

Deep Trans-Omic Network Fusion for Molecular Mechanism of Alzheimer's Disease

Linhui Xie^{a,d,1}, Yash Raj^{b,1}, Pradeep Varathan^{b,d}, Bing He^{b,d}, Meichen Yu^{c,d}, Kwangsik Nho^{c,d}, Paul Salama^a, Andrew J. Saykin^{c,d} and Jingwen Yan^{b,d,*}

^a*Department of Electrical and Computer Engineering, Indiana University Purdue University Indianapolis, Indianapolis, IN, USA*

^b*Department of BioHealth Informatics, Indiana University Purdue University Indianapolis, Indianapolis, IN, USA*

^c*Department of Radiology and Imaging Sciences, Indiana University School of Medicine, Indianapolis, IN, USA*

^d*Indiana Alzheimer's Disease Research Center, Indianapolis, IN, USA*

Accepted 21 March 2024

Pre-press 9 May 2024

Abstract.

Background: There are various molecular hypotheses regarding Alzheimer's disease (AD) like amyloid deposition, tau propagation, neuroinflammation, and synaptic dysfunction. However, detailed molecular mechanism underlying AD remains elusive. In addition, genetic contribution of these molecular hypothesis is not yet established despite the high heritability of AD.

Objective: The study aims to enable the discovery of functionally connected multi-omic features through novel integration of multi-omic data and prior functional interactions.

Methods: We propose a new deep learning model MoFNet with improved interpretability to investigate the AD molecular mechanism and its upstream genetic contributors. MoFNet integrates multi-omic data with prior functional interactions between SNPs, genes, and proteins, and for the first time models the dynamic information flow from DNA to RNA and proteins.

Results: When evaluated using the ROS/MAP cohort, MoFNet outperformed other competing methods in prediction performance. It identified SNPs, genes, and proteins with significantly more prior functional interactions, resulting in three multi-omic subnetworks. SNP-gene pairs identified by MoFNet were mostly eQTLs specific to frontal cortex tissue where gene/protein data was collected. These molecular subnetworks are enriched in innate immune system, clearance of misfolded proteins, and neurotransmitter release respectively. We validated most findings in an independent dataset. One multi-omic subnetwork consists exclusively of core members of SNARE complex, a key mediator of synaptic vesicle fusion and neurotransmitter transportation.

Conclusions: Our results suggest that MoFNet is effective in improving classification accuracy and in identifying multi-omic markers for AD with improved interpretability. Multi-omic subnetworks identified by MoFNet provided insights of AD molecular mechanism with improved details.

Keywords: Alzheimer's disease, deep learning, multi-omics, neural network, systems biology

¹These authors contributed equally to this work.

*Correspondence to: Prof. Jingwen Yan, Department of Bio-Health Informatics, Luddy School of Informatics, Computing,

and Engineering, Indiana University Purdue University Indianapolis, Indianapolis, IN, USA. Tel.: +1 317 278 7668; E-mail: jingyan@iupui.edu.

INTRODUCTION

Alzheimer's disease (AD) is an irreversible progressive neurodegenerative disorder that slowly destroys memory and thinking skills, and ultimately the capability of self-care. It is affecting more than 6.5 million Americans, which is expected to double by 2050 [1]. But by far, there is no clinically validated cure for AD; the molecular mechanism underlying its on-set and progression is not yet established. Two well-known pathological hallmarks of AD are the accumulation of intercellular amyloid- β plaques and intracellular neurofibrillary tangles. A widely accepted hypothetical model of AD pathophysiology starts from the early accumulation of amyloid- β plaques, which precedes spreading of tau, neuronal loss and manifestation of other clinical traits by up to 30 years [2]. Given that, clearance of amyloid plaque inside the brain has been the major target for the development of disease modifying therapies. Other molecular pathways disrupted in AD are also frequently reported like neurotransmitter activity, neuroinflammation and synaptic plasticity [3]. To date, drugs targeting these molecular mechanisms have largely failed in clinical trials, highlighting the need to investigate pathways upstream of plaque and tangle formation, or other distinct mechanisms.

Multi-omic data has been increasingly collected at the individual level, providing a wide range of genetic and molecular features, like single nucleotide polymorphism (SNPs), messenger RNAs (mRNAs), and proteins. These data hold complementary information which could be integrated for improved precision in identifying risk genes and revealing molecular mechanism related to complex diseases like AD. They have shown great potential in advancing the understanding of molecular mechanism underlying complex diseases like various cancers [4]. In addition, databases like Reactome and SNP2TFBS provide critical prior knowledge of regulatory links between SNPs, genes, and proteins [5, 6]. It is important to note, however, that these previous interactions are mostly not specific to any particular tissues, and their tissue-specificity remains largely unknown.

Although not perfect, this prior knowledge could still be leveraged as additional source of evidence on top of multi-omics data, to improve the precision in discovery of molecular mechanism [7, 8]. Several recent studies have confirmed that deep learning models informed by prior network of functional interactions can lead to improved prediction performance and model interpretability. For example, knowledge

primed neural network (KPNN) incorporated the signaling pathways to improve the prediction power of RNA-Seq data [9]. Varmole was proposed for integration of genotype and gene expression data with expression quantitative loci (eQTL) relationships embedded into the neural network [10]. But those deep learning models can only take single or two types of -omics data. Another multi-omics study utilizing prior interactions to inform the sparse logistic regression [7] reported that identified SNPs and genes are mostly eQTLs specific to the tissue where gene expression data was collected, which resolves the concern regarding lack of tissue-specificity in the prior network of interactions.

In this study, we introduce an interpretable deep learning method, multi-omic fused neural network (MoFNet), to jointly model multi-omics data and the prior functional interactions among proteins and genes, and their upstream regulatory SNPs. MoFNet is biologically wired with regulatory relationships of SNPs, genes and proteins that are previously identified in existing literature or provided in public databases. This structure allows MoFNet to handle small sample size with significantly reduced model complexity. Compared to existing knowledge-primed models, MoFNet is capable of handling three-omics types and at the same time explicitly model the information flow from SNPs to genes and proteins. We hypothesize this can better capture the dynamic molecular process underlying AD and mitigate the missing information in the gene/protein expression snapshot. We incorporated a Lasso penalty to guide MoFNet's attention towards the information flow within disease-relevant interactions exclusively. This is anticipated to enhance performance in predicting disease outcomes and potentially uncover more tissue-specific molecular mechanisms. When applied to multi-omics data in the ROS/MAP cohort, MoFNet outperformed all other competing methods across a comprehensive set of evaluation metrics and identified genes/proteins with significantly more functional interactions. It returned three molecular subnetworks in relate to innate immune system, clearance of misfolded proteins, and neurotransmitter release respectively.

MATERIALS AND METHODS

Multi-omic data sets

Genotype, RNA-Seq gene expression and protein expression used for discovery were sourced from

the Religious Orders Study (ROS) and Memory and Aging Project (MAP) cohorts and were collected from prefrontal cortex tissue of postmortem brains [11], as shown in Fig. 1a. The replication data set, also derived from the same brain tissue, was obtained from the Mount Sinai Brain Bank (MSBB) cohort [12]. Details of two cohorts are in Supplementary Text S1 and S2 and detailed preprocessing steps of genotype, RNASeq and protein expression data can be found in Supplementary Text S3, S4, and S5. Shown in Table 1 is the demographic information of all participants.

