Nature Cell and Science 2024;2(4):207–223
https://doi.org/10.61474/ncs.2024.00043

Original Article

OPEN ACCESS

Using a Systems Biology Approach to Construct Genome-Wide Genetic and Epigenetic Networks for Investigating the Pathogenesis and Designing Multi-Molecule Drugs for Frontotemporal Dementia via a Deep Neural Network-based Drug-Target Interaction Model

Wei-Lun Chang and Bor-Sen Chen*

Laboratory of Automatic Control, Signal Processing and Systems Biology, Department of Electrical Engineering, National Tsing Hua University, Hsinchu 300044, Taiwan

***Correspondence to:** Bor-Sen Chen, Laboratory of Automatic Control, Signal Processing and Systems Biology, Department of Electrical Engineering, National Tsing Hua University, Hsinchu 300044, Taiwan. ORCID: https://orcid.org/0000-0003-1644-6106. Tel: +886-35731155, Fax: +886-5715971, E-mail: bschen@ee.nthu.edu.tw

Citation of this article: Chang WL, Chen BS. Using a Systems Biology Approach to Construct Genome-Wide Genetic and Epigenetic Networks for Investigating the Pathogenesis and Designing Multi-Molecule Drugs for Frontotemporal Dementia via a Deep Neural Network-based Drug-Target Interaction Model. *Nat Cell Sci* 2024;2(4):207–223. doi: 10.61474/ncs.2024.00043.

Abstract

Background and objectives: Frontotemporal dementia (FTD) is a slowly progressing neurodegenerative disease. In this study, we employ systems biology methods and genome-wide microarray data to investigate the pathogenetic mechanisms underlying FTD and identify significant biomarkers as potential drug targets. We then utilize a deep neural network (DNN)-based drug-target interaction (DTI) model, along with drug design specifications, to predict and select potential molecular drugs for treating FTD.

Methods: First, we constructed genome-wide genetic and epigenetic networks (GWGENs) for FTD and healthy control through large-scale database mining and system identification methods. We then extracted core GWGENs from these networks using a principal network projection method. Finally, we applied a DNN-based DTI model, trained on DTI databases, to predict potential molecular drugs.

Results: We identified key biomarkers involved in the pathogenesis of FTD, including TAU, GSK-3β, STAT3, *ATG5*, WDR41, and RIPK1. Based on these biomarkers, we selected Iodophenpropit, Clobenpropit, TTNPB, and Probucol as candidate molecules, which were combined into a multi-molecule drug for the potential treatment of FTD.

Conclusions: By combining a DNN-based DTI model with systems biology approaches, we designed a multi-molecular drug treatment for FTD. This approach enables more precise drug design and accurately predicts molecular combinations that may provide the most effective therapeutic outcomes.

Keywords: Frontotemporal dementia (FTD); Systems biology; Signal transduction; Biomarkers; Genome-wide genetic and epigenetic network (GWGEN); Principal network projection; DNN-based DTI model.

Introduction

Frontotemporal dementia (FTD) is a degenerative disease that primarily affects the frontal and anterior temporal lobes. It is the third most common form of primary dementia, following Alzheimer's disease and Lewy body dementia. While the exact cause of FTD remains unclear, genetic factors play a significant role, accounting for approximately 40% of cases. The disease is largely driven by these genetic factors and the degeneration of neurons in these regions. Patients typically experience a decline in various functions, particularly language and social behavior. FTD is a type of early-onset dementia that usually manifests between the ages of 45 and 64, with an average onset around 57.[1] The incidence is

Received: August 29, 2024 | **Revised:** October 10, 2024 | **Accepted:** November 27, 2024 | **Published online:** December 30, 2024

Nat Cell Sci 2024;2(4):207–223
https://doi.org/10.61474/ncs.2024.00043

roughly equal between men and women. Unlike Alzheimer's disease, which predominantly affects memory, FTD rarely involves memory impairment. Instead, early signs include behavioral abnormalities, deficits in executive function, and language difficulties, making it challenging for patients to recognize the condition earlier on.[2]

