

METHODOLOGY

Open Access



MosGraphFlow: a novel integrative graph AI model mining signaling targets from multi-omic data

Heming Zhang¹, Dekang Cao^{1,2}, Tim Xu^{1,2}, Emily Chen^{1,5}, Guangfu Li⁶, Yixin Chen², Philip Payne¹, Michael Province³ and Fuhai Li^{1,4*}

Abstract

Multi-omic dataset can better characterize complex cellular signaling pathways from multiple views compared to individual omic data. However, integrative multi-omic data analysis to rank key disease biomarkers and infer core signaling pathways remains an open problem. In this study, we developed a novel graph AI model, mosGraphFlow, for analyzing multi-omic signaling graphs (mosGraphs), 2) analyzed multi-omic mosGraph datasets of Alzheimer's Disease (AD), and 3) developed a visualization tool to facilitate the visualization of identified disease associated signaling biomarkers and network. The comparison results show that the proposed model not only achieves the best classification accuracy but also identifies important AD disease biomarkers and signaling interactions. In the visualization, the signaling sources are highlighted at specific omic levels to facilitate the understanding of disease pathogenesis. The proposed model can also be applied and expanded for other multi-omic data-driven studies. The code of the model is publicly accessible via GitHub: <https://github.com/FuhaiLiAiLab/mosGraphFlow>.

Keywords Multi-omic, Graph AI, Alzheimer's Disease, Biomarkers discovery

Introduction

The advent of multi-omic data has revolutionized the field of biomedical research by providing a comprehensive view of the complex biological processes underlying various diseases. Unlike single-omic approaches, which focus on a specific type of molecular data such as genomics, transcriptomics, or proteomics, multi-omic data integrates information from multiple molecular layers to measure and characterize the multi-level molecular genotype of diseases. This integrative approach offers a more holistic understanding of cellular signaling pathways, enabling researchers to uncover intricate molecular interactions and regulatory mechanisms. Multi-omic datasets have proven invaluable in identifying essential disease biomarkers and elucidating dysfunctional signaling pathways, particularly in understanding the genetic heterogeneity of diseases at multiple levels. Despite its

*Correspondence:

Fuhai Li

Fuhai.Li@wustl.edu

¹Institute for Informatics, Data Science and Biostatistics (I2DB), Washington University School of Medicine, St. Louis, MO 63110, USA

²Department of Computer Science and Engineering, Washington University in St. Louis, St. Louis, MO 63130, USA

³Division of Statistical Genomics, Washington University School of Medicine, MO 63110 St. Louis, USA

⁴Department of Pediatrics, Washington University School of Medicine, St. Louis, MO 63110, USA

⁵School of Arts and Sciences, University of Rochester, Rochester, NY 14627, USA

⁶Department of Surgery, University of Connecticut School of Medicine, Farmington, CT 06030, USA



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

potential and demonstrated utility, the effective integration and analysis of multi-omic data to identify key disease biomarkers and elucidate core signaling pathways remain significant challenges. Traditional AI models often struggle to fully leverage the richness of multi-omic data due to its complexity and high dimensionality. However, recent advancements in graph AI models have shown promise in addressing these challenges by utilizing graph-based representations to capture the intricate relationships within multi-omic datasets, offering new avenues for biomarker discovery and pathway inference. This approach can be instrumental in enhancing our understanding of disease pathogenesis and in designing more effective therapeutic interventions.

Alzheimer's Disease (AD) is the most prevalent cause of dementia, primarily affecting individuals over the age of 65, though cases in younger individuals starting from around age 40 are increasingly observed. Characterized by progressive cognitive impairment, AD manifests through the hallmark neuropathological features, extracellular amyloid- β plaques and intracellular neurofibrillary tangles (NFT), caused by amyloid- β accumulation and tau hyperphosphorylation [1]. Linked to these hallmarks are blood-brain barrier disruption, mitochondrial impairment, neuroinflammation, synaptic impairment and neuronal loss. The prevalence of AD in America was estimated at 6.7 million in 2023, with projections suggesting a doubling to 13.8 million by 2060 [2]. Despite extensive research in the last century, there remains no cure for AD, and current treatments are symptomatic rather than disease-modifying. With the increasing prevalence of AD driven by an aging population, there is an urgent need for continued research into its pathogenesis and the development of more effective therapeutic interventions.

Given these challenges, leveraging multi-omic data through advanced graph AI models presents a promising frontier in AD research. By integrating multi-omic data with graph-based techniques, researchers can more effectively identify critical disease biomarkers and uncover the core signaling pathways involved in AD. This approach offers the potential to not only enhance our understanding of AD pathogenesis but also pave the way for the development of targeted and disease-modifying treatments. In this study, we explore the application of a graph AI model on multi-omic datasets to identify key biomarkers and signaling interactions in AD, demonstrating its superiority in classification accuracy and its capability to highlight significant molecular mechanisms at various omic levels.

Recently, Graph Neural Networks (GNN) have gained prominence due to their capability to model relationships within graph-structured data [3–6]. And numerous studies have applied the GNN with the integration of the multi-omics data. MOGONET [7] (Multi-Omics

Graph cONvolutional NETworks) initially creates similarity graphs among samples by leveraging each omics data, then employs a Graph Convolutional Network (GCN [3]) to learn a label distribution from each omics data independently. Subsequently, a cross-omics discovery tensor is implemented to refine the prediction by learning the dependency among multi-omics data. MoGCN [8] adopts a similar approach by constructing a patient similarity network using multi-omics data and then using GCN to predict the cancer subtype of patients. GCN-SC [9] utilizes a GCN to combine single-cell multi-omics data derived from varying sequencing methodologies. MOGCL [10] takes this further by exploiting the potency of graph contrastive learning to pretrain the GCN on the multi-omics dataset, thereby achieving impressive results in downstream tasks with fine-tuning. Nevertheless, none of the aforementioned techniques contemplate incorporating structured signaling data like KEGG into the model. Moreover, general GNN models are limited by their expression power, i.e., the low-pass filtering or over-smoothing issues, which hampers their ability to incorporate many layers. The over-smoothing problem was firstly mentioned by extending the propagation layers in GCN [11]. Moreover, theoretical papers using Dirichlet energy showed diminished discriminative power by increasing the propagation layers [12]. And multiple attempts were made to compare the expressive power of the GCNs [13] and it is shown that WL subtree kernel [14] is insufficient for capturing the graph structure. Hence, to improve the expression powerful of GNN, the K -hop information of local substructure was considered in various recent research [13, 15–19]. However, none of these studies was specifically designed to well integrate the biological regulatory network and provide the interpretation with important edges and nodes [20]. In this study, the unique and major contributions of this study are as follows: (1) developed a graph neural network (GNN) model for the mosGraphs, (2) analyzed multi-omic mosGraph datasets of AD, and (3) identified and visualized and evaluated a set of AD associated signaling biomarkers and network.

Materials and methods

Multi-omics datasets of Alzheimer's disease

To study Alzheimer's Disease, multi-omics datasets were sourced from publicly accessible databases, specifically the ROSMAP datasets (Table 1, and Table S1). For more details of data and pre-analysis, please refer to details in Appendix Section A and the pre-analysis protocol [20]. Specifically, upon downloading these datasets, they were transformed into 2-dimensional data frames, structured with columns for sample IDs, sample names, etc., and rows for probes, gene symbols, gene IDs, etc. Integrating multi-omic data with clinical data required identifying identical samples across the datasets. This

Table 1 Dimensions of the ROSMAP multi-omics data before and after the data processing

Database	Before process	After process
Methylation dataset	(420132 probes, 740 samples)	(2099 genes, 437 samples)
Proteomics dataset	(8817 genes, 400 samples)	(2099 genes, 437 samples)
RNAseq dataset	(55889 gene IDs, 640 samples)	(2099 genes, 437 samples)
Copy number variations dataset	(9902 chroms, 1127 samples)	(2099 genes, 437 samples)
Clinical dataset	(16 features, 3584 samples)	(14 features, 437 samples)

process involved standardizing the rows (probes, gene symbols, gene IDs, etc.) into a uniform gene-level format by either aggregating measurements for each gene or removing duplicates caused by gene synonyms. Distinct preprocessing steps were then applied to each omics type to ensure compatibility within the mosGraphFlow model. Methylation data were mapped from CpG sites to the gene level, focusing on the -6 kb to $+3$ kb region around transcription start sites (TSS). The data were divided into five predefined genomic regions (upstream, distal promoter, proximal promoter, core promoter, and downstream regions), followed by integration and padding. Genomics data (mutations and variants), filtered to retain only non-silent mutations, were categorized into deletions (DEL), duplications (DUP), and multiple copy number variations (mCNV), and duplicate genes were aggregated to capture the variation across instances. Transcriptomic data were normalized using the FPKM method, probe IDs were mapped to genes, and duplicate entries were averaged to ensure uniqueness. Proteomics data were standardized using median log2 normalization, and duplicate genes were aggregated to unify gene-level features. All datasets were aligned using the Ensembl reference database and integrated into a unified gene set for further analysis. Finally, sample IDs were cross-referenced across all omics layers to ensure consistency in patient samples, resulting in consistent overlapping samples for all datasets. Once all columns were aligned to standard sample IDs and all rows to standard gene IDs, and the number of samples and genes were unified, the data was ready for integration into graph AI models (see Table 1).

KEGG regulatory network construction

For constructing the knowledge graph, genes were selected by intersecting the multi-omics datasets with gene regulatory networks from the KEGG database, which includes 2241 genes and 21,041 edges. This intersection resulted in 2099 gene entities. To integrate these KEGG pathways into the mosGraphFlow model, protein-protein interactions (PPI) and gene-regulatory pathways

were mapped to the processed multi-omics data by aligning gene identifiers across datasets. For each omics layer, genes with corresponding KEGG annotations were linked to relevant pathway information. This alignment was facilitated by intersecting KEGG pathway genes with the unified gene set from our multi-omic data. The resulting network incorporated not only individual omic features but also their relational interactions as described in KEGG, providing a comprehensive representation of signaling pathways. Edge connections in the graph model were derived directly from KEGG's documented interactions, ensuring that the constructed network reflects real biological processes and supports meaningful pathway-based inference. This integration strategy enables the mosGraphFlow model to leverage biological context in its analysis, enhancing interpretability in the context of AD.

Architecture of the mosGraphFlow

The proposed mosGraphFlow model enhances the analysis and prediction capabilities in multi-omics data, which aims to provide a comprehensive and interpretable analysis of AD dataset. The integrated approach offers a robust solution for multi-omics data analysis with generation of $\mathcal{G} = (V, E)$, where $|V| = n$. In details, there are 3 types of nodes in the graph. $n^{(meth)}$, $n^{(gene)}$ and $n^{(prot)}$ have the same number of nodes and $n = n^{(meth)} + n^{(gene)} + n^{(prot)}$. Furthermore, the whole graph $\mathcal{G}$ can be decomposed into subgraphs $\mathcal{G}'$ and $\mathcal{G}_{PPI}$, where $\mathcal{G}' = \mathcal{G} \setminus \mathcal{G}_{PPI}$; $\mathcal{G}'$ is the internal signaling flow graph which only contains the signaling flows from promoters to proteins; $\mathcal{G}_{PPI}$ is the protein-protein interaction (PPI) graph ($|V_{PPI}| = n^{(prot)}$). Correspondingly, the adjacency matrices A , A' and A_{PPI} for whole graph $\mathcal{G}$, internal graph $\mathcal{G}'$ and PPI graph $\mathcal{G}_{PPI}$ will be generated. And the proposed model can be denoted as $f(\cdot)$, the graph-based deep learning model. It will predict the patient outcome $Y (Y \in \mathbb{R}^{M \times 1})$, constructed with $f(\mathcal{X}, A, A_{in}, S_{PPI}) = Y$, where $\mathcal{X} = \{X^{(1)}, X^{(2)}, \dots, X^{(m)}, \dots, X^{(M)}\}$ ($X^{(m)} \in \mathbb{R}^{n \times d}$) denotes all M data points in the dataset and $X^{(m)}$ is the m -th data points. And $A (A \in \mathbb{R}^{n \times n})$ is the adjacency matrix that demonstrates the node-node interactions, and the element in adjacency matrix A such as a_{ij} indicates an edge from i to j . A' ($A' \in \mathbb{R}^{n \times n}$) is the adjacency matrix which only includes the node interactions from promoters to proteins, corresponding to the graph $\mathcal{G}'$. Regarding the set of subgraphs in the PPI, $S_{PPI} = \{S_1, S_2, \dots, S_p, \dots, S_P\}$ ($S_p \in \mathbb{R}^{n_p \times n_p}$), these subgraphs will partition the whole PPI graph adjacent matrix A_{PPI} into multiple subgraphs with the annotation of each individual signaling pathway, where the vertices in these partitioned subgraphs can be denoted

as $V_1, V_2, \dots, V_p, \dots, V_P$, where $V_{PPI} = \bigcup_{p=1}^P V_p$. In each subgraph S_p , there are nodes interactions between its internal n_p nodes and each subgraph have its own corresponding subgraph node feature matrix $X_p \in \mathbb{R}^{n_p \times d}$.

