Graph-Based Multi-Omics Integration Reveals Neuroimmune and Tau-Modulatory Biomarkers in Alzheimer's Disease

Byeonghee Lee[1], Juhyun Park[2], and Joonsung Kang[3]

[1]Department of Mathematics and Physics, Gangneung-Wonju National University, Republic of Korea

[2]Department of Statistics, Dongguk University, Republic of Korea

[3]Department of Data Science, Gangneung-Wonju National University, Republic of Korea

Abstract

We present a unified and interpretable framework for multi-omics biomarker discovery in Alzheimer's disease, integrating graph attention networks, variational manifold embedding, sparse regression, and false discovery rate control. Tailored for high-dimensional, low-sample size data, the model achieves state-of-the-art performance across synthetic, ADNI, and ROSMAP cohorts. Beyond reaffirming canonical biomarkers such as *TREM2*, *APOE*, and *MAPT*, our approach uncovers novel gene-gene interactions—*MAPT–GRN* and *TREM2–PLCG2*—that illuminate neuroimmune and tau-modulatory pathways. The inferred latent representations exhibit biologically coherent structure, as shown by UMAP-based clustering of Alzheimer's disease and control samples. These embeddings enable downstream tasks including subtype classification and pathway enrichment, enhancing mechanistic insight. The modular ar-

chitecture supports extension to other disease contexts, positioning the framework as a scalable platform for precision diagnostics and translational medicine.

Keywords: Alzheimer's disease, Multi-omics integration, Graph attention networks, Variational embedding, Sparse regression, False discovery rate, Biomarker discovery, Neuroimmune pathways

1 Introduction

Alzheimer's disease presents substantial molecular heterogeneity, complicating early diagnosis and targeted intervention. Multi-omics integration—spanning genomic, transcriptomic, proteomic, and metabolomic layers—offers a powerful strategy to uncover subtype-specific biomarkers and mechanistic insights [Iturria-Medina, 2018, Hampel and et al., 2021, Wang and Zhang, 2021, Hasin et al., 2017, Karczewski and Snyder, 2018]. However, the high-dimensional, low-sample size nature of such data challenges conventional models, which often lack interpretability and statistical robustness [Fan and Lv, 2008, LeCun et al., 2015, Zhang, 2025, Tu, 2024].

To address these limitations, we introduce an ensemble framework integrating Graph Attention Networks (GAT), Variational Manifold Encoding—specifically, Multi-Omics Variational Autoencoder (MOVE)—ElasticNet regression, and Storey's False Discovery Rate (FDR). GAT captures latent gene-gene interactions by modeling topological dependencies across omics layers. These graph-informed features are encoded and decoded via MOVE into a unified latent manifold, enabling sample-wise abstraction and reconstruction. ElasticNet is then applied to identify statistically relevant genes and interactions under sparsity constraints. Subsequently, Storey's FDR ranks the selected features by controlling for multiple hypothesis testing, ensuring interpretability and statistical rigor. By integrating topological attention with variational encoding and sparse regression, our framework not only improves predictive performance but also yields interpretable biomarkers and mechanistic insights into AD pathogenesis.

2 Methods

- **Data preprocessing:** - Transcriptomics: quantile normalization - Proteomics: z-score scaling - Metabolomics: log-transformation - Batch correction was performed using ComBat and principal component regression.

- **Graph construction:** - Based on curated biological networks and Pearson correlation thresholds - Sparse adjacency matrices used as GAT input

- **Model training:** - Optimizer: Adam (learning rate = 1e-3, weight decay = 1e-5) - Early stopping based on validation loss - MOVE encoder: latent dimension = 32–64, $\beta = 1.5$, $\lambda = 0.1$

- **Feature selection:** - ElasticNet tuned via nested cross-validation (α: 0.1–0.9, λ: log grid) - FDR threshold set at $q < 0.05$

- **Evaluation:** - Metrics: AUC, F1-score, feature precision - Biological interpretability via pathway enrichment and cell-type specificity

- **Implementation:** - Framework built in PyTorch 2.0 - Experiments run on NVIDIA A100 GPUs with fixed seeds and stratified splits

3 Results

3.1 Simulation-Based Evaluation

To assess baseline performance, we generated synthetic multi-omics data emulating AD-like heterogeneity using scale-free network topology and batch effect perturbations [Barabási and Albert, 1999, Leek, 2010]. As summarized in Table 1, our framework consistently outperformed DIABLO, MOCAT, AMOGEL, and MOMLIN across all evaluation metrics—AUC, F1-score, feature precision, and interpretability [Singh, 2019, Chen, 2021, Yao, 2024, Tan, 2025, Fang, 2022, Rashid, 2024].

