson GM, Hizel LP, Shah RC, Nag S, Schneider JA, Arnold SE, and Bennett DA (2014). Clinical-pathologic study of depressive symptoms and cognitive decline in old age. Neurology 83, 702–709. [PubMed: 25080520]
114. Bennett DA, Wilson RS, Schneider JA, Bienias JL, and Arnold SE (2004). Cerebral infarctions and the relationship of depression symptoms to level of cognitive functioning in older persons. Am. J. Geriatr. Psychiatry 12, 211–219. [PubMed: 15010350]
115. Yu L, Lutz MW, Wilson RS, Burns DK, Roses AD, Saunders AM, Gaiteri C, De Jager PL, Barnes LL, and Bennett DA (2017). '523 variant and cognitive decline in older persons with ε3/3 genotype. Neurology 88, 661–668. [PubMed: 28108637]
116. Roses AD, Lutz MW, Amrine-Madsen H, Saunders AM, Crenshaw DG, Sundseth SS, Huentelman MJ, Welsh-Bohmer KA, and Reiman EM (2010). A TOMM40 variable-length polymorphism predicts the age of late-onset Alzheimer's disease. Pharmacogenomics J. 10, 375–384. [PubMed: 20029386]
117. Fujita M, Gao Z, Zeng L, McCabe C, White CC, Ng B, Green GS, Rozenblatt-Rosen O, Phillips D, Amir-Zilberstein L, et al. (2024). Cell subtype-specific effects of genetic variation in the Alzheimer's disease brain. Nat. Genet. 56, 605–614. [PubMed: 38514782]
118. Milacic M, Beavers D, Conley P, Gong C, Gillespie M, Griss J, Haw R, Jassal B, Matthews L, May B, et al. (2024). The Reactome Pathway Knowledgebase 2024. Nucleic Acids Res. 52, D672–D678. [PubMed: 37941124]
119. Gene Ontology Consortium; Aleksander SA, Balhoff J, Carbon S, Cherry JM, Drabkin HJ, Ebert D, Feuermann M, Gaudet P, Harris NL, et al. (2023). The Gene Ontology knowledgebase in 2023. Genetics 224, iyad031. 10.1093/genetics/iyad031. [PubMed: 36866529]
120. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. 25, 25–29. [PubMed: 10802651]
121. Bellenguez C, Küçükali F, Jansen IE, Kleineidam L, Moreno-Grau S, Amin N, Naj AC, Campos-Martin R, Grenier-Boley B, Andrade V, et al. (2022). New insights into the genetic etiology of Alzheimer's disease and related dementias. Nat. Genet. 54, 412–436. [PubMed: 35379992]
122. Chen W-T, Lu A, Craessaerts K, Pavie B, Sala Frigerio C, Corthout N, Qian X, Laláková J, Kühnemund M, Voytyuk I, et al. (2020). Spatial Transcriptomics and In Situ Sequencing to Study Alzheimer's Disease. Cell 182, 976–991.e19. [PubMed: 32702314]
123. Keren-Shaul H, Spinrad A, Weiner A, Matcovitch-Natan O, DvirSzternfeld R, Ulland TK, David E, Baruch K, Lara-Astaiso D, Toth B, et al. (2017). A Unique Microglia Type Associated with Restricting Development of Alzheimer's Disease. Cell 169, 1276–1290.e17. [PubMed: 28602351]
124. Srinivasan K, Friedman BA, Etxeberria A, Huntley MA, van der Brug MP, Foreman O, Paw JS, Modrusan Z, Beach TG, Serrano GE, and Hansen DV (2020). Alzheimer's Patient Microglia Exhibit Enhanced Aging and Unique Transcriptional Activation. Cell Rep. 31, 107843. [PubMed: 32610143]
125. Lancichinetti A, Fortunato S, and Radicchi F (2008). Benchmark graphs for testing community detection algorithms. Phys. Rev. E Stat. Nonlin. Soft Matter Phys. 78, 046110. [PubMed: 18999496]
126. Ng B, Tasaki S, Greathouse KM, Walker CK, Zhang A, Covitz S, Cieslak M, Weber AJ, Adamson AB, Andrade JP, et al. (2024). Integration across biophysical scales identifies molecular and cellular correlates of person-to-person variability in human brain connectivity. Nat. Neurosci. 27, 2240–2252. [PubMed: 39482360]

