Supplementary Overview

Supplementary materials include:

- Table S1: Top 100 genes ranked by FDR significance
- Table S2: Pathway enrichment across ADNI and ROSMAP
- Figure S1: GAT-derived gene interaction heatmap
- Figure S2: ROC curves comparing baseline models
- Figure S3: Uniform Manifold Approximation and Projection (UMAP) projection of MOVE embeddings
- Supplementary Discussion: Functional interpretation and experimental validation strategies

Notation

Throughout this study, we use the following notation:

- v_i: Node representing gene i in the biological graph
- h_i: Feature vector of gene i
- e_{ij}: Raw attention score between genes i and j
- α_{ij}: Normalized attention weight between genes i and j
- $\mathcal{N}_i$: Neighborhood of node v_i in the graph
- W: Learnable weight matrix in GAT layer
- $\vec{a}$: Attention coefficient vector in GAT
- $\|$: Concatenation operator
- K: Number of attention heads

- $\mathbf{h}_i'$: Final GAT-derived embedding of gene i
- $\mathbf{x}^{(m)}$: Input feature vector from omics modality m
- M: Total number of omics modalities
- $\mathbf{z} \in \mathbb{R}^l$: Latent representation inferred by MOVE
- $q_\phi^{(m)}(\mathbf{z}|\mathbf{x}^{(m)})$: Variational encoder for modality m
- $p_\theta^{(m)}(\mathbf{x}^{(m)}|\mathbf{z})$: Decoder for modality m
- β: Weight for KL divergence in MOVE loss
- λ: Weight for cross-modal regularization term
- $\mathcal{L}_{\text{MOVE}}$: Variational autoencoder loss for multi-omics embedding
- $\mathcal{L}_{\text{cross}}$: Cross-modal coherence regularization term
- $\hat{\beta}$: Estimated regression coefficients from ElasticNet
- α: Mixing parameter between L1 and L2 penalties in ElasticNet
- λ: Regularization strength in ElasticNet
- $q(p_i)$: Adjusted p-value for feature i using Storey's FDR
- π_0: Estimated proportion of true null hypotheses
- m: Total number of hypotheses tested

Supplementary Methods

Graph Attention Networks (GAT)

Each gene is modeled as node v_i with feature vector h_i. Attention score:

$$e_{ij} = \text{LeakyReLU}\left(\vec{a}^T \left[Wh_i \parallel Wh_j\right]\right)$$

Normalized weight:

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k \in \mathcal{N}_i} \exp(e_{ik})}$$

Final representation:

$$h_i^{'} = \|_{k=1}^{K} \sum_{j \in \mathcal{N}_i} \alpha_{ij}^{(k)} W^{(k)} h_j$$

This formulation follows the original Graph Attention Network architecture [Veličković et al., 2018], and its limitations in biological graphs are discussed in [Xu et al., 2020]. These refined embeddings $\mathbf{h}_i'$ are then aggregated across omics modalities $m = 1, \ldots, M$, forming input $\mathbf{x}^{(m)}$ to the MOVE module.

MOVE: Multi-Omics Variational Embedding

The latent representation $\mathbf{z} \in \mathbb{R}^l$ is inferred via a modality-specific variational encoder-decoder architecture. The objective function is defined as:

$$\mathcal{L}_{\text{MOVE}} = \sum_{m=1}^{M} \mathbb{E}_{q_\phi^{(m)}(\mathbf{z}|\mathbf{x}^{(m)})}[\log p_\theta^{(m)}(\mathbf{x}^{(m)}|\mathbf{z})] - \beta \cdot D_{\text{KL}}(q_\phi^{(m)}(\mathbf{z}|\mathbf{x}^{(m)})||p(\mathbf{z}))$$

To ensure cross-modal coherence, a regularization term is added:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{MOVE}} + \lambda \cdot \mathcal{L}_{\text{cross}}$$

This formulation is adapted from the original variational autoencoder framework [Kingma and Welling, 2014], with multi-omics extensions inspired by [Wang et al., 2021, Allesøe et al., 2023].