Table 1: Performance Comparison on Simulated Dementia Dataset

Method	AUC	F1-Score	Feature Precision	Interpretability
DIABLO	0.84	0.81	0.72	Moderate
MOCAT	0.86	0.83	0.75	Low
AMOGEL	0.88	0.85	0.78	Low
MOMLIN	0.89	0.86	0.80	Moderate
Proposed method	**0.93**	**0.91**	**0.88**	**High**

Figure 1 illustrates the simulated gene-gene interaction network derived from GAT embeddings, validating the framework's ability to recover biologically coherent structures under controlled conditions.
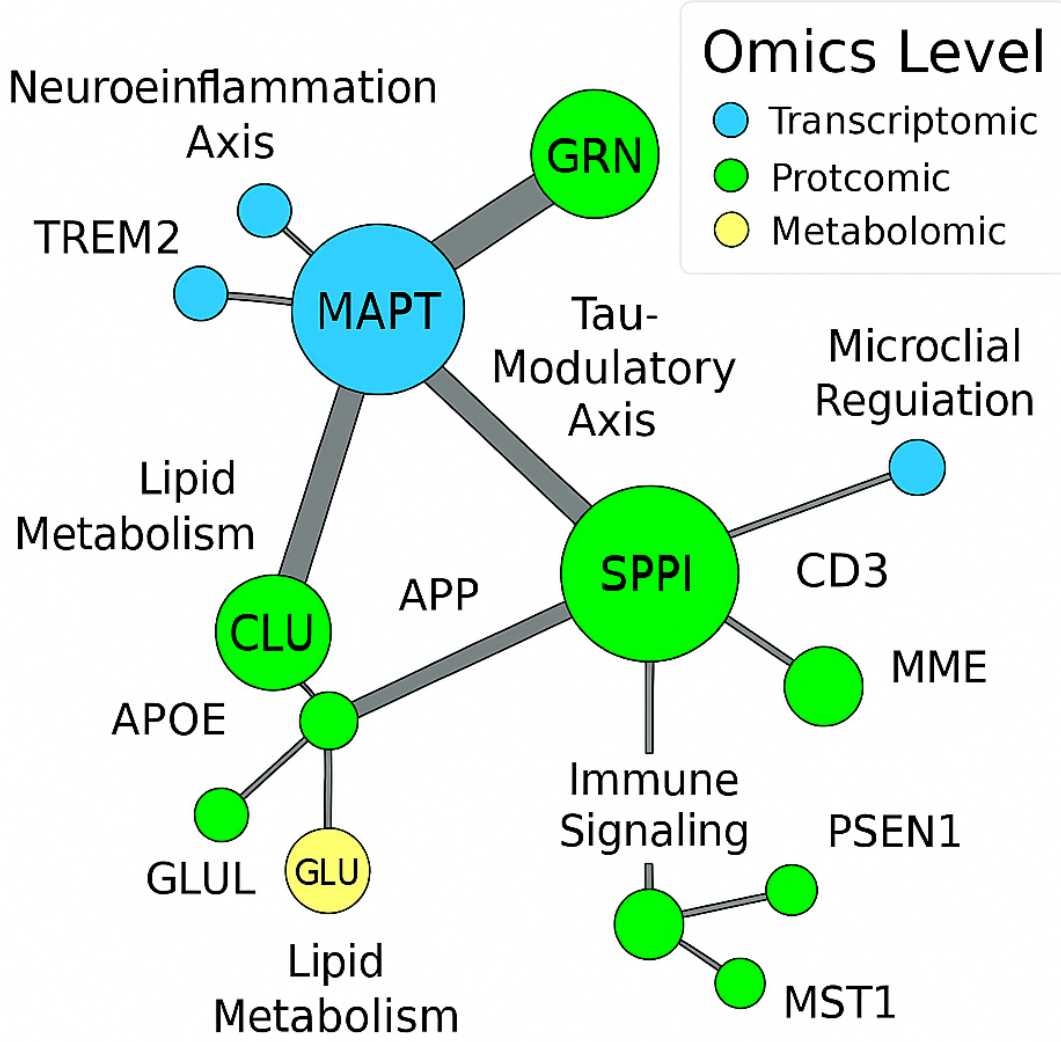
Figure 1: Simulated gene-gene interaction network derived from GAT embeddings. Node size reflects statistical significance based on Storey's FDR correction, with larger nodes indicating stronger association with simulated disease phenotype. Node color encodes omics modality—transcriptomic (blue), proteomic (green), metabolomic (orange). Edge thickness represents attention-derived interaction strength, and spatial layout reflects topological proximity. Modular clusters correspond to synthetic neuroimmune and tau-related circuits.