Highlights

- Integration of seven layers of brain omics data
- Multi-omics factors are robust and reveal consistent biological mechanisms
- Molecular features contributing to factors are potential therapeutic targets
- Clustering reveals 11 molecular subtypes with distinct characteristics

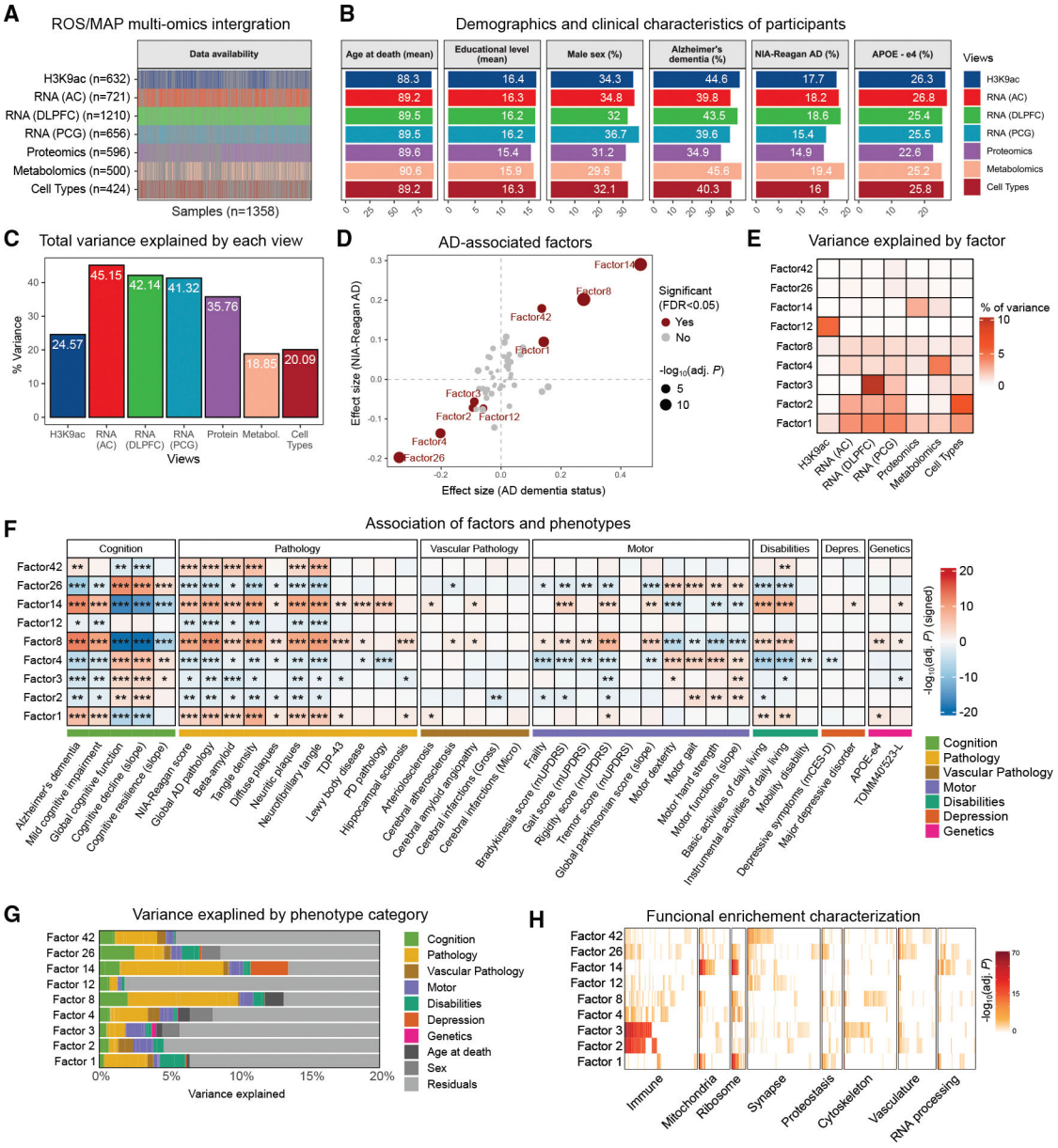


Figure 1. Functional and phenotypic characterization of AD-related factors

(A) Summary of the number of samples for each view. Numbers in parentheses represent the total number of samples, and gray lines represent the missing samples.

(B) Basic demographics and clinical and pathological characterization of individuals by view.

(C) The sum of variance explained from all factors for each view.

(D) Results of the associations of 50 factors with AD dementia (effect size, *x* axis) and NIA-Reagan AD diagnosis (effect size, *y* axis). The size of the dots represents the $-\log_{10}$ of nominal *p*. Red dots represent factors significantly associated after multiple-testing correction (false discovery rate [FDR] < 0.05).

(E) Variance explained by factors across views. The color scale shows the percentage of variance explained, ranging from 0% (white) to 10% (dark orange).

(F) Relationship of factors with covariates via regression analysis. Results were controlled for sex, age, and years of education. The color scale shows the signed log-transformed *p* value, corrected using the Benjamini-Hochberg (BH) FDR method to address multiple comparisons (38 phenotypes × 9 factors), with negative (blue) values indicating a negative association and positive (red) values indicating a positive association. *$-\log_{10}$ (adj. $p < 0.05$), **$-\log_{10}$ (adj. $p < 0.01$), and ***$-\log_{10}$ (adj. $p < 0.001$).

(G) Variance explained for each factor by phenotype category.

(H) Pathway enrichment analysis. Gene set enrichment analysis (GSEA) was performed for each factor, and the resulting terms were assigned to eight pathways. The selection of terms was made using pathway-related keywords. Approximately 31% of the total number of terms were allocated into one of the categories. The color scale shows the *p* values corrected by the BH method.