Internal modular message propagation

In the graph message passing stages of our architecture, we introduced the message passing between the internal links via matrix A' with following formula: $H_{in}^{(m)} = \text{GNN}_{in}(X^{(m)}, A') \# (1)$

, where GNN_{in} is the transformer-based message propagation function [21] and $H_{in}^{(m)} \in \mathbb{R}^{n \times d^{(in)}}$ is the embedded node features for sample m after internal message propagation.

Multi-hop message propagation in signaling pathway subgraphs

Following the internal message passing stages, the local structure for each signaling pathway subgraph can be integrated via following formula:

$$(H_{hop}^{(m)})^p = \text{AVG}[\text{GNN}_{hop}(H_{in}^{(m)}, S_p)] \# (2)$$

, where GNN_{hop} is K -hop attention-based graph neural network [22]; $(H_{hop}^{(m)})^p \in \mathbb{R}^{n_p \times d^{(hop)}}$ is the embedded node features for sample m after K -hop attention-based message propagation and average (AVG) is mean aggregation function from PyTorch package. Based on this, the aggregated node features $(H_{hop}^{(m)})^p \in \mathbb{R}^{n \times d^{(hop)}}$ will be generated by averaging the node information included over multiple sub-signaling pathways in the set S_{PPI} with details of

$$(H_{hop}^{(m)})_i = \frac{1}{\sum_{p=1}^P I[V_i \in S_p]} \sum_{p=1}^P (H_{hop}^{(m)})_i^p \cdot I[V_i \in S_p] \# (3)$$

, where $(H_{hop}^{(m)})_i^p \in \mathbb{R}^{d^{(hop)}}$ is the embedded node information for i -th row in the subgraph S_p from the sample m and $(H_{hop}^{(m)})_i \in \mathbb{R}^{d^{(hop)}}$ will be concatenated for n nodes, resulting the aggregated node features $H_{hop}^{(m)}$ ($H_{hop}^{(m)} \in \mathbb{R}^{n \times d^{(hop)}}$).

Global bi-directional message propagation

Following the message propagation in the multiple internal subgraphs, the global weighted bi-directional message propagation [23] will be performed via formula

$$H_{out}^{(m)} = \text{WeBGNN}(H_{hop}^{(m)}, A) \# (4)$$

$$O^{(m)} = \text{OUT}(H_{out}^{(m)}) \# (5)$$

, where WeBGNN (Weighted Bi-directional Graph Neural Network) [24] is the graph signaling flow framework and $H_{out}^{(m)} \in \mathbb{R}^{n \times d^{(out)}}$. And linear transformation function $g: \mathbb{R}^{d^{(out)}} \rightarrow \mathbb{R}$ will be applied as the output (OUT) function to predict the patient outcome with $O^{(m)}$.

Sample prediction and loss calculation

Based on the patient outcome $O^{(m)}$, softmax function was leveraged for classification by

$$\hat{Y}^{(m)} = \text{softmax}(O^{(m)}) \# (6)$$

, and cross entropy (CE) function was used for loss calculation by

$$L^{(m)} = \text{CE}(Y^{(m)}, \hat{Y}^{(m)}) \# (7)$$

, where $\hat{Y}^{(m)}$ and $Y^{(m)}$ are predicted and original sample types.

Signaling pathway inference

To interpret the underlying mechanisms of AD, the best-performing model was selected after training and validation process, then it was analyzed to extract attention scores from various graphs, which were used to infer signaling pathways related to the disease and key nodes (genes, promoters, and proteins). For each patient, the attention matrices of 1-hop neighbor nodes were calculated in every fold of the cross-validation process. Depending on the specific analysis, patients were stratified into different categories based on either their AD status (AD or non-AD) or gender (female or male), and for each category, the average attention matrices were computed. To quantitatively assess the significance of each node within these networks, the weighted degree of each node for every patient was calculated based on these attention scores, as detailed in the following formula:

$$\bar{W}^{(m)} = \frac{1}{K} \sum_{k=1}^K (W_k^{(m)})^{(1)} \# (8)$$

$$\bar{W}^{(c)} = \frac{1}{|\mathcal{X}^{(c)}|} \sum_{m \in \mathcal{X}^{(c)}} \bar{W}^{(m)} \# (9)$$

$$D_g^{(c)} = \sum_i^n W_{ig}^{(c)} + \sum_j^n W_{gj}^{(c)} \# (10)$$

, where $\left(W_k^{(m)}\right)^{(1)} \in \mathbb{R}^{n^{(prot)} \times n^{(prot)}}$ represents the attention-based matrix extracted from the first hop attention for patient m in the k -th fold; $W_{ig}^{(c)} \in \mathbb{R}$ denotes the element of K -fold averaged attention matrix for patient m in the i -th row and j -th column for patient type c ; $D_g^{(c)} \in \mathbb{R}$ represents the node degree, which quantifies the importance of the node g within the network from the type of patient c .

Afterwards, the unimportant signaling flows in the attention-based matrix for certain type of patient will be filtered out by

$$W_F^{(c)} = F(W^{(c)}, \theta) \# (11)$$

, where $F(\cdot)$ is the filtering mapping function by providing selection of each element in the matrix with

$$F(w, \theta) = \begin{cases} w, & \text{if } w > \theta \\ 0, & \text{if } w \leq \theta \end{cases} \# (12)$$

, where $w \in \mathbb{R}$ is the element in the input matrix and $W_F^{(c)} \in \mathbb{R}^{n^{(prot)} \times n^{(prot)}}$ is the filtered matrix. Hence, the filtered node set for patient type c , $V_F^{(c)}$, will be generated by removing independent nodes and nodes in those small connected components with number of nodes lower than φ , resulting in $|V_F^{(c)}|$ nodes.

Sample-specific network visualization

The distinctions between AD/non-AD or female/male AD patients were made, and the relevant features for each group were identified. Subsequently, p-values for the gene features, such as methylations in promoter nodes, mutations and genes expression in gene nodes and proteins expression in protein nodes were calculated. The p-value calculation for these features was conducted by using the chi-squared test to check the differences between AD/non-AD samples or female/male of AD patients. This statistical method determined whether there were significant differences in the gene features between the samples of AD/non-AD or female/male from AD. By constructing contingency tables and performing the chi-squared test for each gene feature, p-values indicating the statistical significance of the observed differences were obtained. Ultimately, the top T gene features associated with AD or gender were selected based on these p-values.

After finalized important gene features ranked by p-values in top T , the network was pruned by iteratively

removing the nodes which are only connected to one another unimportant node in a linear branch with node recursive algorithm (check details of this algorithm in Appendix Section B). This ensures that each remaining nodes is either linked to an important node or is part of a more complex interaction network, enhancing the purity and reliability of the gene interaction data.

Subsequently, nodes degree were calculated to identify hub node (node degree larger than 2). The set of middle nodes for certain path t which connects two hub nodes u and v can be denoted as $P_{u \rightarrow v}^{(t)} = \{n_1, n_2, \dots, n_r, \dots, n_R\}^{(t)}$, where $\lambda + 1$ is the length of path. Hence, the average edge weight on the path $P_{u \rightarrow v}^{(t)}$ can be generated by

$$O_{u \rightarrow v}^{(t)} = \frac{1}{\lambda} \sum_{r=1}^{\lambda-1} W_{n_r, n_{r+1}}^{(c)} \# (13)$$

, where $W_{n_r, n_{r+1}}^{(c)}$ is the edge weight from node n_r to node n_{r+1} . For all of the paths detected between the hub node u and hub node v , the nodes on the top β paths will be kept. Additionally, p-value middle nodes, which are crucial due to their statistical significance, will be retained along with middle nodes that are adjacent to these p-value nodes.

Results

Overview of the workflow in mosGraphFlow

Figure 1 shows the overview of the workflow in mosGraphFlow. Initially, multi-omic datasets, including methylation, mutation, gene expression, and protein expression, were collected, preprocessed, and aligned to corresponding genes or proteins. These pre-analyzed datasets were subsequently mapped onto signaling networks, such as KEGG pathways, where genes were represented as nodes and their interactions as edges. To encode the multi-omic data, message propagation operations were performed at multiple levels, integrating the data into protein nodes. This process included modular multi-hop message propagation within individual signaling pathways, followed by global propagation across pathways to generate final protein node encodings. These encodings were utilized for sample classification tasks, optimized using a cross-entropy loss function during model training. Finally, attention scores derived from nodes and edges were employed to identify and rank key signaling targets. Sub-signaling networks linking these targets were constructed to provide insights into critical molecular interactions. For detailed methodology, please refer to Materials and Methods section.

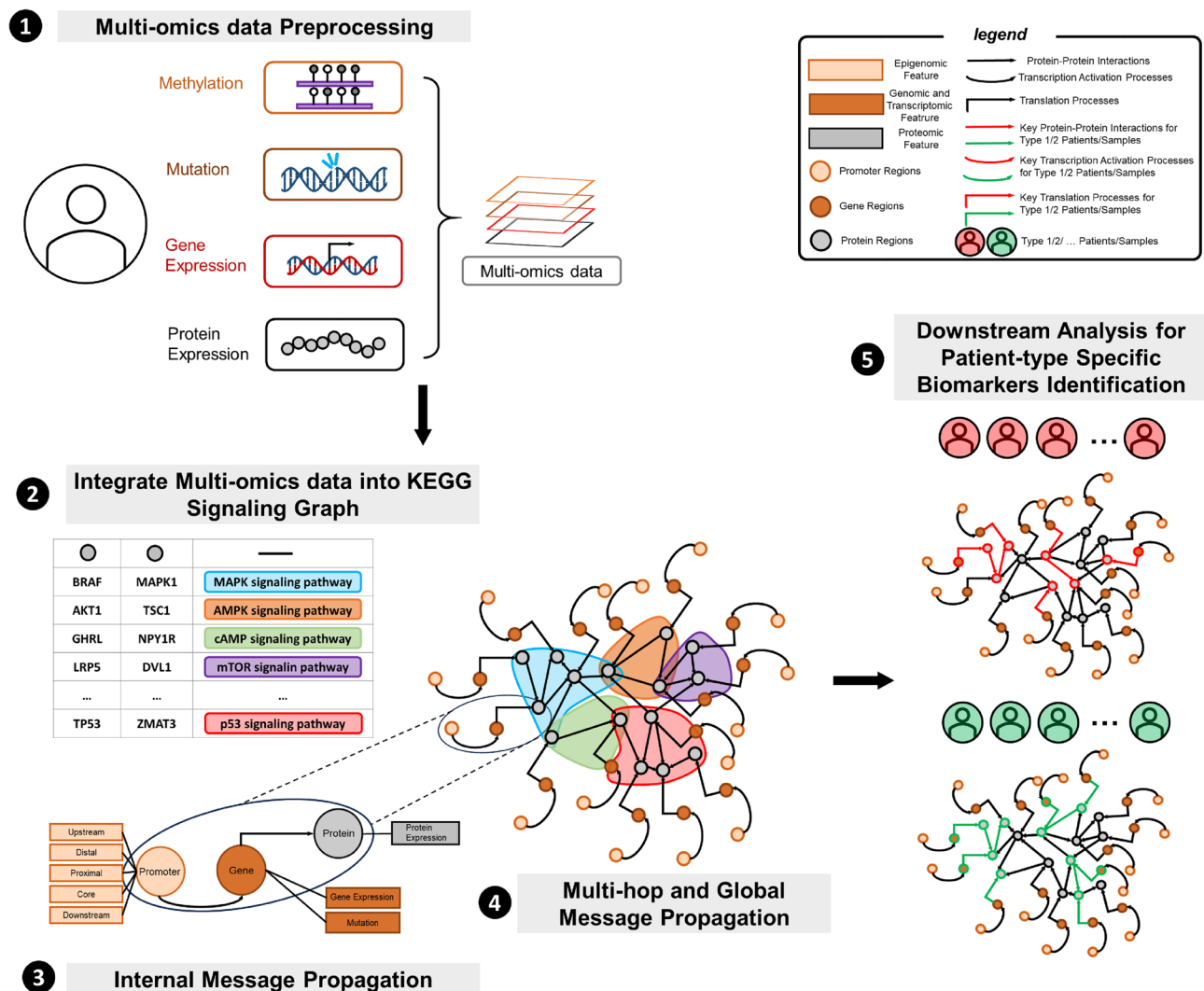


Fig. 1 Overview of workflow in mosGraphFlow. The workflow consists of 5 analysis steps. The multi-omic data were pre-analyzed and integrated with signaling pathways in steps 1–2. The protein embedding was conducted in steps 3–4 via attention-based message propagation operations. Then the key signaling targets and interactions were identified using the node/edge associated attention scores