Currently, FTD cannot be completely cured, and drug development is still in progress. Only a few drugs have been tested in live mouse models, typically by observing whether early injection of the drug in mice with frontotemporal degeneration can prevent the occurrence of motor function abnormalities.[3] Common drugs used to treat FTD include rapamycin, spermidine, carbamazepine, tamoxifen, and other autophagy activators. These drugs help reduce the overexpression of TDP-43 protein.[4] Rapamycin, an mTOR inhibitor, is a central regulator of cell growth, controlling various cellular processes. In tumor cells, this regulatory mechanism can become dysregulated, leading to uncontrolled cell growth. Therefore, rapamycin is used to inhibit mTOR, helping to restore normal cellular function and prevent tumor progression.[5] In addition to rapamycin, drugs such as spermidine, carbamazepine, and tamoxifen can provide some control over FTD. Some individuals may also take antidepressants or antipsychotic medications to alleviate behavioral and emotional issues associated with FTD. However, these come with significant side effects, including neurological damage, drowsiness, dizziness. Currently, common therapies for FTD also include supportive therapies, speech therapy, and cognitive-behavioral therapy. These methods focus on improving patients' quality of life by controlling symptoms and enhancing communication and cognitive abilities.[6]

As noted above, there is currently no single drug or treatment that can completely cure FTD. However, the drug discovery process integrates computational techniques, experimental validation, translational models, and clinical trials to uncover potential therapeutic candidates. Despite considerable progress in biotechnology and an enhanced comprehension of biological systems, drug discovery remains an expensive, lengthy, and inefficient endeavor, with a high failure rate in developing new treatments.[7] Only about 10–20% of candidate drugs successfully progress from the start of clinical trials to market approval, a figure that has remained largely unchanged for decades.[8] Thus, there is a pressing need for a more efficient and systematic approach to drug design. Drug-target interaction (DTI) is a crucial aspect of the drug development process. When a drug binds to its target, such as a protein or gene, it alters the target's biological activity, helping to restore normal function. Predicting drug-target interactions is vital in drug discovery, as it can improve efficiency and reduce costs.[9]

DTI prediction often involves four main types of targets: proteins, diseases, genes, and side effects.[10] Discovering new targets for existing or discontinued drugs—a process known as drug repurposing—is another important aspect of drug discovery. With advancements in pharmacology, the 'multi-target, multi-drug' model has gained widespread acceptance, replacing the traditional 'one target, one drug' approach. One key insight is that drugs often target multiple proteins rather than focusing on just one. Therefore, multi-molecular combination drugs are a current trend in drug development. These drugs can work synergistically to enhance the effectiveness of each drug in the combination and help reduce drug resistance, toxicity, and adverse reactions.[11]

In this study, we developed a workflow, illustrated in Figure 1. Based on the previous discussion, the 'multi-target, multi-drug' model is the prevailing approach. Our primary goal is to optimize multi-target treatment strategies by designing drugs that can simultaneously target multiple biological pathways, maximizing therapeutic efficacy. For example, FTD involves multiple biological signaling pathways and pathological mechanisms, so multi-target drugs are better suited to inhibit various aspects of the disease. Additionally, our research aimed to enhance efficacy while minimizing drug resistance. To address drug resistance, we are investigating how simultaneous interference with multiple targets can reduce the likelihood of pathogens or tumor cells developing resistance. The diversity of targets in multi-molecular drugs makes it more difficult for pathogens or cancer cells to evade treatment, thereby improving overall efficacy.

Materials and methods

To construct the core genome-wide genetic and epigenetic networks (GWGENs), we downloaded the microarray data from the GSE140830 dataset (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE140830) available through the National Center for Biotechnology Information. This dataset includes data from FTD patients and healthy control. For data preprocessing, we first divided the original dataset into five categories: gene, transcription factor (TF), receptor, lncRNA, and miRNA, and then proceeded to rank them.

Research ethics

Ethical approval is not applicable due to the use of the publicly available dataset GSE140830 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE140830).

Using this dataset, we constructed the core GWGENs step by step, as shown in Figure 1.

(I) Construction of Candidate GWGENs: The first step involves creating candidate GWGENs, which include the candidate Protein-Protein Interaction Network (PPIN) and the candidate Gene Regulatory Network (GRN) identified through tree-based mining methods.