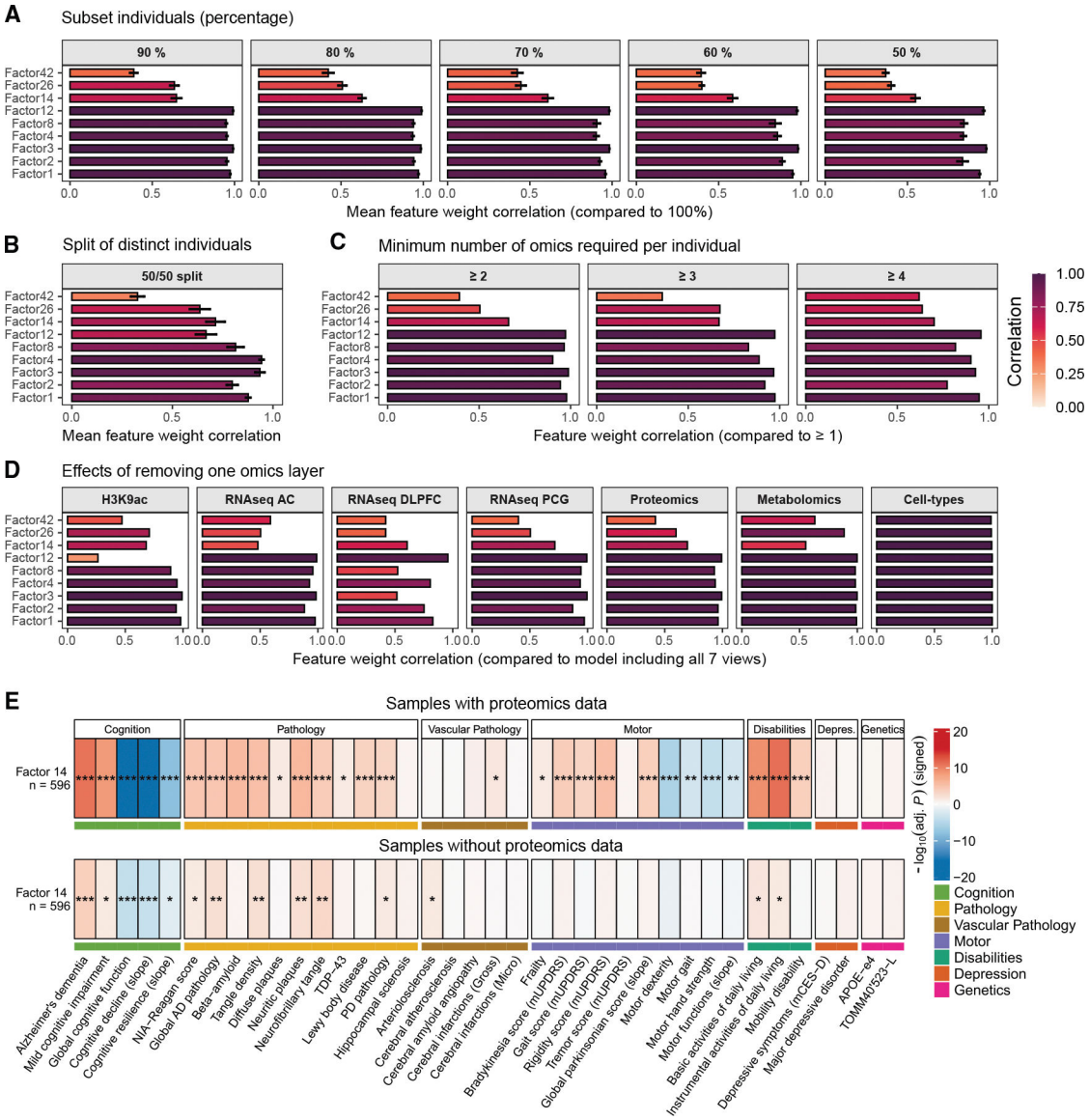


Figure 2. Evaluation of the robustness of AD-related factors under different scenarios

(A) Mean correlation of feature weights between the complete MOFA model (i.e., trained on all samples, as presented in the article) and alternative MOFA models trained on random subsets of decreasing sample sizes (i.e., 90%, 80%, 70%, 60%, and 50%).

(B) Mean correlation of 50%–50% splits comparing feature weights from models generated using independent samples.

(C) Feature weight correlation of MOFA models constructed, allowing different numbers of omics overlap (≥2, ≥3, and ≥4) compared with the model with ≥1.

(D) Feature weight correlation between MOFA models constructed with one missing omics layer (i.e., all samples, features converted to NA).

(E) Relationship of factor 14 with covariates in samples with (top heatmap) and without proteomics data (bottom heatmap). Results were controlled for sex, age, and years of education, and *p* values were corrected using the Benjamini-Hochberg (BH) false discovery

rate (FDR) method to address multiple comparisons. *$-\log_{10}$ (adj. $p < 0.05$), **$-\log_{10}$ (adj. $p < 0.01$), and ***$-\log_{10}$ (adj. $p < 0.001$). The color scale shows the association significance ($-\log_{10}$ adj. p) and the direction of effect, from negative (blue) to positive (red).
(A and B) Random subsampling was performed 15 times with different seeds for each percentage cut while keeping the proportion of omics preserved. In the plots, x axis bars (and color ranges) represent the mean absolute Pearson correlation of feature weights, and the error bar represents the standard deviation over the 15 random comparison sets.
(A–D) For each run, a similarity matrix was estimated by calculating all-vs.-all Pearson correlations between feature weights. Factors were matched by the correlation pair (one-to-many allowed).