Experimental design

Model prediction objectives using the AD multi-omic dataset

In this study, we conducted two independent analysis tasks using the AD multi-omic dataset. Specifically, the first analysis is to uncover the key disease targets and pathways by classifying AD samples from non-AD samples using the multi-omic data. The second analysis is to uncover the sex difference or the sex-/gender-specific key disease targets and pathways of AD by comparing female vs. male AD samples using the multi-omic data. The purpose of sex analysis is to identify the potentially AD associated sex biomarkers using the proposed multi-omic data analysis model because there is clear sex difference in AD. For example, AD risk in women is 2 times higher than in men [25–28]. Women show more severe AD phenotypes and faster memory decline [26, 29–31]. Therefore, it is important to study the pathogenesis and

sex factors in AD. The AD vs. non-AD status is determined using the CERAD score from the phenotypic data. Patients with CERAD scores of 1 or 2 are classified as non-AD (target value 0), while those with scores of 3 or 4 are classified as AD (target value 1). The gender classification is based on the “sex” variable in the dataset, with male patients assigned a target value of 1 and female patients a target value of 0.

Data splitting and cross-validation

We utilized 437 samples from the ROSMAP dataset, categorized by disease status (275 AD, 162 non-AD) and gender (276 females, 161 males). Among the AD samples, there were 177 females and 98 males. To address the significant data imbalance, we performed downsampling for both classification tasks. For the AD vs. non-AD classification, we downsampled the AD samples to match the

Table 2 Model comparison with other GNN network (AD/non-AD)

Models	The Average NLL on Training data	The Average NLL on Testing data	The average accuracy of Training data	The average accuracy of Testing data
Proposed model	0.5947 ± 0.0259	0.6718 ± 0.0294	0.7021 ± 0.0263	0.5919 ± 0.0548
GAT	0.6794 ± 0.0048	0.7320 ± 0.0604	0.5803 ± 0.0038	0.5586 ± 0.0182
GCN	0.6822 ± 0.0079	0.8235 ± 0.1325	0.5686 ± 0.0103	0.5557 ± 0.0084
GIN	0.6869 ± 0.0059	0.7642 ± 0.1340	0.5470 ± 0.0273	0.5121 ± 0.0411
UniMP	0.6665 ± 0.0183	0.8670 ± 0.2616	0.5910 ± 0.0228	0.5741 ± 0.0060

Table 3 Model comparison with other GNN network (female/male)

Models	The Average NLL on Training data	The Average NLL on Testing data	The average accuracy of Training data	The average accuracy of Testing data
Proposed model	0.6507 ± 0.0207	0.6772 ± 0.0092	0.6263 ± 0.0615	0.5918 ± 0.0488
GAT	0.6817 ± 0.0071	0.6914 ± 0.0203	0.6058 ± 0.0294	0.5460 ± 0.0277
GCN	0.6934 ± 0.0035	0.8908 ± 0.3025	0.5497 ± 0.0068	0.5255 ± 0.0115
GIN	0.6884 ± 0.0057	0.7876 ± 0.1397	0.5651 ± 0.0313	0.5204 ± 0.0340
UniMP	0.6629 ± 0.0145	0.6799 ± 0.0614	0.5663 ± 0.0225	0.5407 ± 0.0198

162 non-AD samples, resulting in a dataset with 162 AD and 162 non-AD samples. For the gender classification within the AD samples, we downsampled the female AD samples to match the 98 male AD samples, resulting in a balanced dataset of 98 female and 98 male AD samples. To evaluate model performance and prevent overfitting, we split the dataset into training and testing sets using a 5-fold cross-validation approach. Specifically, a 5-fold split is employed, where the data is divided into 5 subsets. In each fold, one subset is designated as the test set, while the remaining four subsets serve as the training set. This approach ensures that each data point is used for both training and testing, providing a robust measure of the model's generalizability.

Model hyperparameters

The model was implemented using PyTorch and PyTorch Geometric, with the Adam optimizer employed for training. In the K -hop message propagation stage, the number of hops was set as 3. For the AD classification task, the initial learning rate was set to 0.002, and the training epochs were empirically set to 80. For the gender classification task within AD samples, the initial learning rate was set to 0.001, and the training epochs were set to 50. The hidden dimension was set to 10, and the leaky ReLU parameter was configured to 0.1. The output dimension was initially 30, which was subsequently reduced to 1 dimension through max pooling over the receptive field in the final pooling layer.

Model evaluation and comparison

Tables 2 and 3 present the accuracy and negative log-likelihood (NLL) loss values for both the training and testing datasets, corresponding to AD/non-AD and female/male

classifications, respectively. The results demonstrate that the model achieved comparable performance across both datasets. Additionally, the proposed model was compared with other widely used models, namely GCN [32] GAT [5] GIN [6] and UniMP [33] and significantly outperformed them in both classification tasks.

Identified signaling targets and pathways of AD

Attention-based analyses were utilized to identify and refine critical molecular features associated with AD. The workflow integrates multi-omic data, signaling networks, and advanced propagation techniques to uncover key signaling pathways and molecular targets driving disease pathology. It highlights the statistical and network-based processes that led to the discovery of core pathways and distinct gene-protein interactions, emphasizing their implications for understanding the disease and developing potential therapeutic targets. In details, by setting an edge threshold θ as 0.176, and a small component threshold φ as 15, we identified 82 and 81 potential important protein nodes for AD and NOAD, respectively ($|V_F^{(AD)}| = 81$ and $|V_F^{(non-AD)}| = 82$). Then, by filtering out the p-value larger than 0.2 calculated from Mann-Whitney U test, the top 70 gene features associated with AD were selected. By pruning Linear branches and keeping the nodes via top 2 ($\beta = 2$) paths between hub nodes, we filtered out non-essential nodes, reducing the number of potential important protein nodes to 67 for AD and 66 for non-AD. The corresponding gene weights $D_g^{(c)}$ for these top 70 ($T = 70$) gene features for AD/non-AD ($c = \text{AD}; \text{non-AD}$) were calculated. These top 70 gene features and gene importance score are shown in Fig. 2 and detailed in Table 4.

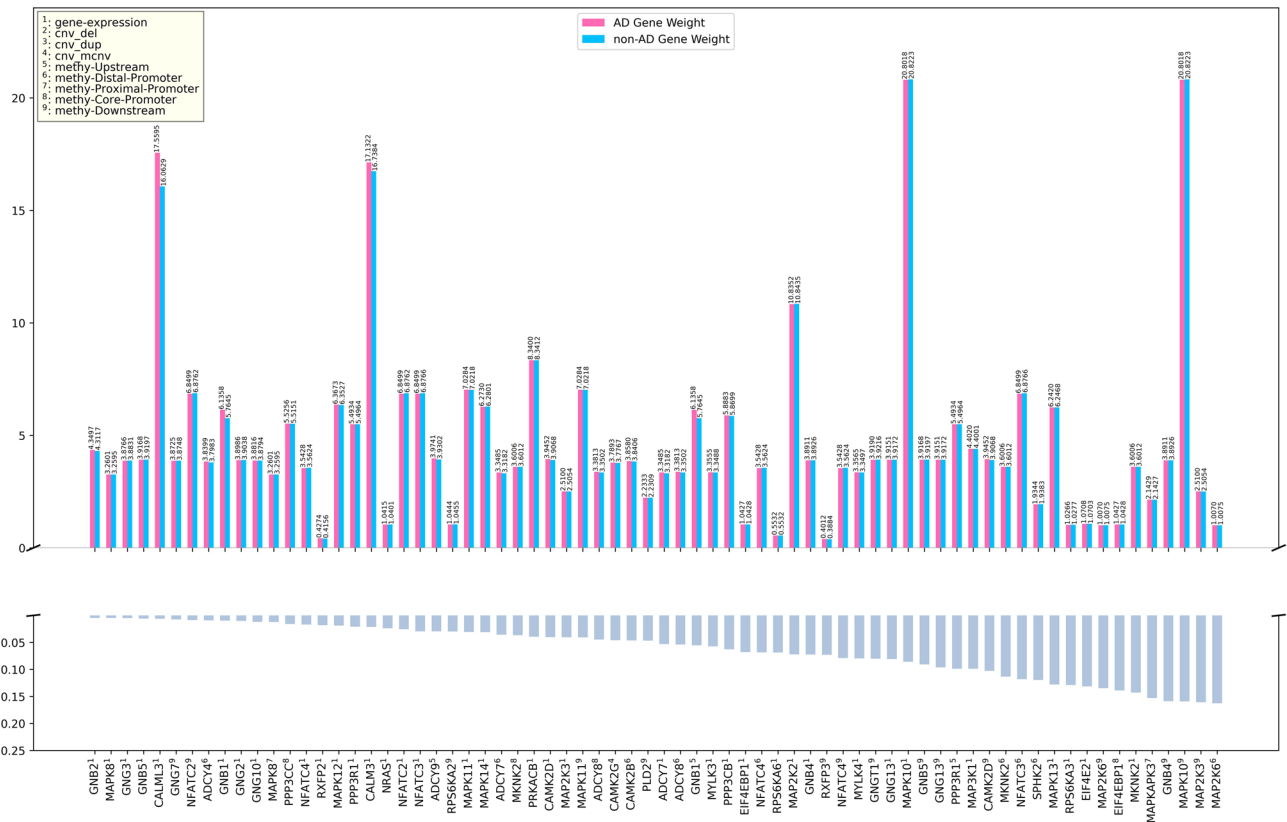


Fig. 2 Bar chart displaying the gene weights (importance scores) in AD and non-AD samples, ranking by their p-values

Figures 3 and 4 demonstrate the top 70 significant nodes within the signaling network interactions for AD samples. These networks highlight key gene features, with node sizes representing the relative importance of each feature based on their p-values. Figures 5 and 6 focus on the inferred core signaling networks, showcasing the top 70 gene features specific to female and male subjects, respectively. The analysis identifies significant genetic mutation, transcriptomic, and methylation features associated with these gene nodes. Figure 7 further details the top 70 gene features and their calculated weights, offering insights into their contributions to the network. In this analysis, through node optimization process, Like the above, we identified 62 potential important protein nodes for females and 75 for males. Notably, we observed a significant overlap between the protein nodes selected from the AD signaling networks and those from the female signaling networks. Specifically, there are 35 overlapping protein nodes between the 67 protein nodes identified in the AD signaling networks and the 62 protein nodes identified in the female signaling networks (see Appendix B Table S2). Furthermore, there is an overlap of 8 gene features between the top 70 AD/non-AD gene features and the top 70 female/male gene features (see Appendix B Table S3)

Pathway enrichment analysis of the selected AD targets

Pathway enrichment analysis was conducted on the top 70 genes associated with AD using ShinyGO 0.80 and the KEGG pathway database. This analysis revealed the top 20 signaling pathways involving these genes (see Fig. 8), enhancing our biological understanding of their roles in AD pathogenesis. To gain a comprehensive view of the complex nature of these signaling pathways, we utilized a Sankey diagram to visualize the interconnectedness between the top 70 genes and their associated pathways (see Fig. 9). The KEGG pathway database categorizes signaling pathways into seven broad categories: metabolism, genetic information processing, environmental information processing, cellular processes, organismal systems, human diseases, and drug development. However, for a more detailed focus on function or disease-related aspects, specific categories are highlighted (see Fig. 9).