(II) Identification of Real GWGENs: To remove false positive data, we established all possible regression systems for each node and used the least squares method for system identification. This method applies a system order detection approach to identify the real GWGENs for FTD and healthy control using whole genome microarray data.

(III) Extraction of Core GWGENs: We employed the Principal Network Projection (PNP) method, which uses singular value decomposition to select the top 6,000 nodes with the highest projection values in the GWGENs. These 6,000 nodes, which exhibit the strongest projections on significant singular vector structures accounting for 85% of the real GWGENs, are considered the core GWGENs for further analysis.

(IV) Designing a multi-molecule drug for the treatment of FTD: After identifying the core GWGENs, we annotated them using KEGG pathways to identify core signaling pathways in FTD and healthy control. Based on various research records, we selected significant biomarkers related to critical pathogenic mechanisms as drug targets, i.e., TAU, GSK-3β, STAT3, *ATG5*, WDR41, and RIPK1. Using the deep neural

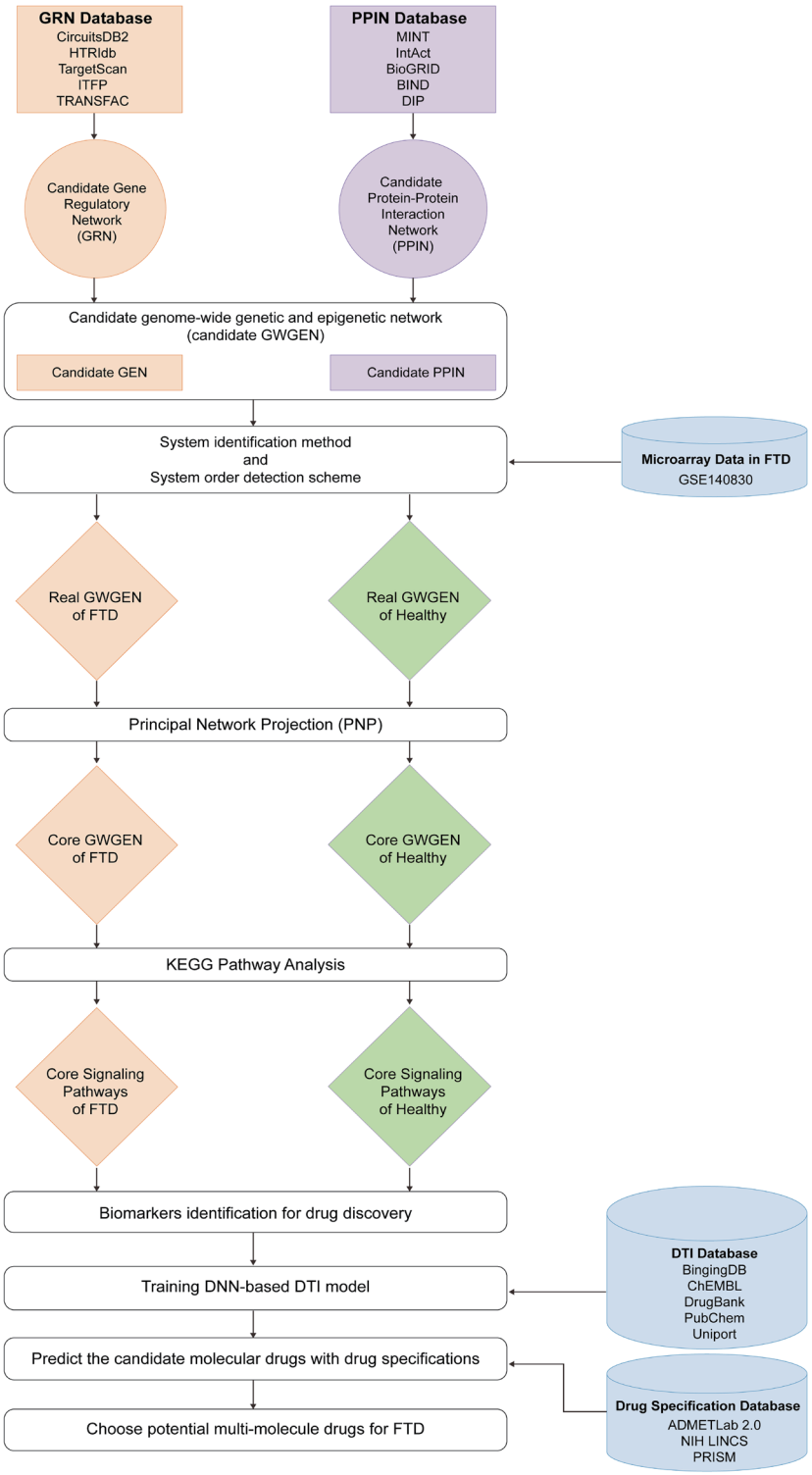

Fig. 1. The flowchart of systems drug discovery for therapeutic treatment of frontotemporal dementia (FTD). (I) Big data mining from gene regulatory network (GRN) and protein-protein interaction network (PPIN) databases to construct candidate genome-wide genetic and epigenetic networks (GWGENs); (II) Microarray data of FTD and healthy control are employed to construct real GWGENs of FTD and healthy control using a system identification method; (III) Principal network projection (PNP) method is used to extract core GWGENs from real GWGENs for Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway annotation to identify the core signaling pathways of FTD and healthy control and to identify significant biomarkers; (IV) A deep neural network (DNN)-based drug-target interaction (DTI) model, trained with DTI databases, is used to predict candidate molecular drugs, which are screened based on drug design specifications to obtain potential multi-molecule drugs for FTD.

Nat Cell Sci 2024;2(4):207–223
https://doi.org/10.61474/ncs.2024.00043

network (DNN)-based DTI model trained on DTI databases, we predicted and screened potential molecular drugs by design specifications to combine them into a multi-molecule drug for treating FTD through targeting these significant biomarkers.