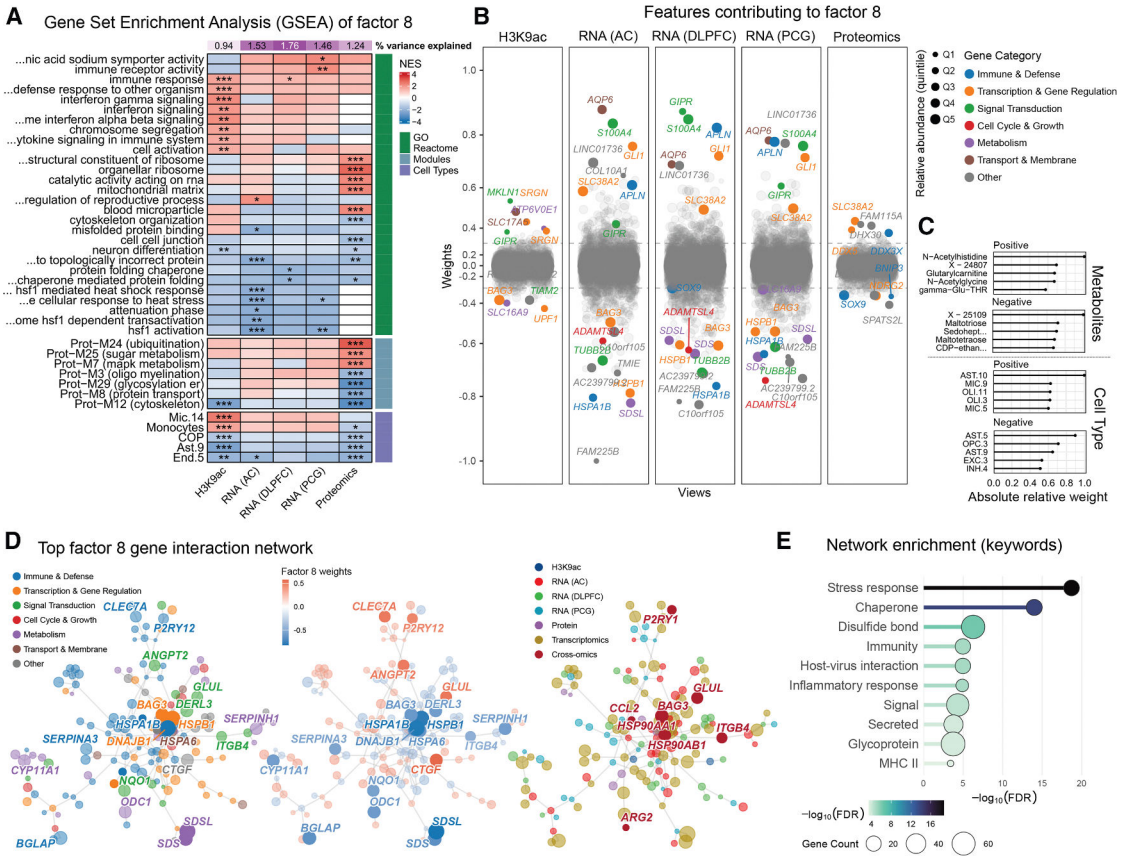


Figure 3. Characterization of factor 8 functional enrichment and feature contributions

(A) Gene set enrichment analysis (GSEA) of factor 8, showing the most significant terms for each view. The *p* values were corrected using the Benjamini-Hochberg (BH) method to address multiple comparisons. Terms were selected based on the lowest adjusted *p* values for each view. “GO” and “Reactome” denote terms from the Gene Ontology and Reactome databases, respectively. “Modules” correspond to co-expression protein modules described in Johnson et al.[24] “Cell Types” refers to subpopulations of cell lineages as determined by Green et al.[36] The color scale shows the normalized enrichment score (NES) direction and strength from negative (blue) to positive (red). *$-\log_{10}$ (adj. $p < 0.05$), **$-\log_{10}$ (adj. $p < 0.01$), and ***$-\log_{10}$ (adj. $p < 0.001$). Term names were truncated to 37 characters to maintain consistent figure size.

(B) Jitter plot showing genes with weights in each view, scaled across all views (cell types and metabolomics are not shown). Genes from selected pathways and with an absolute weight greater than 0.3 are highlighted (colored by category). Circle sizes represent the average expression of the feature, shown as quintiles.