Pre-screening of -omics features

Given the high dimension of genotype and transcriptome data, this analysis was designed to specifically target the molecular mechanism related to the 186 peptides measured in ROS/MAP. These peptides correspond to 126 unique proteins, which were manually selected by the ROS/MAP team as AD-relevant. On top of that, we took a bottom-up approach to further narrow down the number of SNPs and genes. This is also expected to mitigate the effect of small sample size. First, peptides were used as initial seeds to help select a subset of related SNPs and genes for subsequent analysis (Fig. 1b). In the proteomic layer, 186 peptides measured in the ROS/MAP cohort were mapped to 126 unique genes (gene set A), which were found to functionally interact with 954 genes (gene set B) in the Reactome database [5]. Among these 1,080 (126 + 954) genes, 743 without missing RNA-seq data were included to represent the transcriptomic layer in our model. In the genomic layer, we identified SNPs located on the upstream of these 743 genes within the boundary of 5K base pairs. To ensure the functional connection of selected SNPs and their downstream genes, we included only SNPs significantly associated with the transcription factor-binding activity, based on the SNP2TFBS database [6]. Taken together, 822 SNPs, 743 genes, and 186 peptides were included for the subsequent predictive task. The functional relationships used to filter the genes and SNPs formed a trans-omic network and will be embedded into the architecture of MoFNet to guide the search for molecular subnetworks related to AD (Fig. 1c).

Prediction outcomes

Extracted SNP genotype, gene expression and protein expression data were used to classify AD patients

from cognitive normals (CNs). For all the participants included in this study, their clinical diagnosis at the time of brain tissue collection was used to indicate their disease status. In this case, the diagnosis time aligns with the -omics data collection time.

MoFNet: a deep multi-omic fused neural network

MoFNet is an extension from Varmole [10] and knowledge-primed neural network (KPNN) [9]. Both of them confirmed that deep learning informed by prior functional interactions could lead to significant improvement in both prediction performance and model interpretability. However, KPNN takes only RNA-Seq data as input and Varmole can additionally take genotype. Neither of them is capable of taking genotype, RNA-Seq and protein expression altogether with prior interactions. MoFNet addressed this limitation with a novel multi-layer architecture, as shown in Fig. 1c. The first transparent layer has 1,565 input nodes, corresponding to 822 SNPs and 743 genes respectively, and 743 output nodes, corresponding to 743 enhanced genes. Links in the first transparent layer were added if one SNP (as an input node) is connected to one gene (as an output node) in the prior trans-omic network. In addition, we added links between duplicated genes nodes (i.e., gene A node in the input and gene A node in the output). In this case, each output gene node will have integrated information from its upstream SNPs and its originally measured expression level. We assume that not all upstream SNPs are equally informative and may be specific to different tissues. Therefore, L1 regularization was applied for the first transparent layer such that output gene nodes will not integrate information from SNPs that are not relevant to the prediction outcomes. That is, links between irrelevant SNPs and their downstream genes will mostly get zero weight. For the second transparent layer, we have 929 input nodes (i.e., 743 enhanced genes and 186 original peptides), and 186 output nodes corresponding to 186 enhanced peptides. Links were added if one gene (as an input node) is connected to one peptide (as an output node) as indicated in the prior trans-omic network. We also added links between the duplicated peptides (i.e., peptide A node in the input and peptide A node in the output). Taken together, the enhanced peptide nodes, as the output of the second transparent layer, will have integrated information from its corresponding genes, their functional interactors and upstream SNPs. After that, we have fully connected layers to classify the AD patients from cognitive normal sub-

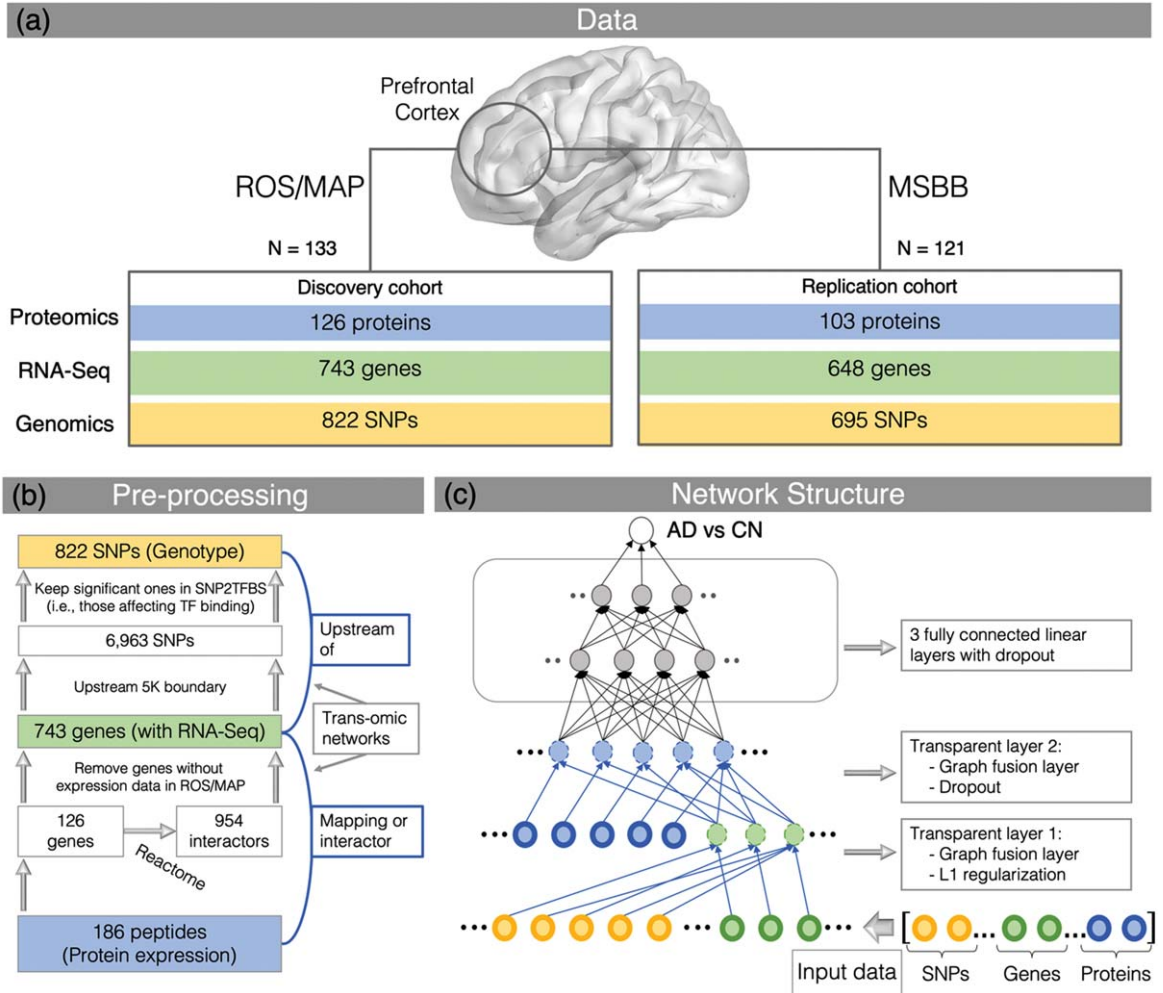


Fig. 1. (a) Multi-omic data used for discovery and replication, both from prefrontal cortex tissue. (b) Steps to filter SNPs and genes using peptides as seeds. (c) Architecture of the proposed MoFNet model, with the prior trans-omic network in (b) embedded in the first two transparent layers.