Literature validations of pathways associated with AD

Various functional pathways are intricately involved in the pathogenesis of AD, making them promising therapeutic targets for intervention. Dysfunction of these pathways leads to neuronal degeneration, impaired synaptic communication, inflammatory damage, and cognitive decline in AD. These interconnected mechanisms

Table 4 Top 70 gene features associated with AD

Gene	Feature	AD-Associated Gene Weight	non-AD-Associated Gene Weight	P-Value
GNB2	gene-expression	4.3497	4.3117	0.0046
MAPK8	gene-expression	3.2601	3.2595	0.0047
GNG3	gene-expression	3.8766	3.8831	0.0049
GNB5	gene-expression	3.9168	3.9197	0.0059
CALML3	gene-expression	17.5595	16.0629	0.0062
GNG7	methy-Downstream	3.8725	3.8748	0.0075
NFATC2	methy-Downstream	6.8499	6.8762	0.0088
ADCY4	methy-Distal-Promoter	3.8399	3.7983	0.0091
GNB1	gene-expression	6.1358	5.7645	0.0097
GNG2	gene-expression	3.8986	3.9038	0.0101
GNG10	gene-expression	3.8816	3.8794	0.0120
MAPK8	methy-Proximal-Promoter	3.2601	3.2595	0.0122
PPP3CC	methy-Core-Promoter	5.5256	5.5151	0.0158
NFATC4	gene-expression	3.5428	3.5624	0.0168
RXFP2	gene-expression	0.4274	0.4156	0.0181
MAPK12	gene-expression	6.3673	6.3527	0.0189
PPP3R1	gene-expression	5.4934	5.4964	0.0209
CALM3	gene-expression	17.1322	16.7384	0.0214
NRAS	gene-expression	1.0415	1.0401	0.0237
NFATC2	gene-expression	6.8499	6.8762	0.0256
NFATC3	gene-expression	6.8499	6.8766	0.0296
ADCY9	methy-Upstream	3.9741	3.9302	0.0296
RPS6KA2	methy-Downstream	1.0444	1.0455	0.0297
MAPK11	gene-expression	7.0284	7.0218	0.0309
MAPK14	gene-expression	6.2730	6.2801	0.0310
ADCY7	methy-Distal-Promoter	3.3485	3.3182	0.0356
MKNK2	methy-Core-Promoter	3.6006	3.6012	0.0367
PRKACB	gene-expression	8.3400	8.3412	0.0394
CAMK2D	gene-expression	3.9452	3.9068	0.0402
MAP2K3	gene-expression	2.5100	2.5054	0.0405
MAPK11	methy-Downstream	7.0284	7.0218	0.0407
ADCY8	methy-Core-Promoter	3.3813	3.3502	0.0446
CAMK2G	cnv_mcnv	3.7893	3.7767	0.0458
CAMK2B	methy-Distal-Promoter	3.8580	3.8406	0.0464
PLD2	methy-Downstream	2.2333	2.2309	0.0466
ADCY7	gene-expression	3.3485	3.3182	0.0532
ADCY8	methy-Distal-Promoter	3.3813	3.3502	0.0537
GNB1	methy-Upstream	6.1358	5.7645	0.0552
MYLK3	gene-expression	3.3555	3.3488	0.0573
PPP3CB	gene-expression	5.8883	5.8699	0.0628
EIF4EBP1	gene-expression	1.0427	1.0428	0.0679
NFATC4	methy-Distal-Promoter	3.5428	3.5624	0.0684
RPS6KA6	gene-expression	0.5532	0.5532	0.0686
MAP2K2	gene-expression	10.8352	10.8435	0.0721
GNB4	gene-expression	3.8911	3.8926	0.0723
RXFP3	methy-Downstream	0.4012	0.3884	0.0730
NFATC4	methy-Downstream	3.5428	3.5624	0.0789
MYLK4	gene-expression	3.3565	3.3497	0.0795
GNGT1	methy-Downstream	3.9190	3.9216	0.0801
GNG13	gene-expression	3.9151	3.9172	0.0808
MAPK10	gene-expression	20.8018	20.8223	0.0858
GNB5	methy-Downstream	3.9168	3.9197	0.0909
GNG13	methy-Downstream	3.9151	3.9172	0.0963

Table 4 (continued)

Gene	Feature	AD-Associated Gene Weight	non-AD-Associated Gene Weight	P-Value
PPP3R1	methy-Upstream	5.4934	5.4964	0.0986
MAP3K1	gene-expression	4.4020	4.4001	0.0988
CAMK2D	methy-Downstream	3.9452	3.9068	0.1028
MKNK2	methy-Distal-Promoter	3.6006	3.6012	0.1134
NFATC3	methy-Distal-Promoter	6.8499	6.8766	0.1181
SPHK2	methy-Distal-Promoter	1.9344	1.9383	0.1195
MAPK13	gene-expression	6.2420	6.2468	0.1281
RPS6KA3	gene-expression	1.0266	1.0277	0.1289
EIF4E2	gene-expression	1.0708	1.0703	0.1315
MAP2K6	methy-Downstream	1.0070	1.0075	0.1346
EIF4EBP1	methy-Core-Promoter	1.0427	1.0428	0.1389
MKNK2	gene-expression	3.6006	3.6012	0.1430
MAPKAPK3	methy-Proximal-Promoter	2.1429	2.1427	0.1531
GNB4	methy-Downstream	3.8911	3.8926	0.1586
MAPK10	methy-Downstream	20.8018	20.8223	0.1593
MAP2K3	methy-Downstream	2.5100	2.5054	0.1607
MAP2K6	methy-Distal-Promoter	1.0070	1.0075	0.1627

underscore the inherent complexity of AD, revealing a multifaceted network of dysregulated pathways that collectively drive disease progression. Addressing these disruptions through precise, targeted therapeutic strategies is critical to mitigating neuronal damage, slowing cognitive decline, and preserving brain function.

The Apelin signaling pathway regulates apoptosis, autophagy, synaptic plasticity, and neuroinflammation. Apelin-13, a key member of the apelin family, has significant neuroprotective functions that help prevent AD by modulating these cellular processes. The Apelin/APJ system influences several signaling pathways, such as PI3K/Akt, MAPK, and PKA, which are essential for cell proliferation and protection from excitotoxicity [34]. Alterations in apelin expression are linked to inflammatory responses, oxidative stress, Ca²⁺ signaling, and apoptosis, all related to AD pathology [35]. The intersection of the Apelin signaling pathway with the WNT signaling pathway suggests a broader regulatory network influencing AD-associated pathologies [36]. The apelinergic system's involvement in brain physiology, including its protective effects against neurological disorders, highlights its importance in maintaining cognitive function and preventing neurodegeneration [37]. The oxytocin (Oxt) signaling pathway influences social behavior, neuroinflammation, and cognitive function. Oxt administration has been shown to reverse learning and memory impairments in AD models, suggesting its potential as a therapeutic target [38]. Key mechanisms include inhibiting microglial activation and reducing inflammatory cytokine levels by blocking the ERK/p38 MAPK and COX-2/iNOS NF- κ B pathways, which prevents cognitive impairment and delays hippocampal atrophy [39]. Oxt also reduces brain inflammation and

corrects memory deficits by promoting A β deposition in dense core plaques, offering neuroprotective effects [40]. Chronic intranasal Oxt administration restores cognitive functions, reduces acetylcholinesterase activity, and lowers levels of β -amyloid and Tau proteins [41]. These effects are supported by decreased hippocampal ERK1/2 and GSK3 β levels, reduced neuronal death, low caspase-3 activity, and improved histopathological profiles, highlighting Oxt's potential in modulating AD pathology [41].

Long-term Potentiation (LTP) is essential for synaptic strength and memory formation [42]. In AD, A β accumulation disrupts LTP by interfering with calcium signaling and receptor function, impairing synaptic plasticity and contributing to cognitive decline [43]. This pathway's deterioration reflects the challenges in memory retention seen in AD, making it a critical focus for therapies aimed at bolstering synaptic function and mitigating memory loss. Oocyte Meiosis contributes to cellular health through mechanisms regulating cell cycle progression and chromosomal stability. Although traditionally associated with reproductive biology, recent studies suggest that dysregulated meiosis-related pathways may contribute to neurodegenerative diseases, including AD [44]. In AD, disruptions in meiosis-like pathways lead to abnormal cell cycle re-entry in neurons, a process that contributes to tau phosphorylation and A β accumulation. By targeting proteins involved in oocyte meiosis pathways, therapeutic approaches could help control aberrant cell cycle activities, potentially reducing neuronal stress and protecting against AD-related cellular damage [45].

The development and progression of AD are also intricately linked to dysfunctions in neurotransmitter systems, including glutamatergic, cholinergic, GABAergic, and dopaminergic synapses. Glutamatergic synapses,

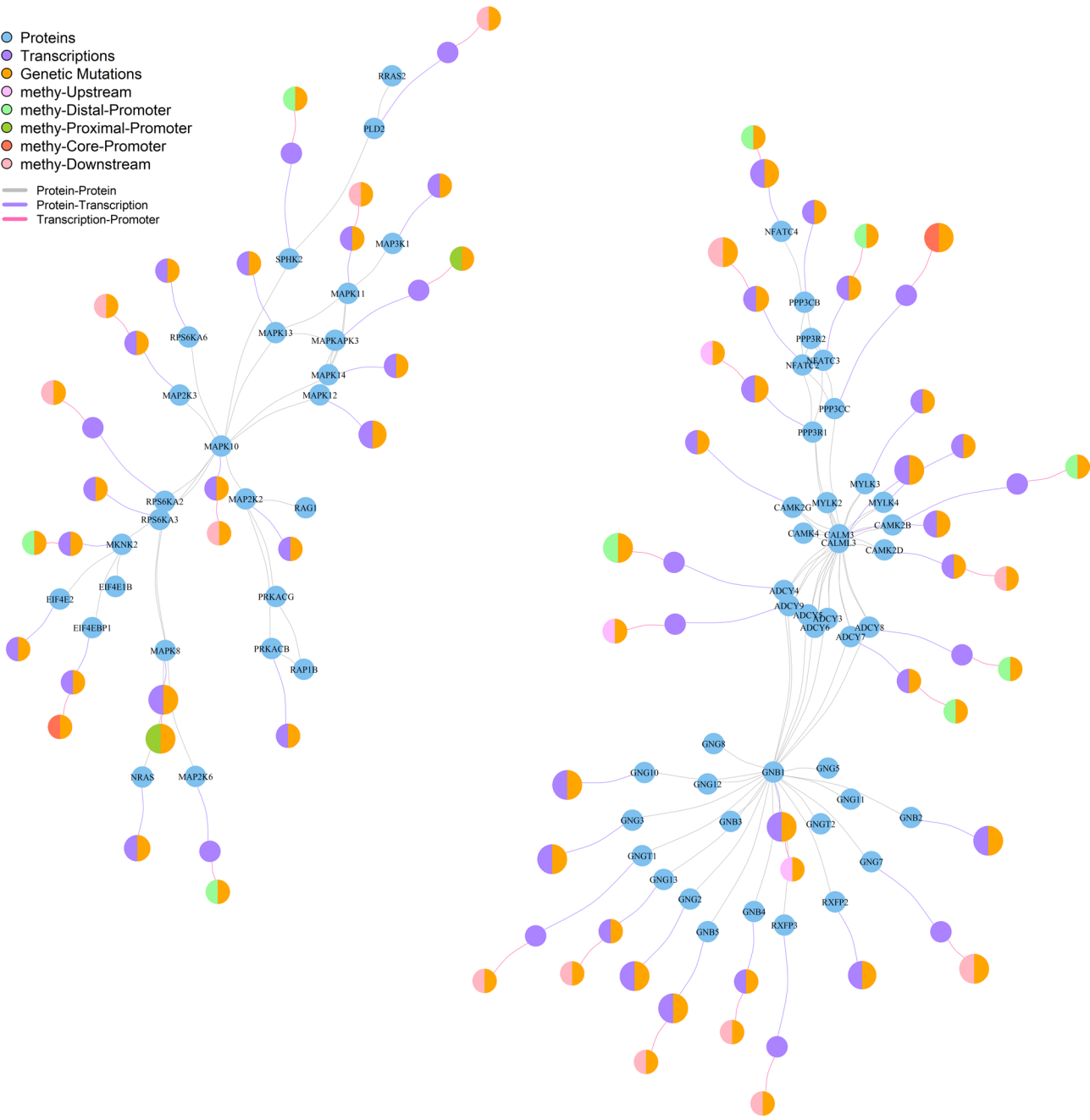


Fig. 3 Top 70 important nodes signaling network interaction in non-AD samples. Blue nodes represent these 70 important genes and the nodes linked to those blue nodes are significant (p-values smaller than 0.2) node features. The 1st hop Linked nodes are genes with significant genetic mutation features or transcriptomic features. And the 2nd hop linked nodes are genes with significant methylation features. Different sizes of the nodes represent the varying importance of the gene features, with larger nodes indicating greater significance based on their p-values