(C) Top five positive and negative feature weights in cell types and metabolomics. Weights are scaled to each view. Metabolites shown as “X – …” have not been characterized.

(D) STRING protein-protein association network constructed from top factor 8 genes (absolute weights greater than 0.3). Only the largest connected component is shown. Nodes represent genes, and node sizes are proportional to the factor 8 weights (maximum weight per gene). Node colors indicate (left) functional categories, (center) feature weight, and

(right) omics source. Labels on the right denote genes detected in multiple omics layers (e.g., transcriptomics and proteomics).

(E) Network enrichment analysis using genes in the largest connected component as input. Color denotes $-\log_{10}$(FDR) from the Kolmogorov-Smirnov test, and dot size represents the number of genes per term. Results are restricted to UniProt keywords.

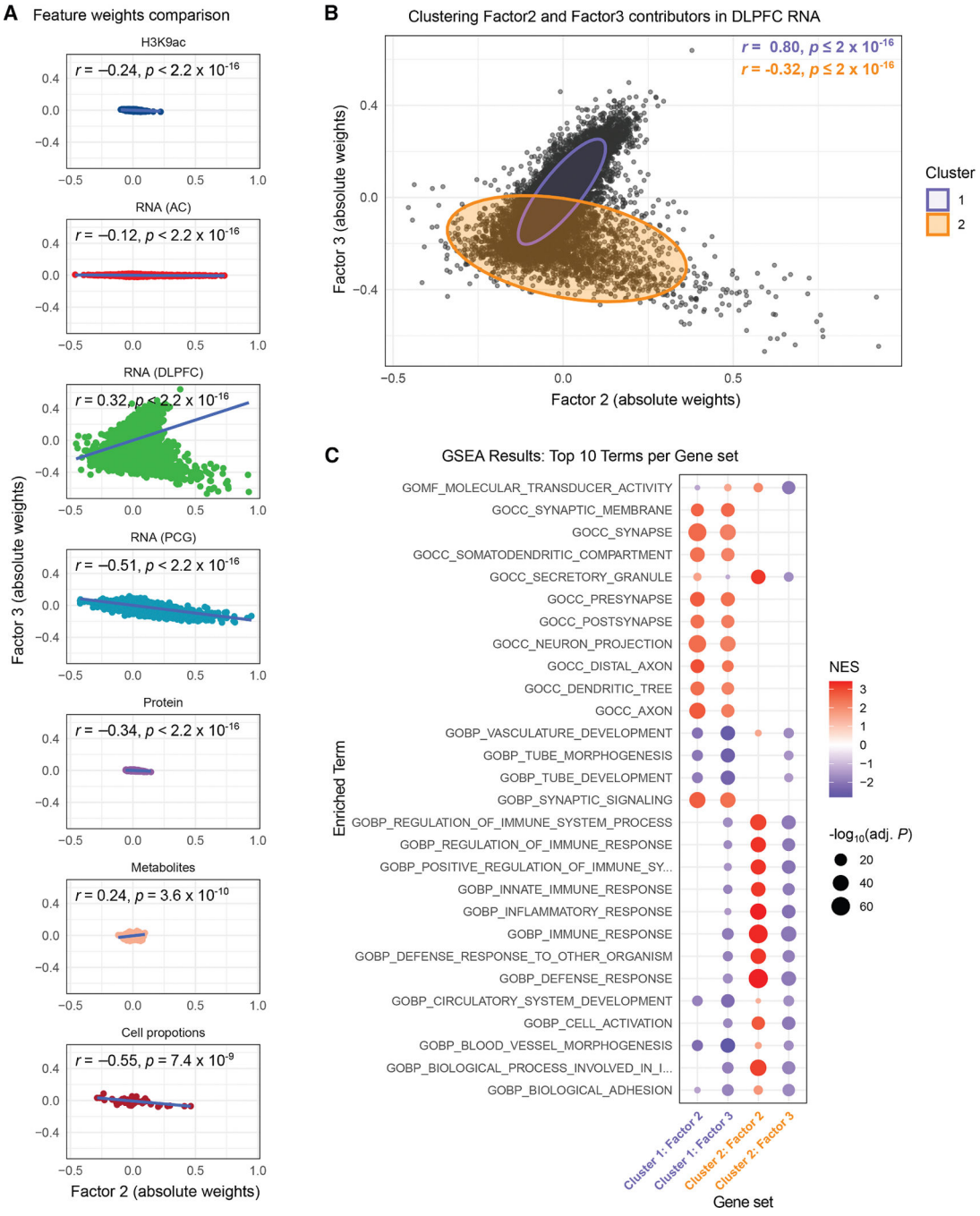


Figure 4. Comparison of features contributing to factors 2 and 3

(A) Feature weights for each view; the *x* axis represents factor 2, while the *y* axis refers to factor 3. The blue line represents the regression line. Pearson R and respective *p* values are shown in each view.

(B) The comparison of features from factors 2 and 3 for the RNA (DLPFC) view. A Gaussian mixture model (GMM) was applied to cluster the data into two clusters: cluster 1 in blue and cluster 2 in red.

(C) Gene Ontology enrichments using GSEA for features contributing to factors 2 and 3 in each GMM cluster. The *p* values were corrected using Benjamini-Hochberg (BH). The color scale shows the normalized enrichment score (NES) direction and strength from negative (blue) to positive (red). The size of the circles represents $-\log_{10}$ (adj. *p*).