Table 1
Demographic information of all participants

Diagnosis	ROS/MAP		MSBB	
	CN	AD	CN	AD
Subject Number	77	56	41	80
Male/Female	35/42	22/34	21/20	23/57
Education (mean $\pm$ std.)	16.7 $\pm$ 3.2	16.8 $\pm$ 3.7	N/A	N/A
Age (mean $\pm$ std.)	83.0 $\pm$ 4.5	86.3 $\pm$ 3.5	83.1 $\pm$ 8.1	85.1 $\pm$ 6.8

jects. To avoid the overfitting problem, we applied dropout and early termination. Detailed architecture of each layer is described below.

1. Input X_1 is the concatenation of the gene expression matrix $G^{n \times g}$ (n samples by g genes), and SNP genotype matrix $S^{n \times s}$ (n samples by s

SNPs). $X_1^{n \times (g+s)} = [G, S]$ where $[\cdot]$ stands for row concatenation.

2. The output from the first transparent layer Z_{11} has the dimension as the number of genes g . Links in this layer indicate the prior functional connections between SNPs and genes,

and between same genes. Functional connections between SNPs and genes were encoded in an adjacency matrix $A_1^{s \times g}$. $A(i, j) = 1$ indicates SNP i is located upstream of gene j and likely to affect the transcription factor binding activity; $A(i, j) = 0$ otherwise. We also added self-connections to genes by adding another adjacency matrix $A_2^{g \times g}$, which is an identity matrix with $A_2(i, j) = 1$. Taken these two adjacency matrices together, the first transparent layer is a ‘Biological DropConnect’ layer [10, 13]. Therefore, weight matrix of this layer W_1 has a sparse structure with a dimension of $(s + g) \times g$. Output of this layer $Z_1 = f\left(X_1 \left(W_1 \odot [A_1^T, A_2^T]^T\right) + b_1\right)$ where $\odot$ is the Hadamard product, $(\cdot)^T$ is the matrix transpose operator, and b_1 is the bias term.

3. The second transparent layer resembles the structure of the second part of the prior trans-omic network, i.e., the functional connections between genes and proteins. The input of this layer is the concatenation of the protein expression (e.g., peptides) data $X_2^{n \times p}$ (n samples by p peptides) and output of the first transparent layer Z_{11} , i.e., $Z_1 = [X_2, Z_{11}]$. The output of the second transparent layer Z_2 has a dimension of the number of peptides. Weight of this layer W_2 has a dimension of $(g + p) \times p$. The adjacency matrix $A_3^{g \times p}$ indicates the functional connections between genes and proteins, where $A_3(i, j) = 1$ if gene i encodes protein j itself or the functional interactor of protein j ; $A_3(i, j) = 0$, otherwise. Similarly, we added self-connections between peptides with an identity adjacency matrix $A_4^{p \times p}$ where $A_4(i, j) = 1$. The output of the second transparent layer is $Z_2 = f\left(Z_1 \left(W_2 \odot [A_3^T, A_4^T]^T\right) + b_2\right)$, where b_2 is the bias term.
4. Three fully connected hidden layers Z_l index by $l \in \{3 \cdots L - 1\}$ were used together with a sigmoid function in the last layer. $Z_L = \sigma(Z_{L-1}W_L + b_L)$, where b_L is the bias term.
1. Finally, we use binary cross-entropy loss to quantify the classification error: $L(y, \hat{y}) = -1/n \sum_{i=1}^n y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)$.

Due to the small sample size, 5-fold cross validation was applied with grid search to tune the parameters. Details of parameter tuning and final parameters can be found in the Supplementary Text S6.

Model interpretation

MoFNet can be interpreted in two ways. First, with L1 penalty, only a few links in the first two layers will get non-zero weight. Secondly, each node will obtain an importance score using integrated gradient, a common method to interpret deep learning models [14, 15]. Importance score of each node measures how much the prediction outcome will change in case of unit change in that node. It is a relative measure for prioritizing the contribution of SNPs, genes, and proteins. The higher the importance score, the larger effect the feature has on the final prediction. Link weight and node importance score will be used to prune prior network and return subnetworks that contribute the most to the disease outcome.

Performance evaluation

To evaluate the effectiveness of the proposed MoFNet, we compared its performance against random forest and four other logistic regression based classification models, using modularity, elastic net, GraphNet and Lasso as penalty terms respectively [16–19]. These sparse logistic regression models were selected because they are designed for both classification and feature selection. Classic classification models, such as support vector machine (SVM) and k nearest neighbor (KNN), do not select features, and therefore are not included for comparison. For the proposed MoFNet, we also evaluated its isoforms structured using shuffled prior trans-omic networks. In total, we repeated shuffling three times and reported the average performance. Two different shuffling strategies were applied: with and without self-connections for gene and protein nodes (e.g., link between gene A node in layer 1 and gene A node in layer 2). Modularity constrained logistic regression (M-logistic) was implemented using Matlab and GraphNet was implemented using R package [20]. Elastic net constrained logistic regression, traditional logistic regression with lasso penalty, and random forest were implemented using the Python scikit-learn package [21]. To provide an unbiased comparison, partition of subjects in all training and testing set was kept identical for all methods. Grid search and 5-fold cross validation for all methods were used to select optimal parameters.

The classification accuracy and the Area Under the Receiver Operating Characteristic (ROC) Curve (AUC) were utilized as primary metrics. Additionally, F1 score, precision, and specificity were

employed for a comprehensive assessment of the model's performance.

RESULTS

Comparison of prediction performance

Shown in Fig. 2 is the performance of MoFNet and other competing methods on test data set. Due to imbalanced case and controls, we reported not only accuracy and AUC, but also F1 score, precision, and specificity metrics to give a comprehensive comparison of performance. In particular, F1 score combines precision and sensitivity (recall) into a single metric, and has been widely used as a major evaluation metric for imbalanced data sets. The proposed MoFNet consistently outperforms other competing models, with highest accuracy, specificity, AUC, and F1-score, indicating its capability in handling the imbalanced data set as well. The MoFNet using shuffled trans-omic network without self-connections remained the worst across all evaluation metrics, which was expected. The MoFNet using shuffled trans-omic network with self-connections scored the second-best. This suggests that the prediction performance of MoFNet was predominantly driven by the genotype, gene expression and protein expression that were originally measured. And indeed, integrated information from functional interactors provided additional information to further enhance the prediction performance.

Multi-omic molecular subnetworks in related to AD

We examined the set of SNPs, genes and/or proteins contributing to the final prediction and their functional connections in the prior network. Among all competing methods, modularity and elastic net constrained logistic regression models identified SNPs, genes, and proteins with a few known functional connections, but not enough to form subnetworks. Multi-omic features selected by other competing methods mostly scattered around the prior network with little known functional connections. In contrast, MoFNet returned multi-omic molecular subnetworks connecting a set of SNPs with genes and proteins. These subnetworks were obtained by pruning the prior network using link coefficients and node importance scores (details in Supplementary Text S7). Final pruned multi-omic network has 169 multi-omic features (32 proteins, 80 genes, and

57 SNPs), with 3 major connected components (i.e., subnetworks) (Fig. 3). Here, node size is made proportional to its importance score. The larger the node size, the more it contributes to the predicted diagnosis. Similarly, edge width is proportional to the weight taken from the transparent layers of MoFNet. The thicker the edge is, the more information flow occurs from SNPs to genes or from genes to proteins. As expected, nodes with top importance scores are mostly proteins since they integrated information from SNPs and genes. One peptide of Apolipoprotein E (*APOE*), a strong risk factor for AD [22], was also identified by MoFNet but not part of top 3 connected components. This is likely due to its limited functional connections in the prior network.