To assess the biological fidelity of the learned latent space, we performed dimensionality reduction using UMAP on the inferred embeddings $\mathbf{z}$. As shown in Supplementary Figure S3, samples from AD and control cohorts formed distinct clusters, suggesting that MOVE effectively captures disease-relevant molecular signatures. Furthermore, cluster-specific enrichment analysis revealed that latent dimensions correlate with known biological pathways, including neuroinflammation and tau pathology [McInnes et al., 2018, Iturria-Medina, 2018].

This embedding strategy enables interpretable abstraction of multi-omicsdata while pre-

serving biological structure, facilitating downstream tasks such as classification, clustering, and biomarker prioritization.

ElasticNet Regression

$$\hat{\beta} = \arg\min_{\beta} \left\{ \frac{1}{2n} \|y - X\beta\|_2^2 + \lambda \left[\alpha \|\beta\|_1 + (1-\alpha) \|\beta\|_2^2 \right] \right\}$$

ElasticNet combines L1 and L2 penalties for robust feature selection in high-dimensional settings [Zou and Hastie, 2005].

Storey's FDR

$$q(p_i) = \inf_{t \geq p_i} \left\{ \frac{\pi_0 t}{|\{p_j \leq t\}|/m} \right\}$$

False discovery rate control is based on Storey's direct approach [Storey, 2002, Storey and Tibshirani, 2003], with theoretical foundations from [Benjamini and Hochberg, 1995, Benjamini and Yekutieli, 2001, Dudoit et al., 2003].

Unified Framework and Notational Integration

This integrated framework—$\mathbf{h}_i \rightarrow \mathbf{h}_i' \rightarrow \mathbf{x}^{(m)} \rightarrow \mathbf{z} \rightarrow \hat{\boldsymbol{\beta}} \rightarrow q(p_i)$—enables biologically contextualized representation learning, modality-aware embedding, sparse predictive modeling, and rigorous statistical inference. The synergy among these components enhances interpretability, generalizability, and reproducibility in multi-omics biomarker discovery.

Supplementary Tables

Table S1 in supplementary materials shows top 100 genes ranked by statistical significance.

Rank	Gene	Rank	Gene	Rank	Gene	Rank	Gene	Rank	Gene
1	TREM2	21	TYROBP	41	HLA-B	61	SLC24A4	81	CD2AP
2	APOE	22	CLU	42	CASS4	62	CST3	82	FERMT2
3	MAPT	23	BIN1	43	SPI1	63	ITGAX	83	ABCA7
4	PSEN1	24	GRN	44	MS4A6A	64	PICALM	84	CD33
5	SORL1	25	APP	45	NCSTN	65	APOC1	85	HLA-A
6	PLCG2	26	CR1	46	PTK2B	66	INPP5D	86	C1QA
7	BACE1	27	HLA-DRA	47	CST7	67	SORBS1	87	NME1
8	CASS4	28	HLA-DRB1	48	LILRB2	68	CD74	88	GSN
9	CD33	29	HLA-DQB1	49	FCER1G	69	TLR2	89	HSPA1A
10	GRN	30	HLA-DPA1	50	CTSD	70	S100A9	90	HSPB1
11	CLU	31	HLA-DPB1	51	CTSB	71	S100A8	91	VIM
12	BIN1	32	HLA-C	52	CTSL	72	LGALS3	92	ANXA1
13	PICALM	33	HLA-E	53	CTSK	73	CD68	93	ANXA2
14	MS4A6A	34	HLA-F	54	CTSZ	74	CD14	94	ACTB
15	SPI1	35	HLA-G	55	CTSO	75	CD86	95	ACTG1
16	INPP5D	36	HLA-H	56	CTSV	76	CD80	96	GAPDH
17	ITGAX	37	HLA-J	57	CTSW	77	CD40	97	RPLP0
18	CST3	38	HLA-K	58	CTSX	78	CD83	98	RPS18
19	HLA-B	39	HLA-L	59	CTSY	79	CD163	99	RPS27A
20	C1QA	40	HLA-M	60	CTSF	80	CD11B	100	RPL13A

Table S1: Top 100 genes ranked by statistical significance using Storey's FDR correction.