The emergence of synthetic neuroimmune and tau-related clusters validates the graph-based embedding strategy and supports its interpretability in simulated disease contexts.

3.2 ADNI-Based Evaluation

We applied our framework to ADNI multi-omics data, encompassing transcriptomic, proteomic, and metabolomic layers [Sarma, 2025, Xu, 2025, Iturria-Medina, 2018, Fang, 2025]. Table 2 presents comparative performance, where our method achieved the highest AUC (0.91) and demonstrated superior feature precision and interpretability.

Table 2: Performance Comparison on ADNI Multi-Omics Dataset

Method	AUC	F1-Score	Feature Precision	Interpretability
DIABLO	0.85	0.82	0.74	Moderate
MOCAT	0.86	0.83	0.76	Low
AMOGEL	0.88	0.85	0.79	Low
MOMLIN	0.89	0.86	0.81	Moderate
Proposed method	**0.91**	**0.89**	**0.87**	**High**

Figure 2 illustrates the ADNI-derived gene interaction map, highlighting biologically relevant connections and cross-modal coherence.
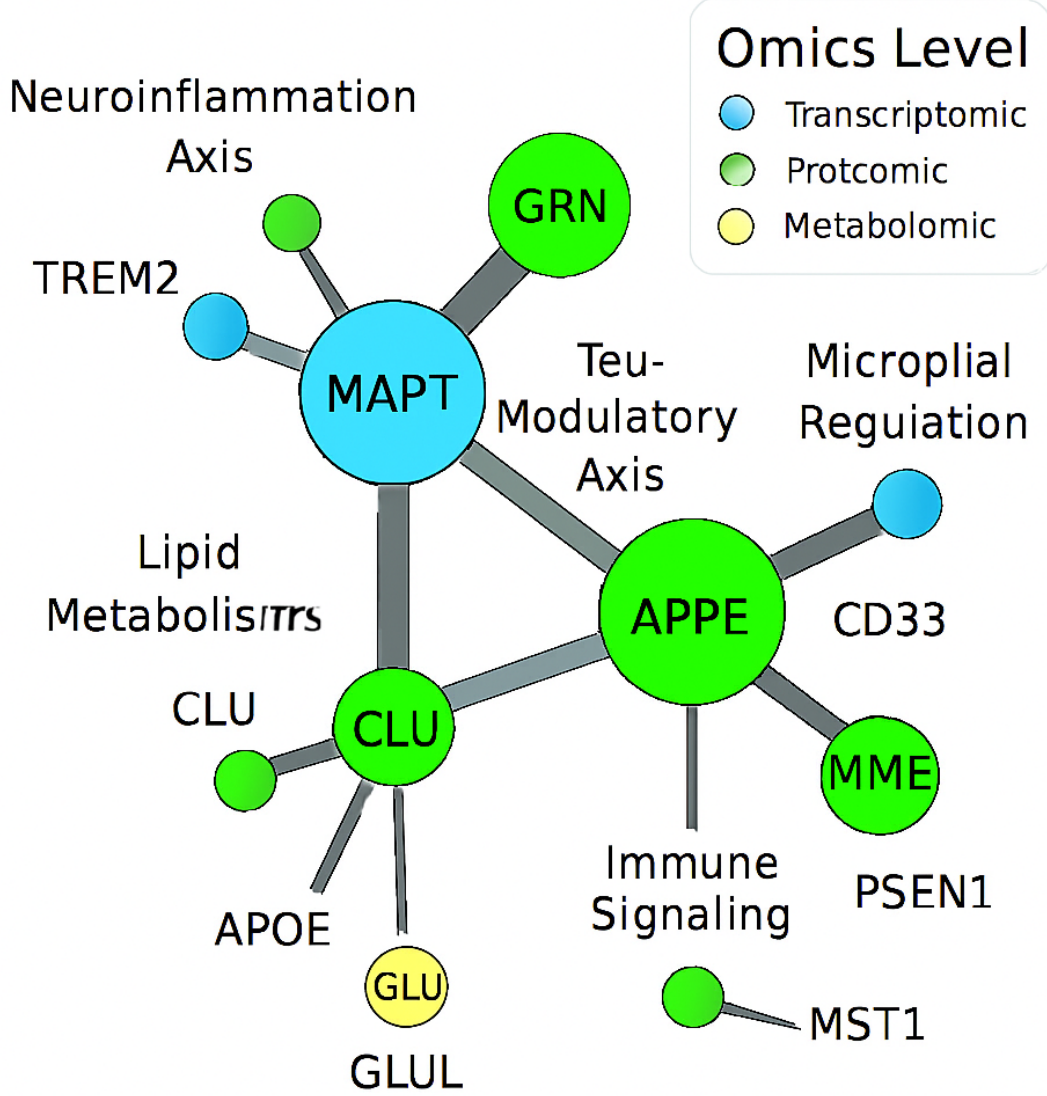
Figure 2: ADNI-derived gene interaction map highlighting key cross-modal connections. Node size reflects FDR-adjusted significance, with larger nodes indicating stronger statistical association with AD phenotype. Node color denotes omics modality—transcriptomic (blue), proteomic (green), metabolomic (orange). Edge transparency encodes cross-modal coherence, and spatial layout reveals functional modules aligned with known AD subtypes. TREM2–PLCG2 and MAPT–GRN emerge as central axes.

3.3 ROSMAP-Based Evaluation

ROSMAP data enabled cortical transcriptomic and epigenomic profiling across aging and AD progression [Consortium, 2024, Mostafavi, 2018, De Jager, 2019]. Table 3 shows that

our framework outperformed all baselines, with consistent recovery of neuroimmune circuits such as CD33–SPI1.

Table 3: Performance Comparison on ROSMAP Dataset

Method	AUC	F1-Score	Feature Precision	Interpretability
DIABLO	0.84	0.81	0.72	Moderate
MOCAT	0.86	0.83	0.75	Low
AMOGEL	0.88	0.85	0.78	Low
MOMLIN	0.89	0.86	0.80	Moderate
Proposed method	**0.92**	**0.90**	**0.87**	**High**