Component-1 has 24 SNPs, 34 genes and 21 proteins, including the amyloid precursor protein *APP* and its corresponding gene. In particular, peptide *APP_2* has the maximum degree, and peptide *Tau_AT8* has a high importance score indicating its significant contribution to the predicted diagnosis of each subject. AT8 and AT100, shown at the end of tau peptides, are antibodies used to recognize tau protein phosphorylated at Ser202/Thr205 and Thr212/Ser214 respectively [23, 24]. Our result also identified other phosphorylated forms of tau to be highly predictive of AD, such as those captured by antibodies PHF1, 12E8, and 77G7 [25, 26]. Several studies have demonstrated that the presence of phosphorylated tau detected by AT8 or AT100, is strongly correlated with the presence of neurofibrillary tangles in AD brains [27, 28]. Activation of *NLRP3* inflammasome is speculated to promote the formation of neurofibrillary tangles and the accumulation of abnormal *tau_AT8* in the brain [29].

Identified SNPs are largely tissue-specific expression quantitative trait loci (eQTL)

We investigated the functional effect of all SNPs in 3 major components on the downstream transcriptome level. In the Brain eQTL Almanac (BRAINEAC) database [30], 50 out of 53 SNPs were found significant as eQTL in the frontal cortex region; the rest 3 SNPs were not found. That means, variations in these SNPs are associated with gene expression levels. Further, we examined whether those SNPs are eQTLs of their directly connected genes in Fig. 3. Among all 50 SNP-gene pairs in Fig. 3, 29 of them have been identified to have significant associations in the frontal cortex tissue (links highlighted as blue in Fig. 3). Top signif-

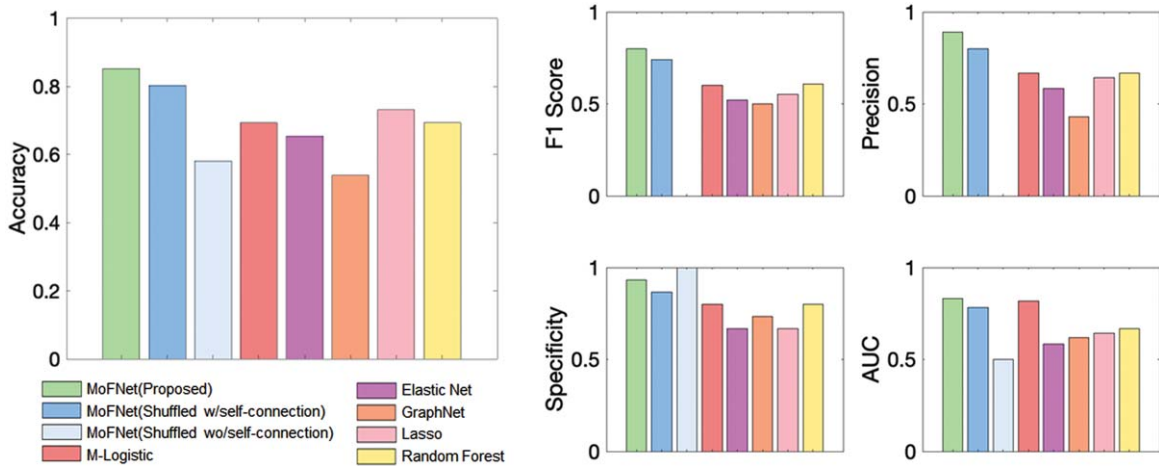


Fig. 2. Performance comparison of MoFNet and other competing methods on test data set. Three MoFNet models were evaluated, structured with prior multi-omic network, shuffled prior network with self-connections and shuffled prior network without self-connections respectively.

icant associations between SNPs and genes from component-1 were listed in Supplementary Table 1. Most significant information flow was found between *VAMP2* gene and its two upstream SNPs, *rs4792267* and *rs59511744*. According to GeneHancer [31], both SNPs are located in a promoter/enhancer region that interacts with genes lysine demethylase 6B (*KDM6B*), CST telomere replication complex component 1 (*CTCI*), phosphoribosylformylglycinamide synthase (*PFAS*), and period circadian regulator 1 (*PER1*). These four genes play essential roles in regulation of synaptic plasticity, telomere maintenance, neurotoxicity, and circadian rhythm respectively, all of which are strong correlates of AD [32–35].

Enriched cell types

We further estimated the cell types related to altered genes and proteins using cell type specific marker genes. Those marker genes usually have selectively high expression in specific cell types, but not in others. Based on CellMarker database [36], we examined all the genes and proteins from 3 major network components to check whether they are marker genes of any cell type. As shown in Supplementary Figure 2, genes/proteins identified by MoFNet are mostly marker genes for astrocyte, microglia, and neuronal cells. Detailed list of marker genes in each component is shown in Supplementary Table 2. Astrocytes are known to provide support and nourishment to neurons, while microglia cells protect against threats and remove damaged cells [37]. Both

astrocytes and microglia are glial cells mediating the neuroinflammation, which has been viewed as a “double-sword” in AD [38]. Cross-talk between astrocyte and microglia has been recently suggested as a potential target for therapeutic intervention in AD [39].

Pathway enrichment analysis

For all the genes and proteins in 3 major network components, we performed enrichment analysis in Reactome pathways using g:Profiler [5, 40]. Shown in Fig. 4 is the map of all enriched Reactome pathways, falling into eight functional groups [41, 42]. Top pathways enriched by genes and proteins in component-1 include notch1 NLR (NOD-like receptors) signaling, eph ephrin cells, release neurotransmitter cycle, and caspase mediated cleavage. NLR signaling pathways are known to be associated with the inflammatory response in AD [43]. Emerging evidence suggests that Eph-Ephrin signaling is associated with both synaptic dysfunction and immune dysregulation, which in turn promotes the progression of AD [44]. Dysregulation of neurotransmitter release has been involved in the development of AD by affecting neural communication and plasticity, which will possibly lead to the death of nerve cells [45]. Lastly, caspase-mediated cleavage of tau is viewed as an early pathological event triggering tangle pathology as a critical toxic moiety underlying neurodegeneration [46, 47].