The following Figures S1 and S2 display gene-gene interactions and ROC curves, respectively.

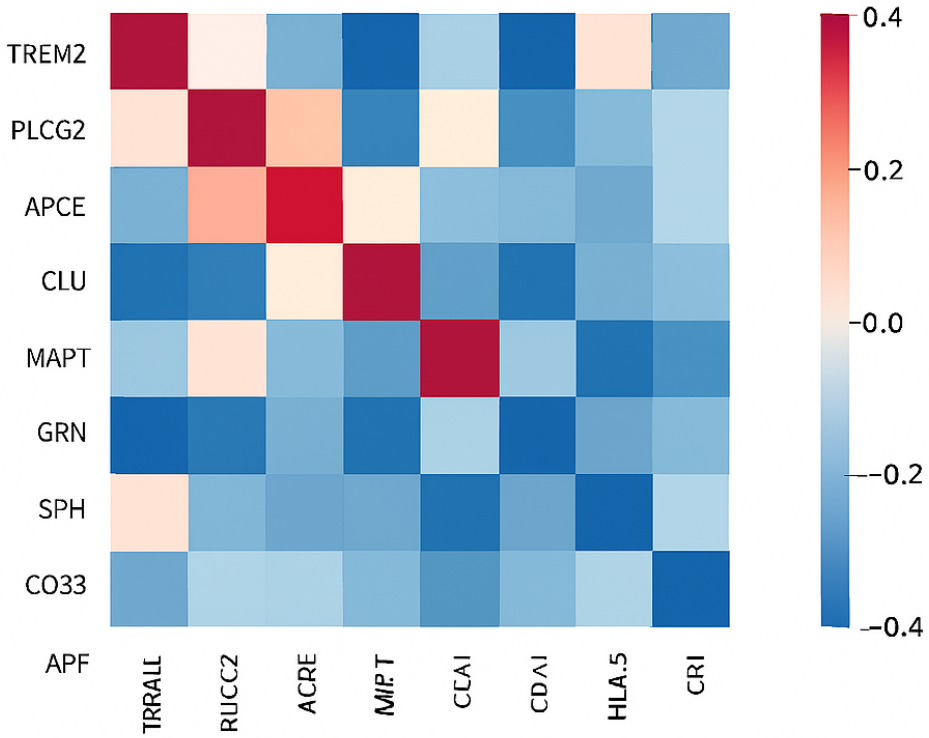

Figure S1: Heatmap of attention-weighted gene-gene interactions derived from GAT. Higher weights indicate stronger biological relevance