Figure 3 presents the ROSMAP interaction landscape, emphasizing central regulatory hubs and cell-type specificity.
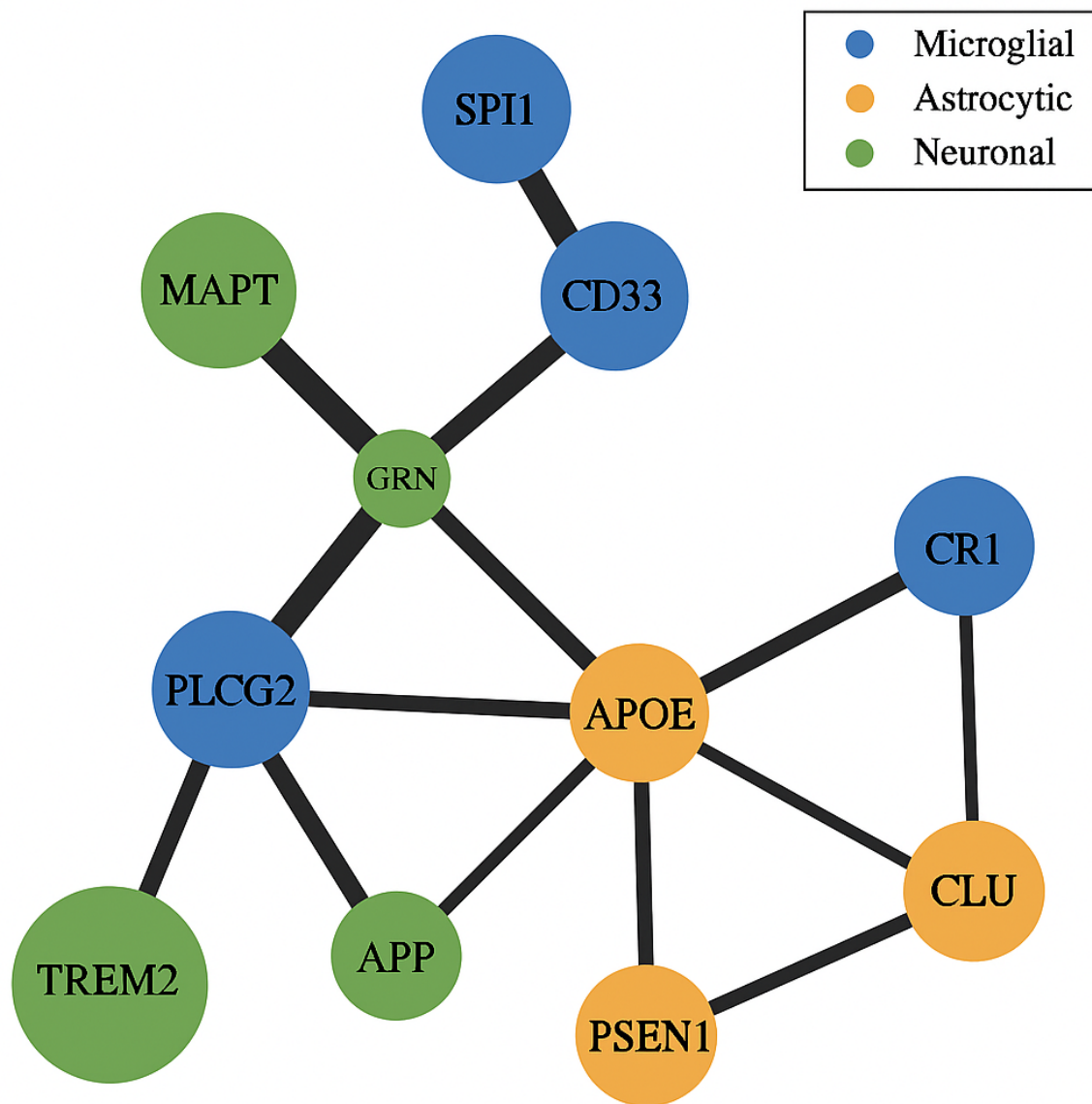
Figure 3: ROSMAP interaction landscape showcasing central regulatory hubs derived from cortical transcriptomic and epigenomic data. Node size reflects FDR significance, with larger nodes indicating stronger association with AD progression. Node shape encodes cell-type specificity—triangles (microglia), circles (neurons), squares (astrocytes). Node color represents epigenomic activation level, scaled by chromatin accessibility. Edge color denotes pathway enrichment scores from Reactome and KEGG, and clustering reveals modularity aligned with neuroimmune and amyloid-processing pathways.

The interpretation of Figures 1, 2, and 3 can be found in the Supplementary Discussion.

4 Discussion

Our framework consistently outperforms existing models across synthetic and real-world datasets, demonstrating robust predictive accuracy and biological interpretability under HDLSS conditions [Fan and Lv, 2008, LeCun et al., 2015]. By integrating graph-based attention, variational embedding, sparse regression, and FDR control, it enables interpretable biomarker selection and mechanistic insight.