essential for cognitive and behavioral functions, are significantly affected in AD. Dysregulated glutamatergic mechanisms contribute to cognitive impairments and disease progression through interactions with neuronal hyperactivity, A β , tau, and glial dysfunction [46]. A β disrupts glutamate receptors like NMDA and AMPA, leading to calcium dyshomeostasis and impaired synaptic plasticity, characterized by suppressed long-term

potentiation and enhanced long-term depression [47]. Additionally, altered glucose metabolism affects glutamate levels, exacerbating synaptic dysfunction in AD [48]. Cholinergic synapses are also critically involved, with cholinergic atrophy accelerating cognitive decline. The cholinergic hypothesis posits that deficits in cholinergic signaling lead to abnormal tau phosphorylation, neuroinflammation, and cell apoptosis [49]. The basal

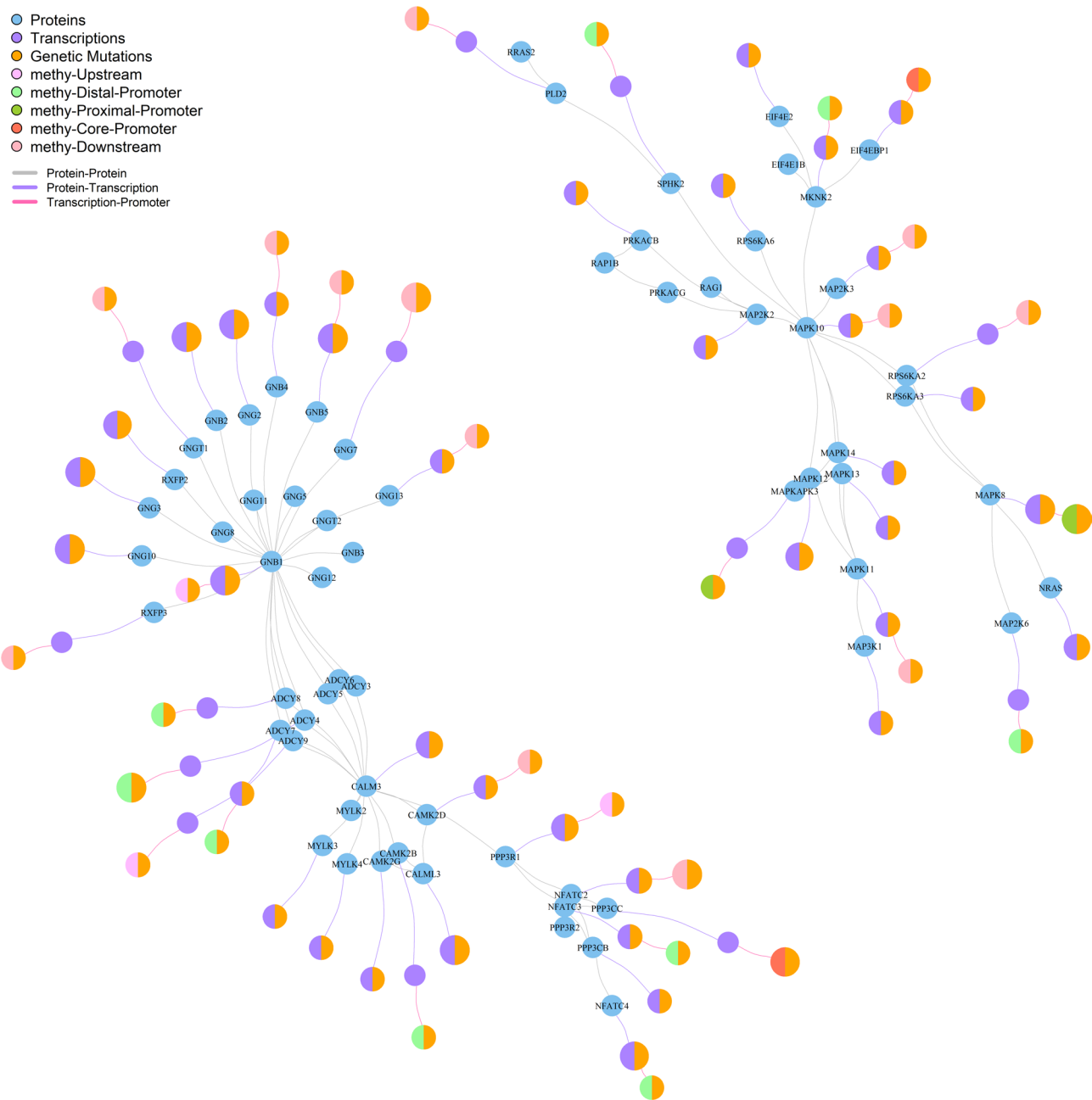


Fig. 4 Top 70 important nodes signaling network interaction in non-AD samples. Blue nodes represent these 70 important genes and the nodes linked to those blue nodes are significant (p-values smaller than 0.2) node features. The 1 st hop Linked nodes are genes with significant genetic mutation features or transcriptomic features. And the 2nd hop linked nodes are genes with significant methylation features. Different sizes of the nodes represent the varying importance of the gene features, with larger nodes indicating greater significance based on their p-values

forebrain cholinergic innervation of cortical areas is particularly vulnerable, and cholinergic receptor regulation is a hallmark of AD progression [50]. GABAergic synapses, responsible for inhibitory signaling, are disrupted in AD due to alterations in the GABA receptor system and perineuronal nets, leading to synaptic hyperactivity and abnormal brain oscillations, contributing to cognitive deficits [51]. Although less studied, dopaminergic synapses also play a role in AD, with D2 dopaminergic

receptors implicated in symptomatology [50]. Synaptic dysfunction is a common pathogenic trait in AD, with synapse loss closely correlating with cognitive decline. The interplay between A β and tau at the synapse exacerbates synaptic deficits, making targeting these dysfunctions crucial for developing therapeutic strategies. Aberrant neurotransmission, including cholinergic, adrenergic, and glutamatergic networks, underpins cognitive decline in AD, with NMDAR dysfunction being

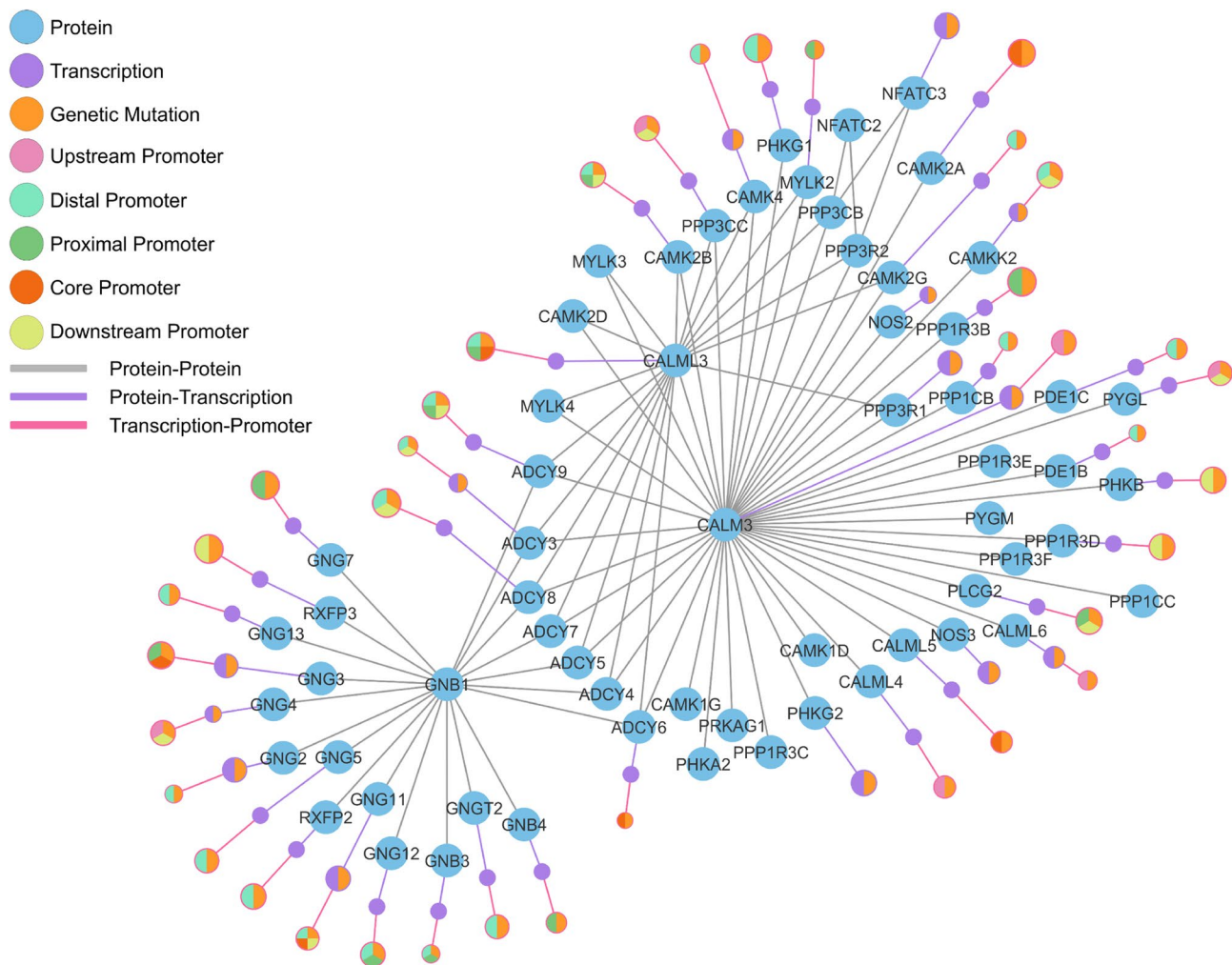


Fig. 5 Top 70 important nodes signaling network interaction in female samples. Blue nodes represent these 70 important genes and the nodes linked to those blue nodes are significant (p-values smaller than 0.2) node features. The 1st hop Linked nodes are genes with significant genetic mutation features or transcriptomic features. And the 2nd hop linked nodes are genes with significant methylation features. Different sizes of the nodes represent the varying importance of the gene features, with larger nodes indicating greater significance based on their p-values

particularly significant. Together, these neurotransmitter system alterations highlight the complexity of AD pathogenesis and the need for targeted interventions to mitigate synaptic dysfunction and cognitive decline.