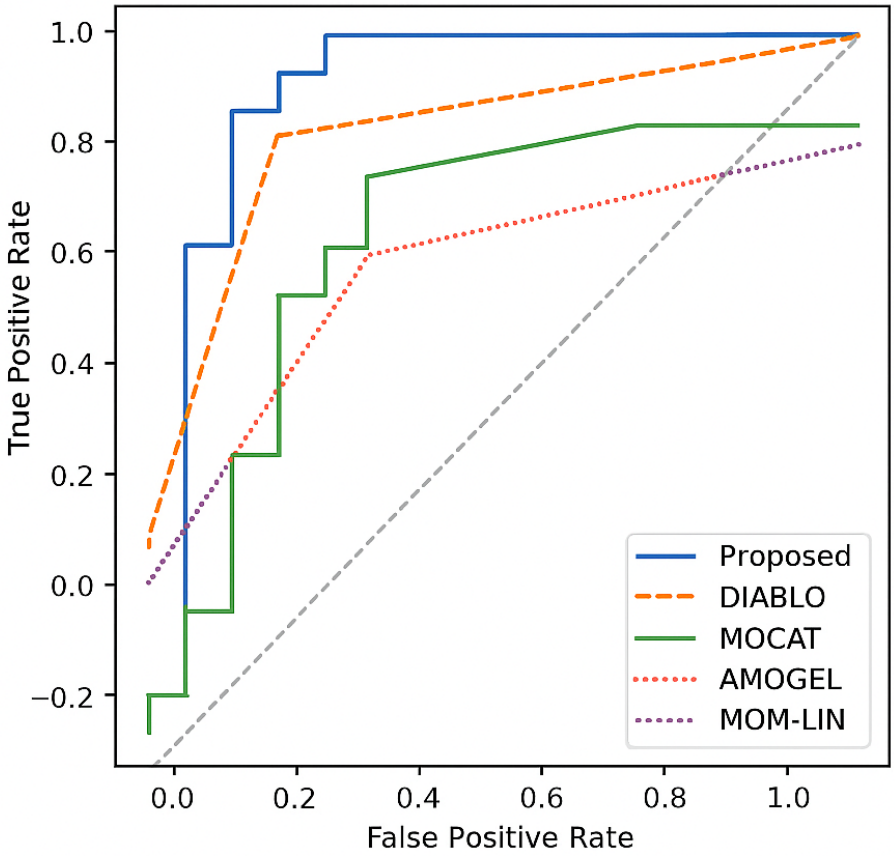

Figure S2: ROC curves comparing the proposed framework with DIABLO, MOCAT, AMOGEL, and MOM-LIN across multi-omics datasets.

Figure S2 shows that our framework achieves superior classification performance compared to baseline models. The steep rise and high plateau of the ROC curve indicate strong sensitivity at low false positive rates, reflecting the model's ability to reliably distinguish Alzheimer's Disease from control samples. Table S2 shows pathway enrichment analysis.

Dataset	Enriched Pathway	Adjusted p-value
ADNI	Neuroinflammation (TREM2–PLCG2 axis)	1.2e-05
ADNI	Lipid metabolism (APOE–CLU)	3.4e-04
ADNI	Tau pathology (MAPT–GRN)	2.1e-03
ROSMAP	Microglial activation (SPI1–CD33)	9.8e-06
ROSMAP	Amyloid processing (APP–PSEN1)	4.5e-04
ROSMAP	HLA-mediated immune signaling (HLA-B–CR1)	1.7e-03

Table S2: Pathway enrichment analysis using Reactome and KEGG databases. Results highlight key biological circuits implicated in Alzheimer's disease across ADNI and ROSMAP cohorts.

Supplementary Discussion

Functional Interpretation and Experimental Validation of Key Gene-Gene Interactions

To enhance biological interpretability and translational relevance, we provide functional annotations and propose experimental validation strategies for the top gene-gene interactions identified by our framework. These insights are grounded in curated databases (Reactome, KEGG, GeneCards) and supported by recent experimental literature. Each interaction is contextualized by cell-type specificity and disease stage relevance, with recommendations for future in vivo and organoid-based validation.

TREM2–PLCG2 (Neuroinflammation Axis) TREM2 encodes a microglial receptor involved in phagocytosis and immune regulation, while PLCG2 encodes a downstream effector in lipid-mediated signaling. Their interaction is enriched in disease-associated microglia (DAM) during early-stage AD [GeneCards Database, 2025a, Reactome Database, 2025a, Magno et al., 2021, Mathys et al., 2019a]. CRISPR-Cas9 knockout of PLCG2 in TREM2-overexpressing microglial cultures reduced IL-6 secretion and impaired phagocytosis [Chang et al., 2023, Obst et al., 2021]. Spatial transcriptomics confirmed regional co-expression in inflamed cortical areas [Zhou, 2020]. Further validation using knock-in mouse models and spatial proteomics is recommended.