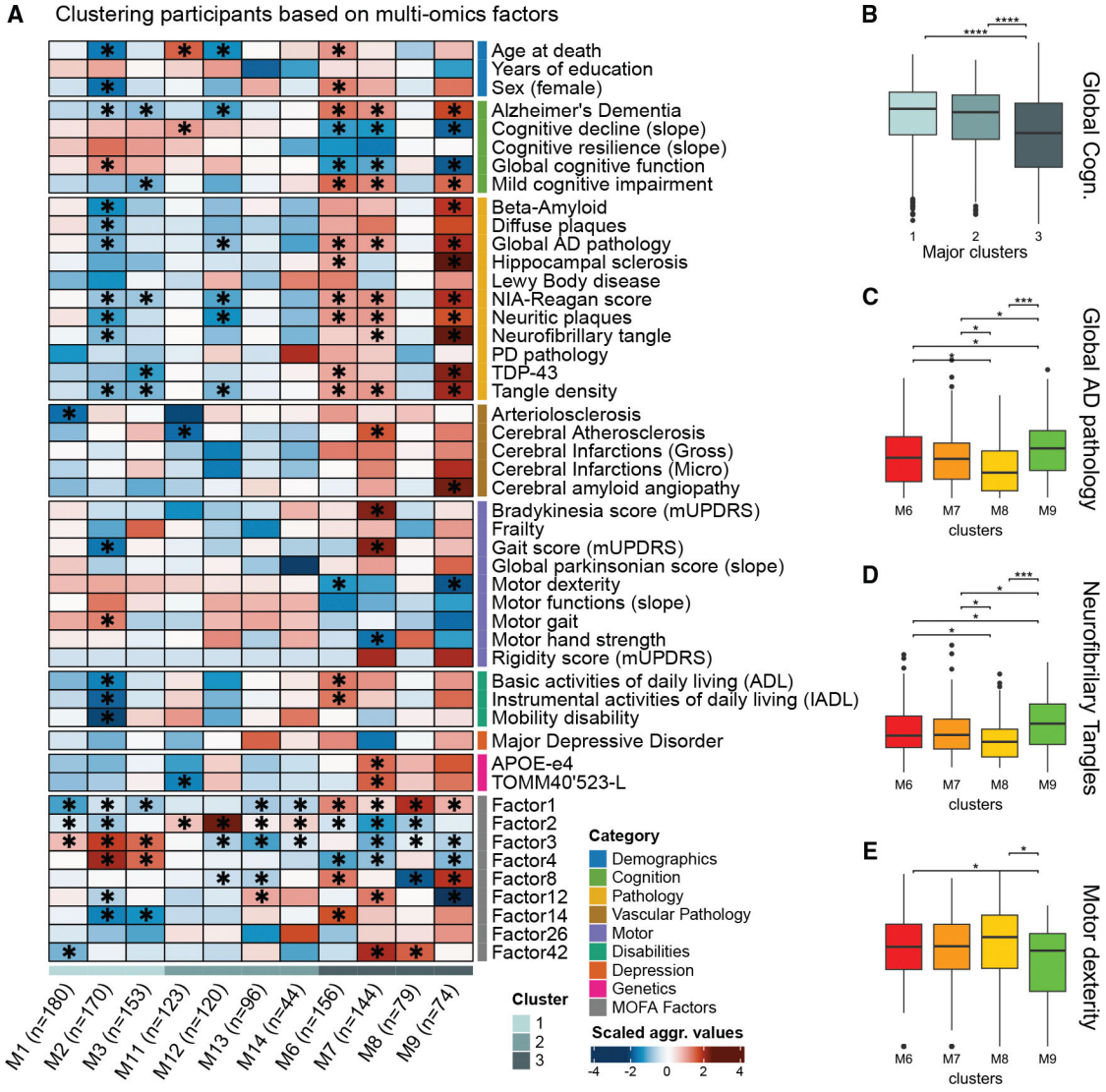


Figure 5. Molecular subtyping of ROS/MAP participants

(A) Association of clusters with Alzheimer's covariates and factors. Color scale shows the scaled aggregated values strength and direction from negative (blue) to positive (red). "Cluster" represents the first level of clustering, with three distinguished clusters. The *x* axis shows clusters from the second level of clustering, along with the number of samples for each cluster in parentheses. The *y* axis shows factors and covariates subdivided in nine categories. The traits with asterisks diverge significantly from the mean or median (for traits with Gaussian distribution) of all clusters after 1,000 permutations.

(B) *t* test comparing global cognitive function across the three groups at the major cluster level.

(C–E) *t* tests comparing clusters M6, M7, M8, and M9 with the following traits: global pathology AD burden (C), neurofibrillary tangle burden (D), and motor dexterity (E). The *p* values were corrected using the Benjamini-Hochberg (BH) false discovery rate (FDR) method. *adj. $p < 0.05$, **adj. $p < 0.01$, ***adj. $p < 0.001$, and ****adj. $p < 0.0001$.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Epigenetics ChIP-Seq (H3K9ac)	Klein et al.[71]	https://www.synapse.org/Synapse:syn4896408