(I) Constructing the candidate GWGEN of FTD and healthy control through big data mining

In this study, the whole-genome microarray dataset with accession number GSE140830 was downloaded from the Gene Expression Omnibus at the National Center for Biotechnology Information. This dataset includes data from 234 blood samples of FTD patients and 248 blood samples of healthy control. Each sample contains expression levels of proteins, receptors, transcription factors, miRNAs, and lncRNAs. These sample data were preprocessed and mined using big data techniques to construct the candidate GWGEN. The candidate GWGEN involves logical and Boolean functions: an interaction between two nodes is recorded as 1, and the absence of interaction is recorded as 0. The candidate GWGEN is divided into the candidate PPIN and the candidate GRN. To construct the candidate PPIN, the following databases were used: MINT, IntAct, BioGRID, BIND, and DIP. To construct the candidate GRN, the following databases were used: CircuitsDB2, HTRIdb, TargetScan, ITFP, and TRANSFAC.

After completing the data preprocessing through the aforementioned databases, we constructed interactive and regulatory models for protein interactions and genetic regulations in the candidate GWGEN. These models included interactions among proteins and regulations among genes, transcription factors, miRNAs, and lncRNAs. The protein interaction and gene regulation models also accounted for random noise caused by model residuals and baseline levels.

For the protein-protein interaction (PPI) model, we constructed the following PPI equations:

$$p_i[n] = \sum_{w=1}^{W_i} \sigma_{iw} p_i[n] p_w[n] + \tau_{i,PPIN} + \varphi_{i,PPIN}[n] \quad \text{for } i = 1, 2, \cdots, I, n = 1, 2, \cdots, N \tag{1}$$

In the protein-protein interaction model, $P_i[n]$ and $P_w[n]$ represent the expression levels of the i-th and w-th proteins in the n-th sample, respectively. The parameter σ_{iw} denotes the interaction strength between the i-th and w-th proteins. $\tau_{i,PPIN}$ represents the baseline expression level of the i-th protein due to unknown interactions caused by histone modifications, such as phosphorylation and acetylation. $\varphi_{i,PPIN}[n]$ signifies the random measurement noise in the expression of the i-th protein in the n-th sample. W_i indicates the total number of interactions with the i-th protein. The letter I stands for the total number of proteins, and N represents the total number of samples.