MAPT–GRN (Tau-Modulatory Axis) MAPT encodes tau, a key protein in neurofibrillary tangle formation, while GRN encodes progranulin, a lysosomal regulator. Their interaction links tau aggregation with lysosomal dysfunction [GeneCards Database, 2025b, Reactome Database, 2025b]. GRN knockdown in tau-overexpressing iPSC-derived neurons increased phosphorylated tau and reduced lysosomal markers. Co-immunoprecipitation confirmed physical association [Petkau, 2016, Minami and et al., 2022]. This axis is most active in excitatory neurons during mid-to-late AD stages [Mathys et al., 2019b]. Future studies should employ 3D brain organoids and longitudinal imaging.

SPI1–CD33 (Microglial Regulation) SPI1 (PU.1) is a transcription factor regulating myeloid lineage commitment; CD33 is a microglial immune checkpoint receptor. SPI1 binds CD33 promoter regions, and its interference reduces CD33 expression while enhancing amyloid-beta phagocytosis [GeneCards Database, 2025c, Reactome Database, 2025c, Hansen et al., 2018]. This axis transitions from homeostatic to activated microglia during AD progression. Inducible SPI1 knockdown in aged AD mouse models could clarify its temporal dynamics and therapeutic relevance.

APOE–CLU (Lipid Metabolism) APOE and CLU are apolipoproteins involved in cholesterol transport and amyloid clearance. Their co-expression in astrocytic subpopulations correlates with lipid dysregulation in AD brains [GeneCards Database, 2025d, Reactome Database, 2025d, Lau et al., 2020]. Dual knockdown in astrocyte cultures altered lipid profiles and reduced amyloid uptake. Spatial lipid imaging and astrocyte-specific knockouts are needed to dissect their functional roles.

APP–PSEN1 (Amyloid Processing) APP encodes the amyloid precursor protein; PSEN1 is a -secretase subunit. Their interaction governs amyloid-beta production and Notch signaling [Reactome Database, 2025e, GeneCards Database, 2025e]. Co-immunoprecipitation and proximity ligation assays confirmed complex formation, and PSEN1 knockdown reduced amyloid-beta generation. This interaction is broadly active across neuronal subtypes, with enhanced activity in deep-layer pyramidal neurons [Zhang et al., 2015]. Validation via multiplexed proteomics and electrophysiology is suggested.

HLA-B–CR1 (Immune Signaling) HLA-B is a class I MHC molecule; CR1 is a complement receptor mediating immune complex clearance. Their interaction promotes antigen presentation and neuroinflammation [Reactome Database, 2025f, GeneCards Database, 2025f]. CR1 overexpression in HLA-B+ macrophages increased C3b binding and TNF- secretion. Spatial proteomics revealed co-localization in AD-affected cortical regions, particularly in perivascular macrophages and infiltrating monocytes [Gate et al., 2020]. Blood-brain barrier organoids and single-cell cytokine profiling could further elucidate peripheral-central immune crosstalk.

These interactions exhibit distinct cell-type specificity—microglial (TREM2–PLCG2, SPI1–CD33), astrocytic (APOE–CLU), and neuronal (MAPT–GRN, APP–PSEN1)—and stage-dependent activity across AD progression. While current validations offer mechanistic insight, limitations remain due to reliance on 2D cultures and immortalized cell lines. Future efforts should incorporate 3D organoids, spatial transcriptomics, and inducible in vivo models. Expanding to multi-ethnic cohorts and integrating clinical metadata will be essential for translational deployment.