Research indicates that HCMV may contribute to poorer cognitive abilities and augment tauopathy by interacting with TRA CDR3 and tau peptides [52]. Additionally, HCMV, along with other herpesviruses, has been shown to impact AD-related processes such as A β formation, neuronal death, and autophagy through virus-host protein-protein interactions [53]. Persistent HCMV infections can lead to the generation of AD hallmarks, including A β plaques and NFTs composed of hyperphosphorylated tau proteins, by exploiting pathways involved in oxidative stress and neuroinflammation [54]. Kaposi Sarcoma-associated Herpesvirus (KSHV), a member of the Herpesviridae family, is known to impact AD-related

processes such as A β formation, neuronal death, and autophagy, which are critical in the pathogenesis of AD [53]. The “infectious hypothesis” of AD suggests that pathogens, including viruses like KSHV, may act as seeds for A β aggregation, leading to plaque formation and cognitive decline [55]. Viral infections, including those caused by herpesviruses, can trigger neuroinflammatory pathways, disrupt the BBB, and activate microglia, leading to neural cell death and neurodegeneration [56]. Specifically, KSHV, along with other herpesviruses, has been shown to influence processes crucial for cellular homeostasis and dysfunction, potentially exacerbating AD pathology through virus-host protein-protein interactions [53]. Additionally, the reactivation of herpesviruses during acute infections, such as SARS-CoV-2, can create a synergistic pathogenic effect, further promoting

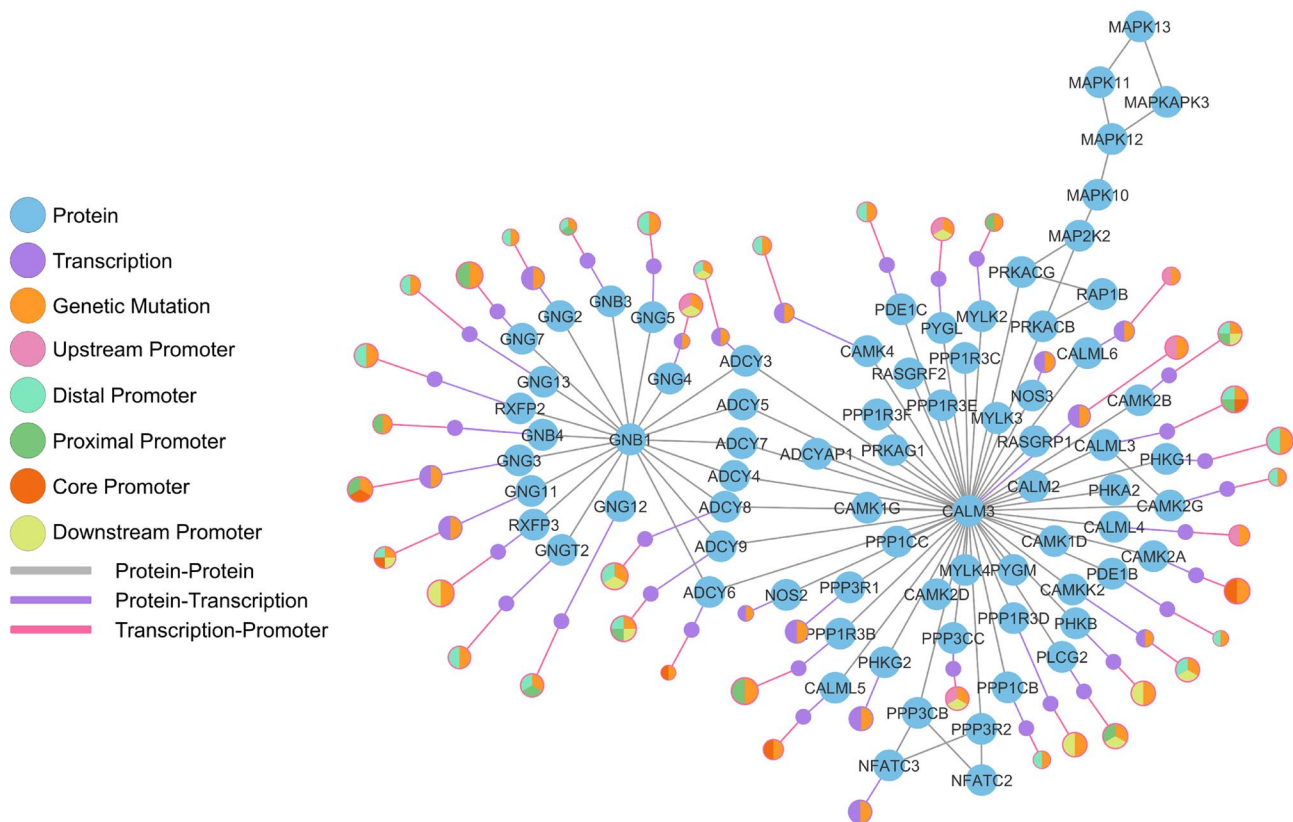


Fig. 6 Top 70 important nodes signaling network interaction in male samples. Blue nodes represent these 70 important genes and the nodes linked to those blue nodes are significant (p-values smaller than 0.2) node features. The 1st hop Linked nodes are genes with significant genetic mutation features or transcriptomic features. And the 2nd hop linked nodes are genes with significant methylation features. Different sizes of the nodes represent the varying importance of the gene features, with larger nodes indicating greater significance based on their p-values

neurodegenerative processes like A β formation and oxidative stress response [53].

Inflammatory mediators significantly regulate TRP channels, which particularly TRPV1 and TRPC6, are involved in neuroinflammation and oxidative stress responses [57, 58]. TRPV1 modulates neuroinflammation by influencing the production of inflammatory mediators and oxidative stress responses [58]. Its activation can rescue microglial dysfunction and restore immune responses, including phagocytic activity and autophagy, through the AKT/mTOR pathway, reducing amyloid pathology and reversing memory deficits in AD models [57, 59]. The regulation of TRP channels by inflammatory mediators also helps maintain the integrity of the BBB and neurovascular coupling, both compromised in AD. TRP channels are activated by reactive oxygen species, linking oxidative stress to neurodegenerative disease progression. This interplay highlights the role of TRP channels in modulating neuroinflammation and oxidative stress, offering promising avenues for AD treatment.

Lipid and Atherosclerosis pathway is critically involved in AD progression due to its impact on vascular health and hemostatic balance [60]. Dysregulated lipid

metabolism and atherosclerosis contribute to blood-brain barrier dysfunction, enabling immune cell infiltration and intensifying neuroinflammation [61]. These vascular changes also exacerbate amyloid-beta (A β) deposition and tau pathology, key markers of AD [62]. Oxidative stress associated with lipid dysregulation accelerates vascular damage, contributing to neurodegeneration. Targeting the lipid and atherosclerosis pathway could therefore help restore vascular health and reduce AD progression driven by hemostatic imbalances [63].

Discussion and conclusion

This study presents a significant advancement in analyzing complex multi-omic data for AD research. Integrating multiple omics layers, i.e., epigenomics, genomics, transcriptomics and proteomics, provides a holistic view of disease-related molecular mechanisms. By embedding these layers into a structured signaling graph, the model achieves substantial improvements in predictive accuracy and biomarker identification, highlighting critical genes and pathways associated with AD pathogenesis. A key strength of our approach lies in its graph-based structure, which captures biological interactions by incorporating

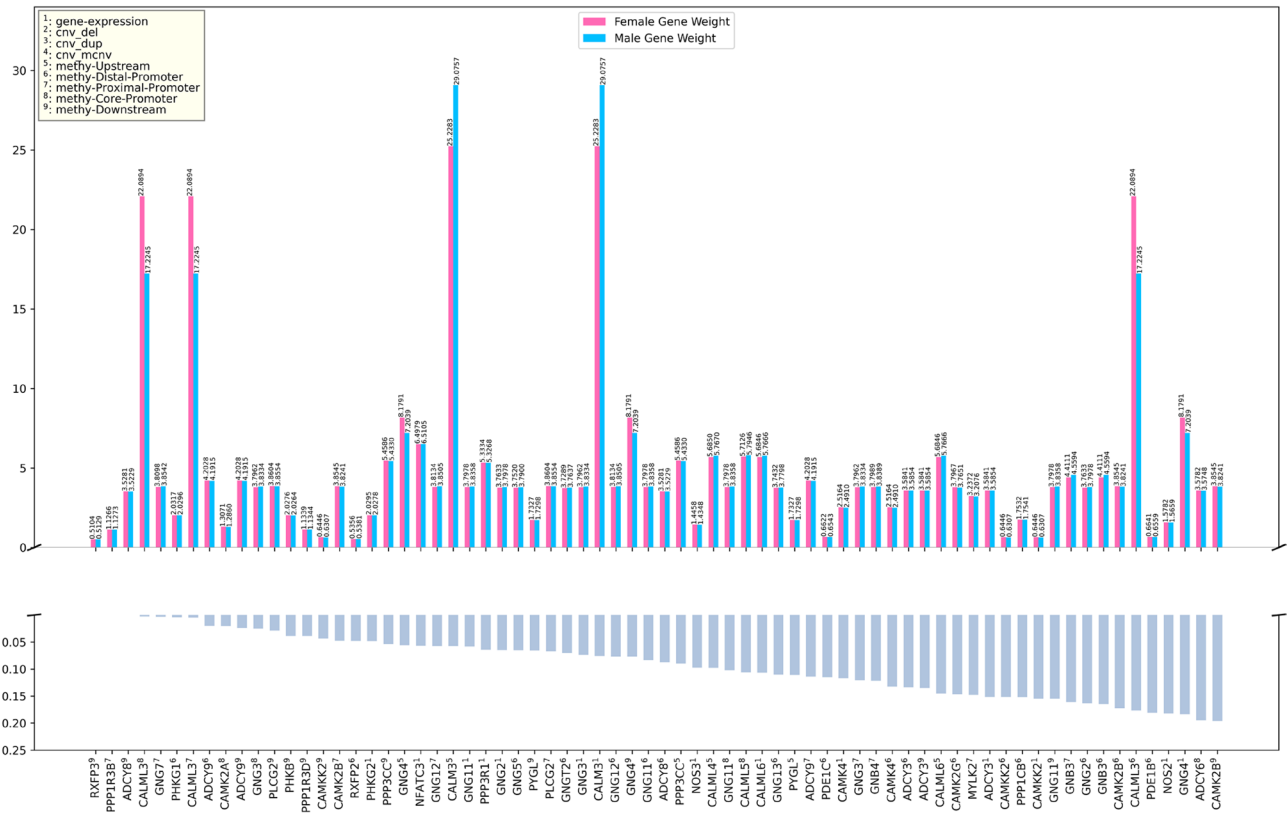


Fig. 7 Bar chart displaying gene weights (importance scores) in female and male from AD samples, ranking by their p-values