For the GRN model, we constructed the following genetic regulatory equations:

$$g_j[n] = \sum_{x=1}^{X_j} \varepsilon_{jx} t_x[n] + \sum_{y=1}^{Y_j} \theta_{jy} l_y[n] - \sum_{z=1}^{Z_j} \mu_{jz} m_z[n] g_j[n] + \tau_j + \varphi_j[n] \quad \text{for } j = 1, 2, \cdots, J, n = 1, 2, \cdots, N \tag{2}$$

In the GRN model, the terms $g_j[n]$, $t_x[n]$, $l_y[n]$, and $m_z[n]$ represent the expression levels of the j-th gene, x-th transcription factor, y-th lncRNA, and z-th miRNA in the n-th sample, respectively. The parameter ε_{jx} denotes the regulatory strength from the x-th transcription factor on the j-th gene. The parameter θ_{jy} represents the regulatory strength from the y-th lncRNA on the j-th gene. The parameter μ_{jz} denotes the regulatory strength from the z-th miRNA on the j-th gene, with μ_{jz} being a positive value due to the negative regulatory role of miRNAs on gene expression. The term τ_j signifies the baseline expression level of the j-th gene due to unknown regulations caused by histone modifications such as phosphorylation and acetylation. The term $\varphi_j[n]$ represents the random noise in the measurement of expression of the j-th gene in the n-th sample. The terms X_j, Y_j and Z_j represent the total number of transcription factors, lncRNAs, and miRNAs regulating the j-th gene, respectively. The letter J stands for the total number of genes, and N represents the total number of samples.

For the lncRNA regulatory model, we constructed the following regulatory equations:

$$l_k[n] = \sum_{x=1}^{X_k} \alpha_{kx} t_x[n] + \sum_{y=1}^{Y_k} \beta_{ky} l_y[n] - \sum_{z=1}^{Z_k} \gamma_{kz} m_z[n] l_k[n] + \tau_k + \varphi_k[n] \quad \text{for } k = 1, 2, \cdots, K, n = 1, 2, \cdots, N \tag{3}$$

In the lncRNA regulatory model, the terms $l_k[n]$, $t_x[n]$, $l_y[n]$, and represent the expression levels of the k-th lncRNA, x-th transcription factor, y-th lncRNA, and z-th miRNA in the n-th sample, respectively. The parameter α_{kx} denotes the regulatory strength from the x-th transcription factor on the k-th lncRNA. The parameter β_{ky} represents the regulatory strength from the y-th lncRNA on the k-th lncRNA. The parameter γ_{kz} denotes the regulatory strength from the z-th miRNA on the k-th lncRNA, with γ_{kz} being a positive value due to the negative regulatory role of miRNAs on lncRNA expression. The term τ_k signifies the baseline expression level of the k-th lncRNA due to unknown regulations caused by histone modifications such as phosphorylation and acetylation. The term $\varphi_k[n]$ represents the random noise in the measurement of expression of the k-th lncRNA in the n-th sample. The terms X_k, Y_k and Z_k represent the total number of transcription factors, lncRNAs, and miRNAs regulating the k-th lncRNA, respectively. The letter K stands for the total number of lncRNAs, and N represents the total number of samples.

For the miRNA regulatory model, we constructed the following regulating equations:

$$m_h[n] = \sum_{x=1}^{X_h} \eta_{hx} t_x[n] + \sum_{y=1}^{Y_h} \pi_{hy} l_y[n] - \sum_{z=1}^{Z_h} \xi_{hz} m_z[n] m_h[n] + \tau_h + \varphi_h[n] \quad \text{for } h = 1, 2, \cdots, H, n = 1, 2, \cdots, N \tag{4}$$