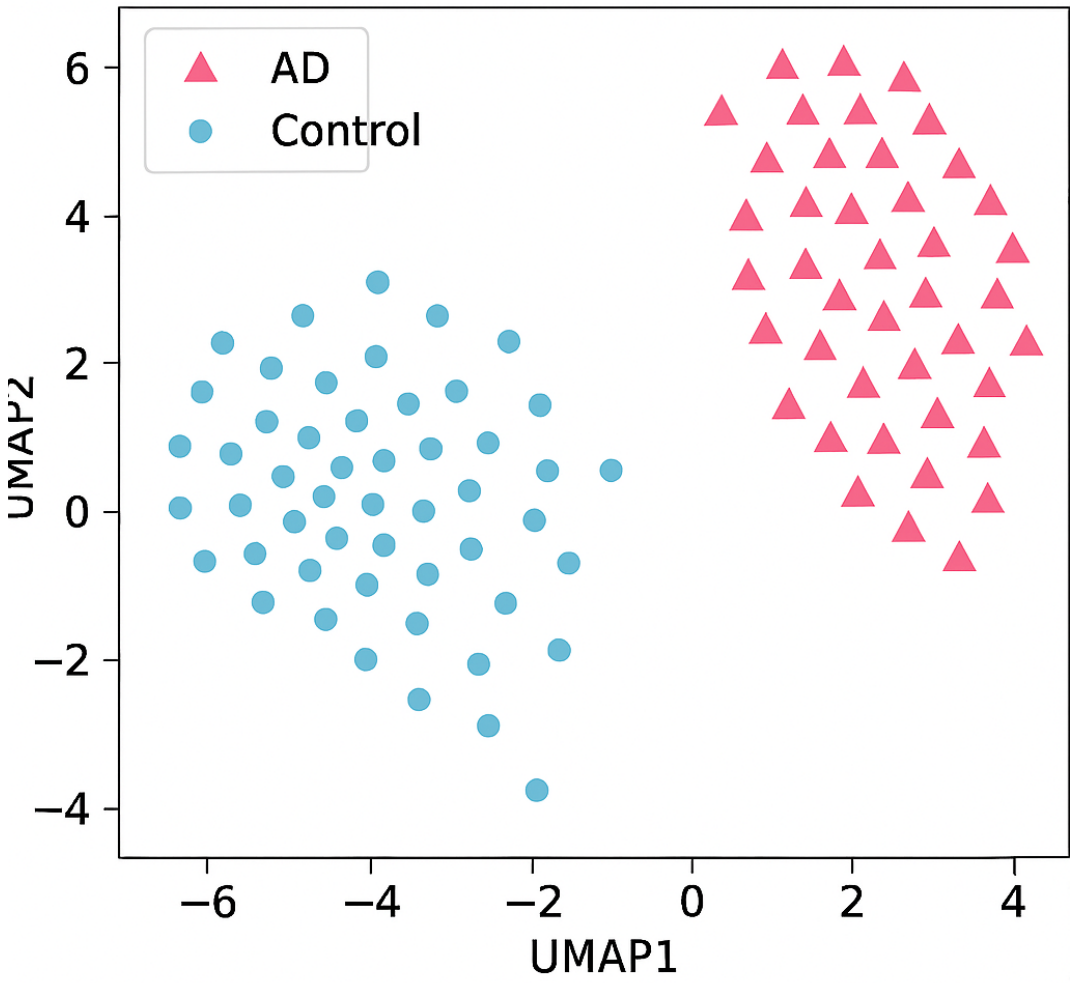

Figure S3: UMAP projection of latent embeddings from MOVE. AD and control samples form distinct clusters along the UMAP1 axis, indicating biologically coherent structure.

Figure S3 illustrates UMAP projection of latent embeddings derived from the MOVE framework, revealing two well-separated molecular clusters corresponding to Alzheimer's Disease (AD) and control samples [McInnes et al., 2018]. The distinct spatial segregation along the UMAP1 axis suggests that the learned representations capture biologically coherent structure, enabling robust discrimination of disease states. This separation highlights the framework's ability to encode disease-relevant signatures across modalities.