RNA-Seq (DLPFC, AC, PCG)	Tasaki et al.[72]	https://www.synapse.org/Synapse:syn3388564
Proteomics TMT	Johnson et al.[73]	https://www.synapse.org/Synapse:syn17015098
Metabolomics	Batra et al.[39]	https://www.synapse.org/Synapse:syn26007830
snRNA-Seq	Green et al.[36]	https://www.synapse.org/Synapse:syn31512863
ROS/MAP phenotype data	RADC Resource Sharing Hub	www.radc.rush.edu
Software and algorithms		
MOFA2 R package v1.10.0	Argelaguet et al.[33]	https://github.com/bioFAM/MOFA2
R language v4.3.1 and v4.1.2	R Core Team[74]	https://www.r-project.org/
arm R package v1.14.4	Gelman et al.[75]	https://cran.r-project.org/package=arm
ordinal R package v2023.12.4	Christensen[76]	https://github.com/runehaubo/ordinal
fgsea R package v1.26.0	Korotkevich et al.[77]	https://bioconductor.org/packages/fgsea/
msigdbr R package v7.5.1	Dolgalev[78]	https://cran.r-project.org/package=msigdbr
rrvgo R package v1.12.2	Sayols[79]	https://www.bioconductor.org/packages/rrvgo/
Original code generated by this study	This paper	https://doi.org/10.5281/zenodo.18842084