Component-2 is a subnetwork centered around protein *FBXO2* (F-box protein 2). It is closely related

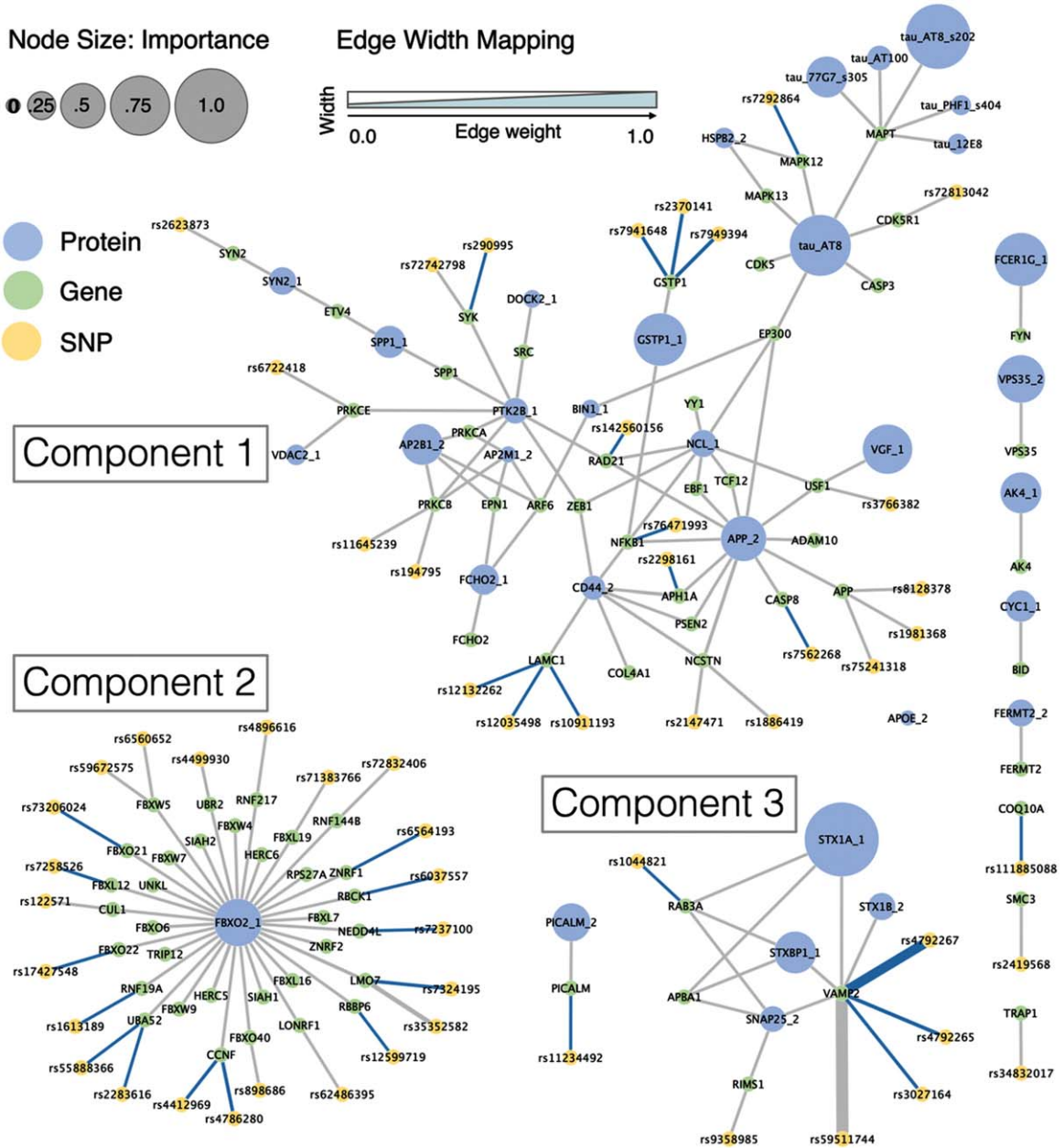
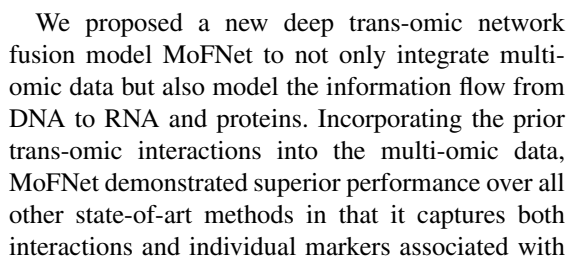


Fig. 3. Molecular subnetworks (with number of nodes ≥ 2) from pruned prior network using the link weights and node importance scores from MoFNet. Node size is proportional to importance score. Link width is proportional to the link weight. Blue links represent SNP pairs where the SNP is a known eQTL of its connected gene in frontal cortex tissue. Suffix of each protein indicates different peptides.

to ubiquitination & proteasome degradation as part of antigen processing (adjusted $p = 1.808e-51$) [5, 40]. Neuronal death in AD has a strong connection with misfolded proteins that aggregate within the brain, e.g. amyloid and tau tangles. Ubiquitination and proteasome degradation is one of the two major pathways that help get rid of unwanted cells or misfolded pro-

teins to prevent their accumulation and to maintain the health of a cell [48].

Component-3 is a small subnetwork centered around protein *STX1A_1*, *SNAP25_2*, *STXB1_1*, and gene *VAMP2*. They are the major components of the SNARE complex, which mediates the fusion process of synaptic vesicles and play an essential role



AD, giving rise to molecular subnetworks for easier interpretation of associated molecular mechanism. Trans-omic paths from SNP to gene and then protein identified by MoFNet suggested that AD may be partly the result of genetic variations due to their cascading effects on the downstream transcriptome and proteome levels. Although none of the prior functional connections was extracted in a tissue specific manner, eQTL analysis showed that MoFNet can accurately capture those tissue specific relationships between SNPs and genes.

MoFNet identified several AD-relevant molecular subnetworks connecting SNPs, genes and proteins, providing a zoomed-in view of possible AD molecular mechanism. Component-3 consists exclusively of SNARE (soluble N-ethylmaleimide-sensitive factor attachment protein receptor) complex priming the synaptic vesicle fusion for neurotransmitter release, including synaptobrevin (also known as vesicle-associated membrane protein 2 (*VAMP2*)) on the synaptic vesicle, syntaxin 1 (*STX1A/STX1B*) on the plasma membrane, and *SNAP25*. These three functions together with *STXBPI* to bring the membranes of synaptic vesicle and plasma membrane into apposition and to enable neurotransmission [49]. The SNARE complex is necessary not only for neuron-neuron communication but also neuron-glia and glia-glia communication [50].

Interestingly, for component-1, enrichment results using EnrichR are also highly relevant to synaptic functions, in addition to signaling pathways and immune system [53]. More specifically, 23 out of 34 genes in component-1 are involved in immune system (adjusted $p = 2.289e - 14$), 25 in signaling transduction (adjusted $p = 1.034e - 15$), 14 in axon guidance (adjusted $p = 2.235e - 12$) and 14 in synaptic function (aggregated from all enriched synaptic GO terms with adjusted $p \leq 0.05$). Out of 14 synaptic function related genes, 10 of them are involved in both immune system and signal transduction and 7 of them are related to neuron axon guidance. This suggests that synaptic function is an integral part of immune response and signal transduction. Top signaling pathways enriched by component-1 like Notch pathways, NTR (neurotrophin) signaling pathways, and NGF (nerve growth factor) are all activated by their corresponding receptors located in the cell membrane. SNARE complex in component-3 along with GTPase will help merge endocytic vesicles that transport these receptors to the plasma membrane [54], which agrees with previous findings that notch signaling is likely regulated by intracellular vesicle

trafficking [54]. These signaling pathways play essential role in normal development of neurons and glia and mediating their cross-talk [55].