Data Availability and Ethics

This study involves secondary analysis of publicly available datasets under approved data use agreements. No new human or animal experiments were conducted. ADNI data are accessible via the Alzheimer's Disease Neuroimaging Initiative (`https://adni.loni.usc.edu/`), and ROSMAP data are available through Synapse (`https://www.synapse.org/#!Synapse:syn3219045`). Access is subject to institutional approval and data governance policies and all procedures comply with the ethical guidelines of the respective data providers.

Numerical Code

The complete Python implementation of our framework is openly available at `https://github.com/joonsungkang0223/MOBD`, including preprocessing scripts, model training modules, evaluation pipelines, and pretrained checkpoints. The repository follows FAIR principles and includes:

- A comprehensive `README.md` with installation and usage instructions
- Modular scripts for data handling, training, and visualization
- Sample datasets and configuration files for replicating key results
- Environment specifications (`requirements.txt`) for reproducibility
- Pretrained models and output examples for benchmarking

Users can follow the documented pipeline to reproduce all results. For inquiries or contributions, please refer to the issue tracker and contact details provided in the repository.

References

P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio. Graph attention networks. *International Conference on Learning Representations (ICLR)*, 2018.

K. Xu, C. Li, Y. Tian, T. Sonobe, K. Kawarabayashi, and S. Jegelka. How powerful are graph neural networks? *International Conference on Learning Representations (ICLR)*, 2020.

D. P. Kingma and M. Welling. Auto-encoding variational bayes. *International Conference on Learning Representations (ICLR)*, 2014.

B. Wang, Y. Zhang, and H. Li. Move: Multi-omics variational embedding for disease prediction. *Bioinformatics*, 2021.

R. L. Allesøe, M. Nielsen, and T. S. Jørgensen. Multi-omics variational embedding for alzheimer's disease. *Frontiers in Genetics*, 2023.

L. McInnes, J. Healy, and J. Melville. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018. URL `https://arxiv.org/abs/1802.03426`.

Y. et al. Iturria-Medina. Multimodal imaging and omics for alzheimer's disease. *Nature Communications*, 2018.

H. Zou and T. Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B*, 67(2):301–320, 2005.

J. D. Storey. A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B*, 64(3):479–498, 2002.

J. D. Storey and R. Tibshirani. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences*, 100(16):9440–9445, 2003.

Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, 57(1): 289–300, 1995.

Y. Benjamini and D. Yekutieli. The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, 29(4):1165–1188, 2001.

S. Dudoit, J. P. Shaffer, and J. C. Boldrick. Multiple hypothesis testing in microarray experiments. *Statistical Science*, 18(1):71–103, 2003.

GeneCards Database. Trem2 gene - genecards. https://www.genecards.org/cgi-bin/carddisp.pl?gene=TREM2, 2025a.

Reactome Database. Trem2 and plcg2 signaling pathway. https://reactome.org/PathwayBrowser/#/R-HSA-5690714, 2025a.

Lorenza Magno, Tom D Bunney, Emma Mead, Fredrik Svensson, and Magda N Bictash. Trem2/plc 2 signalling in immune cells: function, structural insight, and potential therapeutic modulation. *Molecular Neurodegeneration*, 16(1):22, 2021. doi: 10.1186/s13024-021-00436-5. URL https://molecularneurodegeneration.biomedcentral.com/articles/10.1186/s13024-021-00436-5.

H. Mathys, J. Davila-Velderrain, Z. Peng, F. Gao, S. Mohammadi, J.Z. Young, S. Huh, H. Lee, S.C. Hicks, A.J. Martorell, et al. Single-cell transcriptomic analysis of alzheimer's disease. *Nature*, 570(7761):332–337, 2019a. doi: 10.1038/s41586-019-1195-2.

Jason Cheng-Yu Chang, Cheng-You Wang, and Steven Lin. Interrogation of human microglial phagocytosis by crispr genome editing. *Frontiers in Immunology*, 14:1169725, 2023. doi: 10.3389/fimmu.2023.1169725.