In the miRNA regulatory model, the terms $m_h[n]$, $t_x[n]$, $l_y[n]$, and $m_z[n]$ represent the expression levels of the h-th miRNA, x-th transcription factor, y-th lncRNA, and z-th miRNA in the n-th sample, respectively. The parameter η_{hx} denotes the

regulatory strength from the x-th transcription factor on the h-th miRNA. The parameter π_{hy} represents the regulatory strength from the y-th lncRNA on the h-th miRNA. The parameter ξ_{hz} denotes the regulatory strength between the z-th miRNA and the h-th miRNA, with ξ_{hz} being a positive value due to the negative regulatory role of miRNAs on miRNA expression. The term τ_h signifies the baseline expression level of the h-th miRNA due to unknown regulations caused by histone modifications such as phosphorylation and acetylation. The term $\varphi_h[n]$ represents the random noise in the measurement of expression of the h-th miRNA in the n-th sample. The terms X_h, Y_h and Z_h represent the total number of transcription factors, lncRNAs, and miRNAs regulating the h-th miRNA, respectively. The letter H stands for the total number of miRNAs, and N represents the total number of samples.

(II) Constructing the real GWGEN for FTD and healthy control using system identification and system order detection methods

In the previous section, we established four models for the candidate GWGEN, including proteins, genes, lncRNAs, and miRNAs. However, the candidate GWGEN only records whether there is an interaction or regulation between two nodes, while the actual expression levels vary from person to person. Furthermore, false positives may arise due to data mining from large databases. To eliminate false positives in the candidate GWGEN, we constructed the real GWGEN for FTD and the healthy controlby applying system identification and system order detection methods.

To determine the interaction and regulation parameters for the protein interaction and genetic regulatory models, we rewrote Equations (1) to (4) in linear regression form, as shown in Equations (5) to (8).

$$p_i[n] = [p_i[n]p_1[n] \quad \cdots \quad p_i[n]p_{Q_i}[n]1] \times \begin{bmatrix} \sigma_{i1} \\ \vdots \\ \sigma_{iWi} \\ \tau_{i,PPIN} \end{bmatrix} + \varphi_{i,PPIN}[n] \tag{5}$$

$$\triangleq \omega_i[n] \cdot \delta_{i,PPIN} + \varphi_{i,PPIN}[n],$$

$$\text{for i} = 1, 2, \cdots, \text{I, n} = 1, 2, \cdots \text{ N}$$

$$g_j[n] = [t_1[n] \cdots t_{R_j}[n] l_1[n] \cdots l_{S_j}[n]\, m_1[n] g_j[n] \cdots m_{T_j}[n] g_j[n] 1]$$

$$\times \begin{bmatrix} \varepsilon_{j1} \\ \vdots \\ \varepsilon_{jX_j} \\ \theta_{j1} \\ \vdots \\ \theta_{jy_j} \\ \mu_{j1} \\ \vdots \\ \mu_{jZ_j} \\ \tau_j \end{bmatrix} + \varphi_j[n] \tag{6}$$

$$\triangleq \omega_j[n] \cdot \delta_j + \varphi_j[n], \text{ for j} = 1, 2, \cdots, \text{J, n} = 1, 2, \cdots, \text{N}$$

$$l_k[n] = [t_1[n] \cdots t_{R_k}[n] l_1[n] \cdots l_{S_k}[n] m_1[n]\ l_k[n] \cdots m_{T_k}[n] l_k[n] 1]$$

$$\times \begin{bmatrix} \alpha_{k1} \\ \vdots \\ \alpha_{kX_k} \\ \beta_{k1} \\ \vdots \\ \beta_{kY_k} \\ \gamma_{k1} \\ \vdots \\ \gamma_{kZ_k} \\ \tau_k \end{bmatrix} + \varphi_k[n] \tag{7}$$

$$\triangleq \omega_k[n] \cdot \delta_k + \varphi_k[n], \text{ for k} = 1, 2, \cdots, \text{K, n} = 1, 2, \cdots, \text{N}$$

$$m_h[n] = [t_1[n] \cdots t_{R_h}[n] l_1[n] \cdots l_{S_h}[n] m_1[n] m_h[n] \cdots m_{T_h}[n] m_h[n] 1]$$

$$\times \begin{bmatrix} \eta_{h1} \\ \vdots \\ \eta_{hX_h} \\ \pi_{h1} \\ \vdots \\ \pi_{hY_h} \\ \xi_{h1} \\ \vdots \\ \xi_{hZ_h} \\ \tau_h \end{bmatrix} + \varphi_h[n] \tag{8}$$

$$\triangleq \omega_h[n] \cdot \delta_h + \varphi_h[n], \text{ for h} = 1, 2, \cdots, \text{H, n} = 1, 2, \cdots, \text{N}$$