Synaptic dysfunction is among the earliest changes of AD, even before the accumulation of misfolded protein aggregates and neuronal loss [56–58]. Synapse loss in postmortem brains is a strong correlate with cognitive decline [59, 60]. Yet, the connection of synapse dysfunction with amyloid and tau pathology is not fully understood. Synapse is the place where amyloid- β peptides are generated and is the target of the toxic amyloid- β oligomers [61]. Oligomeric amyloid- β is found to co-localize with Apolipoprotein E4 protein, which is associated with significant increase of amyloid- β at synapses [62]. A recent study reported the interaction of tau in neuron with *STX1A*, a member of the SNARE complex, suggesting the localization of tau at sites of presynaptic vesicle fusion [63].

Taken together, both amyloid and tau are likely to localize at the presynaptic vesicle and have domain-specific interactions with synaptic vesicle-associated proteins, interfering with synaptic vesicle function in the early stage of AD. Also, considering that all these genes are preferably expressed in brain according to GTEx [64, 65] and are mostly markers of astrocyte, microglia and neurons (Fig. 4), such interference of synaptic function could possibly cause disruption in neuron-neuron or neuron-glia cross talk and further lead to neuronal and synapse loss in AD. Further investigation of these identified genes/proteins could possibly help decipher the mechanisms underlying synaptic dysfunction in AD, and ultimately inform therapeutic strategies to modify AD progression.

We proposed a new deep multi-omic fusion network to leverage the information flow in prior functional interactions to improve the prediction performance and model interpretability. When applied to AD multi-omic data, MoFNet achieved promising performance in identifying molecular subnetworks associated to AD, along with the upstream regulatory SNPs. These findings have been mostly replicated in an independent dataset, confirming the robustness of the MoFNet even with limited sample size. It's also important to note that while our prior network is not tissue specific, MoFNet is capable of identifying functional interactions that are specific to the tissue where the omics data is collected, which underscores the great potential of this method. As such, MoFNet offers a significant opportunity to advance our understanding of disease molecular mechanisms in AD and beyond.

AUTHOR CONTRIBUTIONS

Linhui Xie (Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Software; Validation; Visualization; Writing – original draft; Writing – review & editing); Yash Raj (Conceptualization; Formal analysis; Methodology; Software; Validation; Writing – original draft); Pradeep Varathan (Conceptualization; Visualization); Bing He (Formal analysis; Investigation; Visualization); Meichen Yu (Conceptualization; Writing – review & editing); Kwangsik Nho (Data curation; Writing – review & editing); Paul Salama, Ph.D. (Methodology; Supervision; Writing – review & editing); Andrew J. Saykin (Data curation; Resources; Writing – review & editing); Jingwen Yan (Conceptualization; Funding acquisition; Investigation; Methodology; Supervision; Visualization; Writing – original draft; Writing – review & editing).

ACKNOWLEDGMENTS

The results published here are in whole or in part based on data obtained from the AMP-AD Knowledge Portal. Study data were provided by the Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago. Data collection was supported through funding by NIA grants P30AG10161, R01AG15819, R01AG17917, R01AG30146, R01AG36836, U01AG32984, U01AG46152, the Illinois Department of Public Health, and the Translational Genomics Research Institute.

FUNDING

This research was supported by NIH grants R21 AG072101, U19 AG074879, U01 AG068057, and NSF CAREER 1942394.

CONFLICT OF INTEREST

Dr. Saykin has received support from Avid Radiopharmaceuticals, a subsidiary of Eli Lilly (in kind contribution of PET tracer precursor) and holds advisory roles with Siemens Medical Solutions USA, Inc., NIH NHLBI, and Eisai. His editorial commitments include serving as Editor-in-Chief for the journal "Brain Imaging and Behavior", and he participates in various NIH/NIA advisory committees.

Dr. Saykin is an Editorial Board Member of this journal but was not involved in the peer-review process of this article nor had access to any information regarding its peer-review.

All other authors have no conflict of interest to report.

DATA AVAILABILITY

Multi-omic data used in this analysis is from the ROS/MAP project and is available after application through the AMP-AD knowledge portal (<https://adknowledgeportal.synapse.org>). The source code is available through GitHub (<https://github.com/JW-Yan/MoFNet>).

SUPPLEMENTARY MATERIAL

The supplementary material is available in the electronic version of this article: <http://dx.doi.org/10.3233/JAD-240098>.

REFERENCES

- [1] Alzheimer's Association (2023) 2023 Alzheimer's disease facts and figures. *Alzheimers Dement* **19**, 1598-1695.
- [2] Dubois B, Feldman HH, Jacova C, Hampel H, Molinuevo JL, Blennow K, DeKosky ST, Gauthier S, Selkoe D, Bateman R, Cappa S, Crutch S, Engelborghs S, Frisoni GB, Fox NC, Galasko D, Habert MO, Jicha GA, Nordberg A, Pasquier F, Rabinovici G, Robert P, Rowe C, Salloway S, Sarazin M, Epelbaum S, de Souza LC, Vellas B, Visser PJ, Schneider L, Stern Y, Scheltens P, Cummings JL (2014) Advancing research diagnostic criteria for Alzheimer's disease: The IWG-2 criteria. *Lancet Neurol* **13**, 614-629.
- [3] Clark C, Dayon L, Masoodi M, Bowman GL, Popp J (2021) An integrative multi-omics approach reveals new central nervous system pathway alterations in Alzheimer's disease. *Alzheimers Res Ther* **13**, 71.
- [4] Menyhart O, Györfy B (2021) Multi-omics approaches in cancer research with applications in tumor subtyping, prognosis, and diagnosis. *Comput Struct Biotechnol J* **19**, 949-960.
- [5] Jassal B, Matthews L, Viteri G, Gong C, Lorente P, Fargat A, Sidiropoulos K, Cook J, Gillespie M, Haw R, Loney F, May B, Milacic M, Rothfels K, Sevilla C, Shamovsky V, Shorser S, Varusai T, Weiser J, Wu G, Stein L, Hermjakob H, D'Eustachio P (2020) The reactome pathway knowledgebase. *Nucleic Acids Res* **48**, D498-D503.
- [6] Kumar S, Ambrosini G, Bucher P (2017) SNP2TFBS – a database of regulatory SNPs affecting predicted transcription factor binding site affinity. *Nucleic Acids Res* **45**, D139-D144.
- [7] Xie L, He B, Varathan P, Nho K, Risacher SL, Saykin AJ, Salama P, Yan J (2021) Integrative-omics for discovery of network-level disease biomarkers: A case study in Alzheimer's disease. *Brief Bioinform* **22**, bbab121.
- [8] Xu J, Mao C, Hou Y, Luo Y, Binder JL, Zhou Y, Bekris LM, Shin J, Hu M, Wang F, Eng C, Oprea TI, Flanagan ME, Pieper AA, Cummings J, Leverenz JB, Cheng F (2022) Interpretable deep learning translation of GWAS