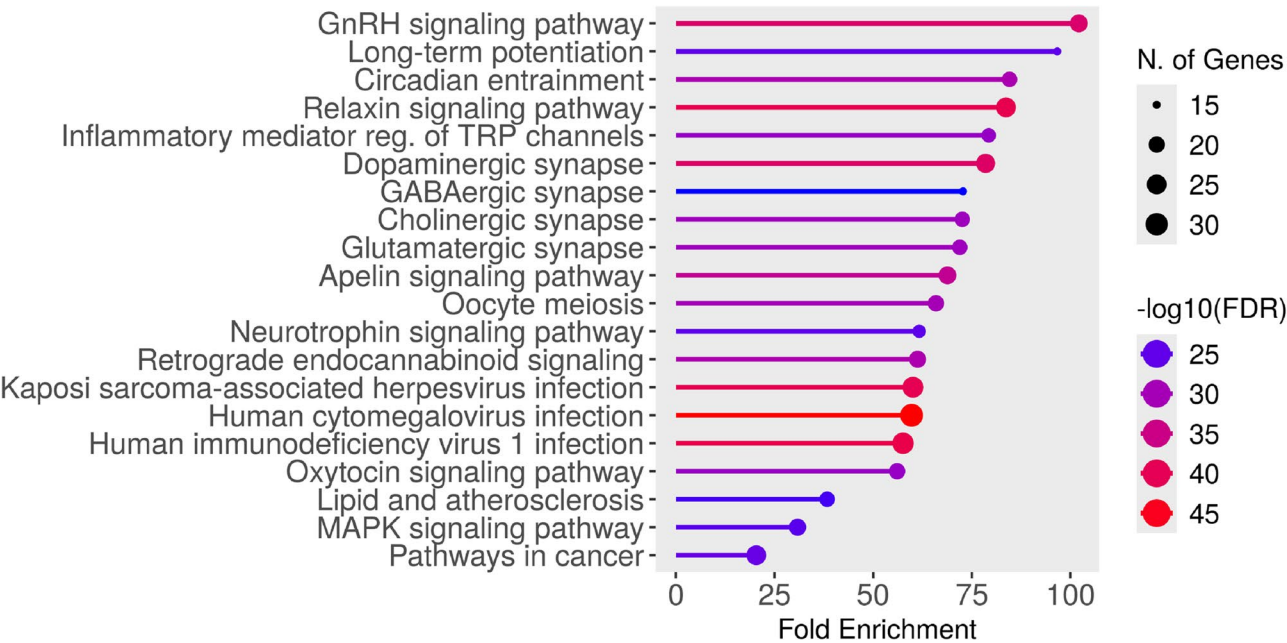


Fig. 8 Lollipop plot showing the negative base-10 logarithm of the False Discovery Rate (FDR) and number of genes of the top 20 signaling pathways based on the top 70 gene features found to be associated with AD. Generated by ShinyGo 0.80 after performing pathway enrichment analysis with FDR cutoff at 0.05

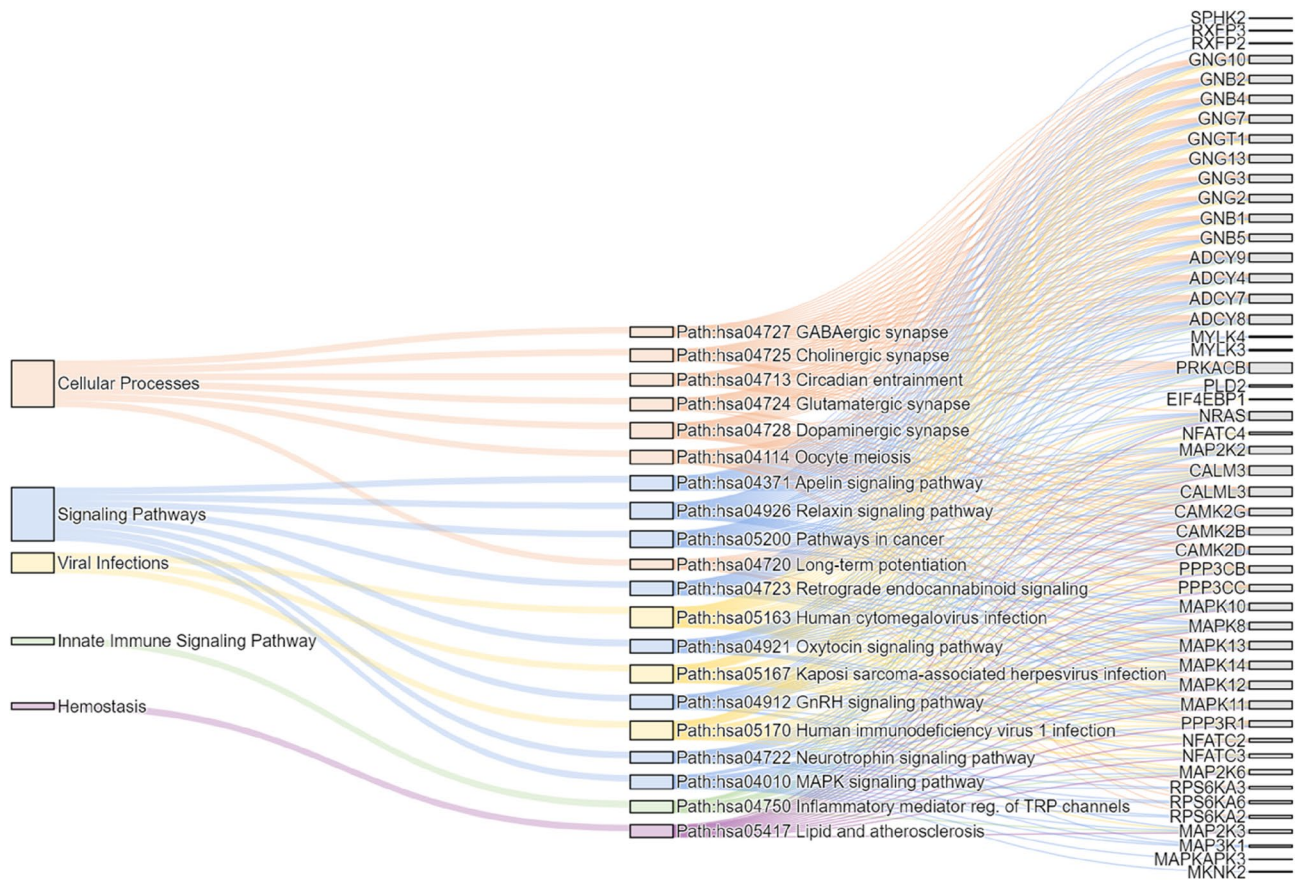


Fig. 9 Sankey diagram illustrating the relationship between the identified signaling pathways and corresponding genes using the top 70 genes features found to be associated with AD

KEGG pathways, enabling the model to leverage the complex interconnections between genes and proteins, often disregarded by traditional approaches. The inclusion of multi-hop message propagation and attention-based mechanisms allows for effective identification of crucial molecular interactions at various omic levels, providing interpretable results that align well with established AD research. Additionally, the model demonstrates robustness in both AD versus non-AD classification and gender-based subgroup classification within the AD samples, suggesting potential gender-specific molecular differences in AD pathology. Pathway enrichment and inferred signaling networks further validated the model by identifying pathways known to play roles in neurodegeneration, inflammation, and synaptic dysfunction in AD. Notably, pathways such as Apelin signaling, oxytocin signaling, and cellular processes like long-term potentiation (LTP) and oocyte meiosis were recurrently highlighted, indicating their significance in neuroinflammation and synaptic impairment within AD. The findings also resonate with observed dysfunctions in neurotransmitter systems (glutamatergic, cholinergic, and GABAergic), critical components in AD pathogenesis. In addition

to the genetic biomarkers, like APOE4, sex is another factor of AD. For example, the AD risk in women is 2 times higher than in men, with more severe AD phenotypes and faster memory decline. Therefore, it is important to study the pathogenesis of sex in AD. However, the pathogenesis and sex factors of AD are complex, and it is not a trivial task to integrate the biomarkers identified by comparing AD vs. control and Female vs. Male. We will systematically investigate the pathogenesis and sex factors of AD in the future work.

In conclusion, this model demonstrates the potential of graph AI for multi-omic data integration and biomarker discovery in AD. By capturing multi-level molecular interactions, it provides a more comprehensive understanding of AD pathogenesis, identifying not only potential biomarkers but also pathways that could serve as therapeutic targets. Its performance, surpassing conventional GNNs, highlights its value in handling the complexities of multi-omic data. This study affirms the potential of multi-omic integration for translational applications in AD research and lays a foundation for future investigations to refine our understanding of disease mechanisms and intervention strategies. Given its

flexibility, the approach can also be adapted for other diseases, making it a valuable tool in precision medicine efforts aimed at identifying disease-specific targets. Future work may focus on expanding its application across diseases and enhancing interpretability by incorporating additional biological context, such as gene-environment interactions and regulatory epigenetic modifications. Moreover, the data quality and diversity might affect the model generalization capability. With more and more multi-omic datasets have been being generated, the proposed new model framework for multi-omic data analysis can be further evaluated and improved in the future work.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s44330-025-00041-8>.

Supplementary Material 1.

Acknowledgements

The results published here are in whole or in part based on data obtained from the AD Knowledge Portal.

Authors' contributions

HZ, DC, TX, EC, GL, YC, PP, MP, FL designed the analysis methodology. HZ, DC, TX, EC, FL conducted the data analysis and drafted the manuscript. All authors read and approved the final manuscript.

Funding

This study was partially supported by NIA R56AG065352, 1R21AG078799-01A1, NINDS 1R01NS132962-01.

Data availability

The datasets are available via the AD Knowledge Portal (<https://adknowledgeportal.org>). The AD Knowledge Portal is a platform for accessing data, analyses, and tools generated by the Accelerating Medicines Partnership (AMP-AD) Target Discovery Program and other National Institute on Aging (NIA)-supported programs to enable open-science practices and accelerate translational learning. The data, analyses and tools are shared early in the research cycle without a publication embargo on secondary use. Data is available for general research use according to the following requirements for data access and data attribution (<https://adknowledgeportal.synapse.org/Data%20Access>). For access to content described in this manuscript see the synapse dataset IDs in Table S1 for details.

Declarations

Ethics approval and consent to participate

Not applicable. This is a second use study of the AD omic datasets. The datasets are available via the AD Knowledge Portal (<https://adknowledgeportal.org>). For the second use of the AD datasets, we submitted The Intended Data Use statement (IDU) and Data Use Certificate (DUC) to AD knowledge data portal and were approved by AD knowledge data portal for the data access and use. The second use was also approved by Washington university school of medicine IRB (Exempt) no: 202506188.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 27 August 2024 / Accepted: 23 July 2025

Published online: 06 October 2025

References

1. Zeliger H. Oxidative Stress: Its Mechanisms and Impacts on Human Health and Disease Onset. Academic Press; 2022:229–238. Chapter 19, Aging mechanism.
2. *MAPPING A BETTER FUTURE FOR DEMENTIA CARE NAVIGATION.*
3. Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. In: 5th International Conference on Learning Representations (ICLR); April 24–26, 2017; Toulon, France. Published online 2017:1–14.
4. Hamilton WL, Ying R, Leskovec J. Inductive representation learning on large graphs. *Adv Neural Inf Process Syst.* 2017;2017:1025–35.
5. Veličković P, Casanova A, Liò P, Cucurull G, Romero A, Bengio Y. Graph attention networks. In: 6th International Conference on Learning Representations (ICLR); April 30–May 3, 2018; Vancouver, BC, Canada. Published online 2018:1–12. <http://arxiv.org/abs/1710.10903>.
6. Xu K, Hu W, Leskovec J, Jegelka S. *HOW POWERFUL ARE GRAPH NEURAL NETWORKS?*
7. Wang T, Shao W, Huang Z, et al. MOGONET integrates multi-omics data using graph convolutional networks allowing patient classification and biomarker identification. *Nat Commun.* 2021;12(1): 3445.
8. Li X, Ma J, Leng L, et al. MoGCN: a multi-omics integration method based on graph convolutional network for cancer subtype analysis. *Front Genet.* 2022;13: 806842.
9. Gao H, Zhang B, Liu L, Li S, Gao X, Yu B. A universal framework for single-cell multi-omics data integration with graph convolutional networks. *Brief Bioinform.* 2023;24(3): bbad081.
10. Rajadhyaksha N, Chitkara A. Graph contrastive learning for multi-omics data. *arXiv.* Published online January 5, 2023. <https://doi.org/10.48550/arXiv.2301.02242>.
11. Li G, Muller M, Thabet A, Ghanem B. DeepGCNs: can GCNs go as deep as cnns?? *Proc IEEE Int Conf Comput Vis.* 2019;2019–Octob:9266–75. <https://doi.org/10.1109/ICCV.2019.00936>.
12. Cai C, Wang Y. A note on over-smoothing for graph neural networks. Published online June 23, 2020. <http://arxiv.org/abs/2006.13318>. Accessed 6 Aug 2025.
13. Abu-El-Hajja S, Perozzi B, Kapoor A, Al-Rfou R, Lee J. MixHop: Higher-order graph convolutional architectures via sparsified neighborhood mixing. In: Chaudhuri K, Salakhutdinov R, eds. *Proceedings of the 36th International Conference on Machine Learning (ICML 2019)*. PMLR. 2019;97:21–9.
14. Morris C, Ritzert M, Fey M, et al. *Weisfeiler and Leman go neural: higher-order graph neural networks.* *Proceedings of the AAAI Conference on Artificial Intelligence.* 2019. <https://doi.org/10.1609/aaai.v33i01.33014602>.
15. Chien E, Peng J, Li P, Milenkovic O. Adaptive universal generalized PageRank graph neural network. *arXiv.* <https://doi.org/10.48550/arXiv.2006.07988>. Published online June 13, 2020.
16. Nikolentzos G, Dasoulas G, Vazirgiannis M. K-hop graph neural networks. *Neural Netw.* 2020;130:195–205.
17. Brossard R, Frigo O, Dehaene D. Graph convolutions that can finally model local structure. *arXiv.* <https://doi.org/10.48550/arXiv.2011.15069>. Published online November 29, 2020.
18. Wang G, Ying R, Huang J, Leskovec J. Multi-hop attention graph neural network. *arXiv.* <https://doi.org/10.48550/arXiv.2009.14332>. Published online September 30, 2020.
19. Feng J, Chen Y, Li F, Sarkar A, Zhang M. *How Powerful Are K-Hop Message Passing Graph Neural Networks.*
20. Zhang H, Cao D, Chen Z et al. MosGraphGen: a novel tool to generate multi-omic signaling graphs to facilitate integrative and interpretable graph AI model development. <https://doi.org/10.1101/2024.05.15.594360>
21. Shi Y, Huang Z, Feng S, Zhong H, Wang W, Sun Y. Masked label prediction: unified message passing model for Semi-Supervised classification. Published online September 8, 2020. <http://arxiv.org/abs/2009.03509>
22. Zhang H, Goedegebuure SP, Ding L et al. M3NetFlow: a novel multi-scale multi-hop multi-omics graph AI model for omics data integration and interpretation. <https://doi.org/10.1101/2023.06.15.545130>
23. Zhang H, Chen Y, Payne P, Li F. Mining signaling flow to interpret mechanisms of synergy of drug combinations using deep graph neural networks. <https://doi.org/10.1101/2021.03.25.437003>