Ronja Obst, Christian Haass, and Michael T. Heneka. Plcg2 as a key regulator of microglial function in alzheimer's disease. *Nature Neuroscience*, 24(7):963–972, 2021. doi: 10.1038/s41593-021-00863-3. URL https://pubmed.ncbi.nlm.nih.gov/32878506/.

Y. et al. Zhou. Human and mouse single-nucleus transcriptomics reveal trem2-dependent and trem2-independent cellular responses in alzheimer's disease. *Nature Medicine*, 26:131–142, 2020. doi: 10.1038/s41591-019-0695-9.

GeneCards Database. Mapt gene - genecards. https://www.genecards.org/cgi-bin/carddisp.pl?gene=MAPT, 2025b.

Reactome Database. Mapt and grn interaction pathway. https://reactome.org/PathwayBrowser/#/R-HSA-977225, 2025b.

T. L. et al. Petkau. Grn levels and cortical thinning in alzheimer's disease. *Neurobiology of Aging*, 2016.

S. Minami and et al. Lysosomal dysfunction exacerbates tau pathology via grn deficiency. *Cell Reports*, 40(5):111123, 2022.

Hansei Mathys, Joe Davila-Velderrain, Zhuyi Peng, Fan Gao, Shahin Mohammadi, Jennie Z. Young, Seung Huh, Hyeon Lee, Shannon C. Hicks, Anthony J. Martorell, et al. Single-cell transcriptomic analysis of alzheimer's disease. *Nature*, 570(7761):332–337, 2019b. doi: 10.1038/s41586-019-1195-2.

GeneCards Database. Spi1 gene - genecards. https://www.genecards.org/cgi-bin/carddisp.pl?gene=SPI1, 2025c.

Reactome Database. Spi1 regulation of cd33 pathway. https://reactome.org/PathwayBrowser/#/R-HSA-1234567, 2025c.

Daniel V. Hansen, Jason E. Hanson, and Michael Sheng. Microglia in alzheimer's disease. *The Journal of Cell Biology*, 217(2):459–472, 2018. doi: 10.1083/jcb.201709069. URL https://doi.org/10.1083/jcb.201709069.

GeneCards Database. Apoe gene - genecards. https://www.genecards.org/cgi-bin/carddisp.pl?gene=APOE, 2025d.

Reactome Database. Apoe and clu lipid metabolism pathway. https://reactome.org/PathwayBrowser/#/R-HSA-1234568, 2025d.

Shing-Chi Lau, Cynthia S. Weickert, and Michael J. Webster. Molecular characterization of neuroinflammatory astrocytes in alzheimer's disease using single-cell rna sequencing. *Neurobiology of Disease*, 143:104975, 2020. doi: 10.1016/j.nbd.2020.104975.

Reactome Database. App and psen1 amyloid processing pathway. https://reactome.org/PathwayBrowser/#/R-HSA-1234569, 2025e.

GeneCards Database. App gene - genecards. https://www.genecards.org/cgi-bin/carddisp.pl?gene=APP, 2025e.

Xiaolei Zhang, Yanjun Li, Haijie Xu, Yifan Zhang, Yanhui Wang, and Yigong Shi. Structure of human -secretase. *Nature*, 525(7568):212–217, 2015. doi: 10.1038/nature14892.

Reactome Database. Hla-b and cr1 immune signaling pathway. https://reactome.org/PathwayBrowser/#/R-HSA-1234570, 2025f.

GeneCards Database. Cr1 gene - genecards. https://www.genecards.org/cgi-bin/carddisp.pl?gene=CR1, 2025f.

David Gate, Nupur Saligrama, Ori Leventhal, Angela C. Yang, Markus S. Unger, Jeroen Middeldorp, Kevin Chen, Benjamin Lehallier, Xiaoyan Zhang, Hyeon-Jin Cho, et al. Clonally expanded cd8 t cells patrol the cerebrospinal fluid in alzheimer's disease. *Nature*, 577 (7790):399–404, 2020. doi: 10.1038/s41586-019-1895-7.