- and multi-omics findings to identify pathobiology and drug repurposing in Alzheimer's disease. *Cell Rep* **41**, 111717.
- [9] Fortelny N, Bock C (2020) Knowledge-primed neural networks enable biologically interpretable deep learning on single-cell sequencing data. *Genome Biol* **21**, 190.
 - [10] Nguyen ND, Jin T, Wang D (2021) Varmole: A biologically drop-connect deep neural network model for prioritizing disease risk variants and genes. *Bioinformatics* **37**, 1772-1775.
 - [11] Bennett DA, Schneider JA, Arvanitakis Z, Wilson RS (2012) Overview and findings from the religious orders study. *Curr Alzheimer Res* **9**, 628-645.
 - [12] Wang M, Beckmann ND, Roussos P, Wang E, Zhou X, Wang Q, Ming C, Neff R, Ma W, Fullard JF, Hauberg ME, Bendl J, Peters MA, Logsdon B, Wang P, Mahajan M, Mangravite LM, Dammer EB, Duong DM, Lah JJ, Seyfried NT, Levey AI, Buxbaum JD, Ehrlich M, Gandy S, Katsel P, Haroutunian V, Schadt E, Zhang B (2018) The Mount Sinai cohort of large-scale genomic, transcriptomic and proteomic data in Alzheimer's disease. *Sci Data* **5**, 180185.
 - [13] Wan L, Zeiler M, Zhang S, Cun YL, Fergus R (2013) *Regularization of Neural Networks using DropConnect*, PMLR, Proceedings of Machine Learning Research.
 - [14] Sundararajan M, Taly A, Yan Q (2017) *Axiomatic Attribution for Deep Networks*, PMLR, Proceedings of Machine Learning Research.
 - [15] Kokhlikyan N, Miglani V, Martin M, Wang E, Alsallakh B, Reynolds J, Melnikov A, Kliushkina N, Araya C, Yan S, Reblitz-Richardson O (2020) Captum: A unified and generic model interpretability library for PyTorch. arXiv:2009.07896.
 - [16] Newman ME (2006) Modularity and community structure in networks. *Proc Natl Acad Sci U S A* **103**, 8577-8582.
 - [17] Zou H, Hastie T (2005) Regularization and variable selection via the elastic net. *J R Stat Soc Series B Stat Methodol* **67**, 301-320.
 - [18] Grosenick L, Klingenberg B, Katovich K, Knutson B, Taylor JE (2013) Interpretable whole-brain prediction analysis with GraphNet. *Neuroimage* **72**, 304-321.
 - [19] Tibshirani R (1996) Regression shrinkage and selection via the lasso. *J R Stat Soc Series B Methodol* **58**, 267-288.
 - [20] Chen L, Liu H, Kocher JP, Li H, Chen J (2015) glmgraph: An R package for variable selection and predictive modeling of structured genomic data. *Bioinformatics* **31**, 3991-3993.
 - [21] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V (2011) Scikit-learn: Machine learning in Python. *J Mach Learn Res* **12**, 2825-2830.
 - [22] Liu CC, Liu CC, Kanekiyo T, Xu H, Bu G (2013) Apolipoprotein E and Alzheimer disease: Risk, mechanisms and therapy. *Nat Rev Neurol* **9**, 106-118.
 - [23] Goedert M, Jakes R, Vanmechelen E (1995) Monoclonal antibody AT8 recognises tau protein phosphorylated at both serine 202 and threonine 205. *Neurosci Lett* **189**, 167-169.
 - [24] Zheng-Fischhofer Q, Biernat J, Mandelkow EM, Illenberger S, Godemann R, Mandelkow E (1998) Sequential phosphorylation of Tau by glycogen synthase kinase-3 β and protein kinase A at Thr212 and Ser214 generates the Alzheimer-specific epitope of antibody AT100 and requires a paired-helical-filament-like conformation. *Eur J Biochem* **252**, 542-552.
 - [25] Avila J (2006) Tau phosphorylation and aggregation in Alzheimer's disease pathology. *FEBS Lett* **580**, 2922-2927.
 - [26] Gordon-Krajcer W, Kozniowska E, Lazarewicz JW, Ksiezak-Reding H (2007) Differential changes in phosphorylation of tau at PHF-1 and 12E8 epitopes during brain ischemia and reperfusion in gerbils. *Neurochem Res* **32**, 729-737.
 - [27] Grundke-Iqbal I, Iqbal K, Tung YC, Quinlan M, Wisniewski HM, Binder LI (1986) Abnormal phosphorylation of the microtubule-associated protein tau (tau) in Alzheimer cytoskeletal pathology. *Proc Natl Acad Sci U S A* **83**, 4913-4917.
 - [28] Brion JP (1998) Neurofibrillary tangles and Alzheimer's disease. *Eur Neurol* **40**, 130-140.
 - [29] Ising C, Venegas C, Zhang S, Scheiblich H, Schmidt SV, Vieira-Saecker A, Schwartz S, Albasset S, McManus RM, Tejera D, Griep A, Santarelli F, Brosseron F, Opitz S, Stunden J, Merten M, Kaye R, Golenbock DT, Blum D, Latz E, Buee L, Heneka MT (2019) NLRP3 inflammasome activation drives tau pathology. *Nature* **575**, 669-673.
 - [30] Ramasamy A, Trabzuni D, Guelfi S, Varghese V, Smith C, Walker R, De T, Consortium UKBE, North American Brain Expression C, Coin L, de Silva R, Cookson MR, Singleton AB, Hardy J, Ryten M, Weale ME (2014) Genetic variability in the regulation of gene expression in ten regions of the human brain. *Nat Neurosci* **17**, 1418-1428.
 - [31] Fishilevich S, Nudel R, Rappaport N, Hadar R, Plaschkes I, Iny Stein T, Rosen N, Kohn A, Twik M, Safran M, Lancet D, Cohen D (2017) GeneHancer: Genome-wide integration of enhancers and target genes in GeneCards. *Database (Oxford)* **2017**.
 - [32] Wang Y, Khandelwal N, Liu S, Zhou M, Bao L, Wang JE, Kumar A, Xing C, Gibson JR, Wang Y (2022) KDM6B cooperates with Tau and regulates synaptic plasticity and cognition via inducing VGLUT1/2. *Mol Psychiatry* **27**, 5213-5226.
 - [33] Feng X, Hsu SJ, Kasbek C, Chaiken M, Price CM (2017) CTC1-mediated C-strand fill-in is an essential step in telomere length maintenance. *Nucleic Acids Res* **45**, 4281-4293.
 - [34] Foguth R, Sepulveda MS, Cannon J (2020) Per- and polyfluoroalkyl substances (PFAS) neurotoxicity in sentinel and non-traditional laboratory model systems: Potential utility in predicting adverse outcomes in human health. *Toxics* **8**, 42.
 - [35] Musiek ES, Holtzman DM (2016) Mechanisms linking circadian clocks, sleep, and neurodegeneration. *Science* **354**, 1004-1008.
 - [36] Hu C, Li T, Xu Y, Zhang X, Li F, Bai J, Chen J, Jiang W, Yang K, Ou Q, Li X, Wang P, Zhang Y (2022) CellMarker 2.0: An updated database of manually curated cell markers in human/mouse and web tools based on scRNA-seq data. *Nucleic Acids Res* **51**, D870-D876.
 - [37] Dong Y, Benveniste EN (2001) Immune function of astrocytes. *Glia* **36**, 180-190.
 - [38] Singh D (2022) Astrocytic and microglial cells as the modulators of neuroinflammation in Alzheimer's disease. *J Neuroinflammation* **19**, 206.
 - [39] McAlpine CS, Park J, Griciu A, Kim E, Choi SH, Iwamoto Y, Kiss MG, Christie KA, Vinegoni C, Poller WC, Mindur JE, Chan CT, He S, Janssen H, Wong LP, Downey J, Singh S, Anzai A, Kahles F, Jorfi M, Feruglio PF, Sadreyev RI, Weissleder R, Kleinstiver BP, Nahrendorf M, Tanzi RE, Swirski FK (2021) Astrocytic interleukin-3 programs microglia and limits Alzheimer's disease. *Nature* **595**, 701-706.
 - [40] Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, Vilo J (2019) g:Profiler: A web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res* **47**, W191-W198.