24. Zhang H, Chen Y, Payne P, Li F. Using deepsignalingflow to mine signaling flows interpreting mechanism of synergy of cocktails. *NPJ Syst Biol Appl*. 2024;10(1): 92.
25. 2022 Alzheimer's disease facts and figures. *Alzheimer's & Dementia*. 2022;18(4):700–789. <https://doi.org/10.1002/alz.12638>
26. Mauvais-Jarvis F, Bairey Merz N, Barnes PJ, et al. Sex and gender: modifiers of health, disease, and medicine. *Lancet*. 2020;396(10250):565–82. [https://doi.org/10.1016/S0140-6736\(20\)31561-0](https://doi.org/10.1016/S0140-6736(20)31561-0).
27. Li F, Eteleeb AM, Buchser W, et al. Weakly activated core neuroinflammation pathways were identified as a central signaling mechanism contributing to the chronic neurodegeneration in Alzheimer's disease. *Front Aging Neurosci*. 2022. <https://doi.org/10.3389/fnagi.2022.935279>.
28. Li F, Oh I, Kumar S, et al. Loss of estrogen unleashing neuro-inflammation increases the risk of Alzheimer's disease in women. *bioRxiv*. Published online January 1, 2022:2022.09.19.508592.
29. Verheijen J, Slegers K. *Understanding Alzheimer Disease at the Interface between Genetics and Transcriptomics*. Vol 34; 2018. <https://doi.org/10.1016/j.tig.2018.02.007>
30. Riedel BC, Thompson PM, Brinton RD, Age, APOE and sex: triad of risk of alzheimer's disease. *J Steroid Biochem Mol Biol*. 2016;160:134–47. <https://doi.org/10.1016/j.jsbmb.2016.03.012>.
31. Del-Aguila J, Fernández Hernández MV, Schindler S, et al. Assessment of the genetic architecture of alzheimer's disease risk in rate of memory decline. *J Alzheimer's Disease*. 2018;62:745–56. <https://doi.org/10.3233/JAD-170834>.
32. Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. *arXiv*. <https://doi.org/10.48550/arXiv.1609.02907>. Published online September 9, 2016.
33. Shi Y, Huang Z, Feng S, Zhong H, Wang W, Sun Y. *Masked Label Prediction: Unified Message Passing Model for Semi-Supervised Classification*.
34. Kinjo T, Higashi H, Uno K, Kuramoto N. Apelin/Apelin receptor system: molecular characteristics, physiological roles, and prospects as a target for disease prevention and pharmacotherapy. *Curr Mol Pharmacol*. 2021;14(2):210–9. <https://doi.org/10.2174/1874467213666200602133032>.
35. Luo H, Han L, Xu J. Apelin/APJ system: a novel promising target for neurodegenerative diseases. *J Cell Physiol*. 2020;235(2):638–57. <https://doi.org/10.1002/jcp.29001>.
36. Kostas WW, Brafman DA. The multifaceted role of WNT signaling in Alzheimer's disease onset and age-related progression. *Cells*. 2023. <https://doi.org/10.3390/cells12081204>.
37. Ivanov MN, Stoyanov DS, Pavlov SP, Tonchev AB. Distribution, function, and expression of the apelinergic system in the healthy and diseased mammalian brain. *Genes*. 2022;13(11): 2172. <https://doi.org/10.3390/genes13112172>.
38. Takahashi J, Yamada D, Nagano W, Saitoh A. The role of oxytocin in Alzheimer's disease and its relationship with social interaction. *Cells*. 2023. <https://doi.org/10.3390/cells12202426>.
39. Ye C, Cheng M, Ma L, et al. Oxytocin nanogels inhibit innate inflammatory response for early intervention in Alzheimer's disease. *ACS Appl Mater Interfaces*. 2022;14(19):21822–35. <https://doi.org/10.1021/acsami.2c00007>.
40. Clara Selles M, Fortuna JT, de Faria YP et al. Oxytocin attenuates microglial activation and restores social and non-social memory in the APP/PS1 mouse model of alzheimer's disease. <https://doi.org/10.1101/2022.05.07.491031>
41. El-Ganainy SO, Soliman OA, Ghazy AA, et al. Intranasal oxytocin attenuates cognitive impairment, β -amyloid burden and tau deposition in female rats with Alzheimer's disease: interplay of ERK1/2/GSK3 β /Caspase-3. *Neurochem Res*. 2022;47(8):2345–56. <https://doi.org/10.1007/s11064-022-03624-x>.
42. Bliss TVP, Collingridge GL. A synaptic model of memory: long-term potentiation in the hippocampus. *Nature*. 1993. <https://doi.org/10.1038/361031a0>.
43. Palop JJ, Mucke L. Amyloid- β -induced neuronal dysfunction in alzheimer's disease: from synapses toward neural networks. *Nat Neurosci*. 2010;13(7):812–8. <https://doi.org/10.1038/nn.2583>.
44. Koseoglu MM, Norambuena A, Sharlow ER, Lazo JS, Bloom GS. Aberrant neuronal cell cycle re-entry: the pathological confluence of Alzheimer's disease and brain insulin resistance, and its relation to cancer. *J Alzheimers Dis*. 2019;67(1):1–11. <https://doi.org/10.3233/JAD-180874>.
45. Martin L, Latypova X, Wilson CM, et al. Tau protein kinases: involvement in alzheimer's disease. *Ageing Res Rev*. 2013;12(1):289–309. <https://doi.org/10.1016/j.arr.2012.06.003>.
46. Pinky PD, Pfitzer JC, Senfeld J, et al. Recent insights on glutamatergic dysfunction in Alzheimer's disease and therapeutic implications. *Neuroscientist*. 2023;29(4):461–71. <https://doi.org/10.1177/10738584211069897>.
47. Zhang H, Jiang X, Ma L, et al. Role of A β in Alzheimer's-related synaptic dysfunction. *Front Cell Dev Biol*. 2022;10. <https://doi.org/10.3389/fcell.2022.964075>.
48. Bukke VN, Archana M, Villani R, et al. The dual role of glutamatergic neurotransmission in alzheimer's disease: from pathophysiology to pharmacotherapy. *Int J Mol Sci*. 2020;21(20):1–29. <https://doi.org/10.3390/ijms21207452>.
49. Chen ZR, Huang JB, Yang SL, Hong FF. Role of cholinergic signaling in Alzheimer's disease. *Molecules*. 2022. <https://doi.org/10.3390/molecules27061816>.
50. Lombardero L, Llorente-Ovejero A, Manuel I, Rodríguez-Puertas R. Chapter 28 - Neurotransmitter receptors in Alzheimer's disease: from glutamatergic to cholinergic receptors. In: Martin CR, Preedy VR, editors. *Genetics, Neurology, Behavior, and Diet in Dementia*. Academic; 2020. p. 441–56. <https://doi.org/10.1016/B978-0-12-815868-5.00028-1>.
51. Ali AB, Islam A, Constanti A. The fate of interneurons, GABAA receptor subtypes and perineuronal nets in Alzheimer's disease. *Brain Pathol*. 2023. <https://doi.org/10.1111/bpa.13129>.
52. Blanck G, Huda TI, Chobrutskiy BI, Chobrutskiy A. CMV as a factor in the development of Alzheimer's disease? *Med Hypotheses*. 2023. <https://doi.org/10.1016/j.mehy.2023.111140>.
53. Onisiforou A, Zanos P. From viral infections to alzheimer's disease: unveiling the mechanistic links through systems bioinformatics. <https://doi.org/10.1101/2023.12.05.570187>
54. Athanasios E, Gargalionis AN, Anastassopoulou C, Tsakris A, Boufidou F. New Insights into the Molecular Interplay between Human Herpesviruses and Alzheimer's Disease—A Narrative Review. *Brain Sci*. 2022;12(8). <https://doi.org/10.3390/brainsci12081010>.
55. Piotrowski SL, Tucker A, Jacobson S. The elusive role of herpesviruses in Alzheimer's disease: current evidence and future directions. *NeuroImmune Pharmacology and Therapeutics*. 2023;2(3):253–66. <https://doi.org/10.1515/nip-2023-0011>.
56. Du C. Virus-induced alzheimer's disease: potential roles of viral infections in AD neuropathogenesis from two aspects: aberrant protein accumulations with neuroinflammatory response and virus-induced ablation of adult neurogenesis. *AIP Conf Proc*. 2022;2511(1):020063. <https://doi.org/10.1063/5.0093889>.
57. Lu R, He Q, Wang J. TRPC channels and alzheimer's disease. *Advances Experimental Med Biology*. Published Online 2017:73–83. https://doi.org/10.1007/978-94-024-1088-4_7
58. Duitama M, Vargas-López V, Casas Z, Albarracín SL, Sutachan JJ, Torres YP. TRP channels role in pain associated with neurodegenerative diseases. *Front Neurosci*. 2020. <https://doi.org/10.3389/fnins.2020.00782>.
59. Lu J, Zhou W, Dou F, Wang C, Yu Z. TRPV1 sustains microglial metabolic reprogramming in alzheimer's disease. *EMBO Rep*. 2021;22(6). <https://doi.org/10.15252/embr.202052013>.
60. Williams DM, Finan C, Schmidt AF, Burgess S, Hingorani AD. Lipid lowering and Alzheimer disease risk: a Mendelian randomization study. *Ann Neurol*. 2020;87(1):30–9. <https://doi.org/10.1002/ana.25642>.
61. Montagne A, Barnes SR, Sweeney MD, et al. Blood-brain barrier breakdown in the aging human hippocampus. *Neuron*. 2015;85(2):296–302. <https://doi.org/10.1016/j.neuron.2014.12.032>.
62. Tao QQ, Lin RR, Chen YH, Wu ZY. Discerning the role of blood brain barrier dysfunction in Alzheimer's disease. *Ageing Dis*. 2022;13(5):1391–404. <https://doi.org/10.14336/AD.2022.0130-1>.
63. Cai Z, Yan LJ. *Protein Oxidative Modifications: Beneficial Roles in Disease and Health*.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.