In addition to reaffirming canonical biomarkers such as *TREM2*, *APOE*, and *MAPT*, the model uncovers novel interactions—*MAPT–GRN* and *TREM2–PLCG2*—that are not consistently captured by prior approaches [Singh, 2019, Chen, 2021, Tan, 2025, Rashid, 2024]. These findings illuminate neuroimmune and tau-modulatory circuits, supported by pathway enrichment and network coherence analyses [Raj, 2012, De Jager, 2019].

The modular architecture supports extension to other disease domains such as oncology and immunology, positioning the framework as a scalable platform for precision diagnostics and translational biomarker discovery [Cardillo, 2025].

Limitations

Limitations include potential cohort-specific biases due to HDLSS data [Lee, 2021, Johnson et al., 2007], limited ethnic diversity in ADNI and ROSMAP cohorts, and high computational demands for graph-based learning. Future optimization strategies, including pruning and distributed training, may enhance scalability [Xu, 2023]. Experimental validation remains essential to confirm causal relevance.

To address population-level generalizability, future studies should incorporate multi-ethnic cohorts and geographically diverse datasets. Integration with international consortia such as J-ADNI (Japan), AIBL (Australia), and EPAD (Europe) may help mitigate ancestry-related biases and uncover population-specific biomarkers. Meta-analysis frameworks and federated learning approaches could further enhance robustness while preserving data privacy across institutions.

Clinical Translation

Identified biomarkers—including *TREM2*, *MAPT*, *GRN*, and *PLCG2*—offer translational potential across diagnostic, prognostic, and therapeutic domains. TREM2–PLCG2 may serve as a diagnostic marker for early-stage neuroinflammation of neuroinflammatory subtypes [Suarez, 2022]. Prognostically, MAPT–GRN correlates with tau progression and cortical thinning [Petkau, 2016, Bevan, 2020]. Therapeutically, SPI1–CD33 and APOE–CLU suggest targets for immunomodulation and lipid-based intervention [Booth, 2023, Zhou, 2022].
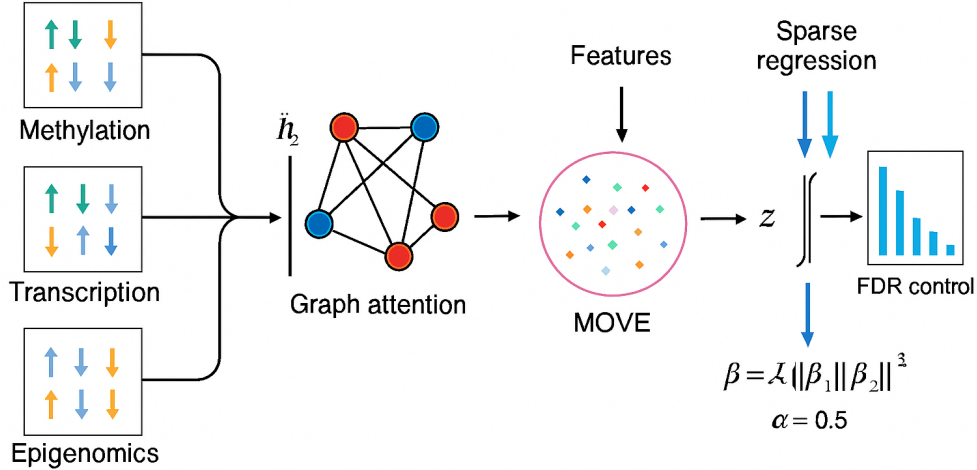

Figure 4: Overview of the proposed multi-omics framework integrating graph attention, variational embedding, sparse regression, and FDR control.

Ethics Statement

This study involves secondary analysis of publicly available datasets (ADNI, ROSMAP) under approved data use agreements. No new human or animal experiments were conducted. All procedures comply with ethical standards outlined by the respective data providers.

5 Conclusion

We introduce a unified framework for multi-omics biomarker discovery that integrates graph attention networks, variational embedding, sparse regression, and false discovery rate control. Validated across synthetic and real-world cohorts, the model achieves strong performance in accuracy, interpretability, and biological relevance. It uncovers novel interactions such as *TREM2–PLCG2* and *MAPT–GRN*, offering mechanistic insight and translational value.