- [41] Merico D, Isserlin R, Stueker O, Emili A, Bader GD (2010) Enrichment map: A network-based method for gene-set enrichment visualization and interpretation. *PLoS One* **5**, e13984.
- [42] Kucera M, Isserlin R, Arkhangorodsky A, Bader GD (2016) AutoAnnotate: A Cytoscape app for summarizing networks with semantic annotations. *F1000Res* **5**, 1717.
- [43] Kong W, Zhang J, Mou X, Yang Y (2014) Integrating gene expression and protein interaction data for signaling pathway prediction of Alzheimer's disease. *Comput Math Methods Med* **2014**, 340758.
- [44] Ganguly D, Thomas JA, Ali A, Kumar R (2022) Mechanistic and therapeutic implications of EphA-4 receptor tyrosine kinase in the pathogenesis of Alzheimer's disease. *Eur J Neurosci* **56**, 5532-5546.
- [45] Xu Y, Yan J, Zhou P, Li J, Gao H, Xia Y, Wang Q (2012) Neurotransmitter receptors and cognitive dysfunction in Alzheimer's disease and Parkinson's disease. *Prog Neurobiol* **97**, 1-13.
- [46] Gamblin TC, Chen F, Zambrano A, Abbrah A, Lagalwar S, Guillozet AL, Lu M, Fu Y, Garcia-Sierra F, LaPointe N, Miller R, Berry RW, Binder LI, Cryns VL (2003) Caspase cleavage of tau: Linking amyloid and neurofibrillary tangles in Alzheimer's disease. *Proc Natl Acad Sci U S A* **100**, 10032-10037.
- [47] Rissman RA, Poon WW, Blurton-Jones M, Oddo S, Torp R, Vitek MP, LaFerla FM, Rohn TT, Cotman CW (2004) Caspase-cleavage of tau is an early event in Alzheimer disease tangle pathology. *J Clin Invest* **114**, 121-130.
- [48] Schmidt MF, Gan ZY, Komander D, Dewson G (2021) Ubiquitin signalling in neurodegeneration: Mechanisms and therapeutic opportunities. *Cell Death Differ* **28**, 570-590.
- [49] Chen XQ, Zuo X, Becker A, Head E, Mobley WC (2023) Reduced synaptic proteins and SNARE complexes in Down syndrome with Alzheimer's disease and the Dp16 mouse Down syndrome model: Impact of APP gene dose. *Alzheimers Dement* **19**, 2095-2116.
- [50] Vadisiute A, Meijer E, Szabo F, Hoerder-Suabedissen A, Kawashita E, Hayashi S, Molnar Z (2022) The role of snare proteins in cortical development. *Dev Neurobiol* **82**, 457-475.
- [51] Solas M, Puerta E, J Ramirez M (2015) Treatment options in Alzheimer's disease: The GABA story. *Curr Pharm Design* **21**, 4960-4971.
- [52] Jo S, Yarishkin O, Hwang YJ, Chun YE, Park M, Woo DH, Bae JY, Kim T, Lee J, Chun H, Park HJ, Lee DY, Hong J, Kim HY, Oh SJ, Park SJ, Lee H, Yoon BE, Kim Y, Jeong Y, Shim I, Bae YC, Cho J, Kowall NW, Ryu H, Hwang E, Kim D, Lee CJ (2014) GABA from reactive astrocytes impairs memory in mouse models of Alzheimer's disease. *Nat Med* **20**, 886-896.
- [53] Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, McDermott MG, Monteiro CD, Gundersen GW, Ma'ayan A (2016) Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* **44**, W90-97.
- [54] Yamamoto S, Charng WL, Bellen HJ (2010) Endocytosis and intracellular trafficking of Notch and its ligands. *Curr Top Dev Biol* **92**, 165-200.
- [55] Zhang Y, Li B, Cananzi S, Han C, Wang LL, Zou Y, Fu YX, Hon GC, Zhang CL (2022) A single factor elicits multilineage reprogramming of astrocytes in the adult mouse striatum. *Proc Natl Acad Sci U S A* **119**, e2107339119.
- [56] Nilsson J, Gobom J, Sjodin S, Brinkmalm G, Ashton NJ, Svensson J, Johansson P, Portelius E, Zetterberg H, Blennow K, Brinkmalm A (2021) Cerebrospinal fluid biomarker panel for synaptic dysfunction in Alzheimer's disease. *Alzheimers Dement (Amst)* **13**, e12179.
- [57] Selkoe DJ (2002) Alzheimer's disease is a synaptic failure. *Science* **298**, 789-791.
- [58] Masliah E, Mallory M, Alford M, DeTeresa R, Hansen LA, McKeel DW, Jr., Morris JC (2001) Altered expression of synaptic proteins occurs early during progression of Alzheimer's disease. *Neurology* **56**, 127-129.
- [59] DeKosky ST, Scheff SW (1990) Synapse loss in frontal cortex biopsies in Alzheimer's disease: Correlation with cognitive severity. *Ann Neurol* **27**, 457-464.
- [60] Scheff SW, Price DA, Schmitt FA, DeKosky ST, Mufson EJ (2007) Synaptic alterations in CA1 in mild Alzheimer disease and mild cognitive impairment. *Neurology* **68**, 1501-1508.
- [61] Pelucchi S, Gardoni F, Di Luca M, Marcello E (2022) Synaptic dysfunction in early phases of Alzheimer's Disease. *Handb Clin Neurol* **184**, 417-438.
- [62] Koffie RM, Hashimoto T, Tai HC, Kay KR, Serrano-Pozo A, Joyner D, Hou S, Kopeikina KJ, Frosch MP, Lee VM, Holtzman DM, Hyman BT, Spires-Jones TL (2012) Apolipoprotein E4 effects in Alzheimer's disease are mediated by synaptotoxic oligomeric amyloid-beta. *Brain* **135**, 2155-2168.
- [63] Tracy TE, Madero-Perez J, Swaney DL, Chang TS, Moritz M, Konrad C, Ward ME, Stevenson E, Huttenhain R, Kauwe G, Mercedes M, Sweetland-Martin L, Chen X, Mok SA, Wong MY, Telpoukhovskaia M, Min SW, Wang C, Sohn PD, Martin J, Zhou Y, Luo W, Trojanowski JQ, Lee VMY, Gong S, Manfredi G, Coppola G, Krogan NJ, Geschwind DH, Gan L (2022) Tau interactome maps synaptic and mitochondrial processes associated with neurodegeneration. *Cell* **185**, 712-728 e714.
- [64] Baxter PS, Dando O, Emelianova K, He X, McKay S, Hardingham GE, Qiu J (2021) Microglial identity and inflammatory responses are controlled by the combined effects of neurons and astrocytes. *Cell Rep* **34**, 108882.
- [65] Consortium GT (2015) Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **348**, 648-660.