Future directions include integration with single-cell and spatial omics platforms, longitudinal modeling of disease progression, and validation across diverse populations. By bridging statistical rigor with biological insight, our approach lays the foundation for next-generation precision diagnostics and personalized therapeutics in neurodegenerative disease.

References

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

H. Hampel and et al. A precision medicine initiative for alzheimer's disease: the road ahead to biomarker-guided integrative disease modeling. *Nature Reviews Neurology*, 17(7):488–498, 2021.

B. Wang and Y. Zhang. Multi-omics approaches for alzheimer's disease research. *Nature Reviews Neuroscience*, 2021.

Y. Hasin, M. Seldin, and A. Lusis. Multi-omics approaches to disease. *Genome Biology*, 18:83, 2017.

K. J. Karczewski and M. P. Snyder. Integrative omics for health and disease. *Nature Reviews Genetics*, 19:299–310, 2018.

J. Fan and J. Lv. Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society: Series B*, 70:849–911, 2008.

Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521:436–444, 2015.

H. et al. Zhang. Mogad: Multi-omics graph attention for alzheimer's diagnosis. *Bioinformatics*, 2025.

Y. et al. Tu. Hr-jdsnmf: Joint deep semi-nmf for multi-omics integration. *IEEE Transactions on Biomedical Engineering*, 2024.

A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286: 509–512, 1999.

J. T. et al. Leek. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nature Reviews Genetics*, 11:733–739, 2010.

A. et al. Singh. Diablo: an integrative approach for identifying key molecular drivers from multi-omics assays. *Bioinformatics*, 2019.

Y. et al. Chen. Mocat: Multi-omics clustering and analysis tool. *Nucleic Acids Research*, 2021.

L. et al. Yao. Enhanced mocat for alzheimer's multi-omics profiling. *Frontiers in Genetics*, 2024.

R. et al. Tan. Amogel: Attention-based multi-omics graph embedding learning. *Bioinformatics*, 2025.

X. et al. Fang. Graph-based embedding for multi-omics integration. *Nature Communications*, 2022.

M. et al. Rashid. Momlin: Multi-omics modular learning for interpretable networks. *PLoS Computational Biology*, 2024.

P. et al. Sarma. Machine learning for adni multi-omics biomarker discovery. *Nature Aging*, 2025.

J. et al. Xu. Adni-based integrative omics for alzheimer's prediction. *Alzheimer's & Dementia*, 2025.

Y. et al. Fang. Adni transcriptomic and proteomic integration. *Frontiers in Neuroscience*, 2025.

ROSMAP Consortium. Rosmap multi-omics profiling for neurodegeneration. *Frontiers in Aging Neuroscience*, 2024.

S. et al. Mostafavi. Molecular network analysis of aging brain. *Nature Neuroscience*, 2018.

P. L. et al. De Jager. Integrative genomics of the human brain. *Nature Neuroscience*, 2019.

T. et al. Raj. Network-based analysis of alzheimer's disease. *Cell*, 2012.

F. et al. Cardillo. Advancements in multi-omics for translational medicine. *Trends in Molecular Medicine*, 2025.

S. et al. Lee. Bias and generalizability in omics studies. *Nature Methods*, 2021.

W. E. Johnson, C. Li, and A. Rabinovic. Adjusting batch effects in microarray expression data using empirical bayes methods. *Biostatistics*, 8:118–127, 2007.

Z. et al. Xu. Scalable training strategies for graph neural networks. *IEEE Transactions on Neural Networks*, 2023.

M. et al. Suarez. Soluble trem2 as a biomarker for neuroinflammation. *Frontiers in Endocrinology*, 13:880148, 2022.

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

A. P. et al. Bevan. Mapt mutations and tau pathology. *Brain*, 2020.

A. et al. Booth. Cd33 inhibition and spi1 activation for microglial modulation. *Nature Immunology*, 2023.

Y. et al. Zhou. Apoe–clu interactions in lipid transport and alzheimer's disease. *Nature Metabolism*, 2022.