In these linear regression equations, the regression vectors $\omega_i[n]$, $\omega_j[n]$, $\omega_k[n]$, and $\omega_h[n]$ represent the expression levels of the i-th protein, j-th gene, k-th lncRNA, and h-th miRNA in the n-th sample, respectively. The parameter vector $\delta_{i,PPIN}$ denotes the interaction abilities of the i-th protein. The parameter vectors δ_j, δ_k, δ_h denote the regulatory abilities and basal levels of the j-th gene, k-th lncRNA, and h-th miRNA, respectively. The terms $\varphi_{i,PPIN}[n]$, $\varphi_j[n]$, $\varphi_k[n]$, and $\varphi_h[n]$ represent the random noise in the expression of the i-th protein, j-th gene, k-th lncRNA, and h-th miRNA in the n-th sample.

Equations (5) to (8) can be expanded by considering all sample data and rewritten as follows.

$$\begin{bmatrix} p_i[1] \\ p_i[2] \\ \vdots \\ p_i[N] \end{bmatrix} = \begin{bmatrix} \omega_i[1] \\ \omega_i[2] \\ \vdots \\ \omega_i[N] \end{bmatrix} \cdot \delta_{i,PPIN} + \begin{bmatrix} \varphi_{i,PPIN}[1] \\ \varphi_{i,PPIN}[2] \\ \vdots \\ \varphi_{i,PPIN}[N] \end{bmatrix}, \text{ for i} = 1, 2, \cdots, \text{I} \tag{9}$$

$$\begin{bmatrix} g_j[1] \\ g_j[2] \\ \vdots \\ g_j[\text{N}] \end{bmatrix} = \begin{bmatrix} \omega_j[1] \\ \omega_j[2] \\ \vdots \\ \omega_j[N] \end{bmatrix} \cdot \delta_j + \begin{bmatrix} \varphi_j[1] \\ \varphi_j[2] \\ \vdots \\ \varphi_j[N] \end{bmatrix}, \text{ for j} = 1, 2, \cdots, \text{J} \tag{10}$$

Nat Cell Sci 2024;2(4):207–223
https://doi.org/10.61474/ncs.2024.00043

$$\begin{bmatrix} l_k[1] \\ l_k[2] \\ \vdots \\ l_k[\mathrm{N}] \end{bmatrix} = \begin{bmatrix} \omega_k[1] \\ \omega_k[2] \\ \vdots \\ \omega_k[N] \end{bmatrix} \cdot \delta_k + \begin{bmatrix} \varphi_k[1] \\ \varphi_k[2] \\ \vdots \\ \varphi_k[N] \end{bmatrix}, \text{for k} = 1, 2, \cdots, \mathrm{K} \quad (11)$$

$$\begin{bmatrix} m_h[1] \\ m_h[2] \\ \vdots \\ m_h[\mathrm{N}] \end{bmatrix} = \begin{bmatrix} \omega_h[1] \\ \omega_h[2] \\ \vdots \\ \omega_h[N] \end{bmatrix} \cdot \delta_h + \begin{bmatrix} \varphi_h[1] \\ \varphi_h[2] \\ \vdots \\ \varphi_h[N] \end{bmatrix}, \text{for h} = 1, 2, \cdots, \mathrm{H} \quad (12)$$

Equations (9) to (12) can be simplified as follows.

$$P_i = \Omega_i \cdot \Delta_i + \Psi_i, \text{for i} = 1, 2, \cdots, \mathrm{I} \quad (13)$$

$$G_j = \Omega_j \cdot \Delta_j + \Psi_j, \text{for j} = 1, 2, \cdots, \mathrm{J} \quad (14)$$

$$L_k = \Omega_k \cdot \Delta_k + \Psi_k, \text{for k} = 1, 2, \cdots, \mathrm{K} \quad (15)$$

$$M_h = \Omega_h \cdot \Delta_h + \Psi_h, \text{for h} = 1, 2, \cdots, \mathrm{H} \quad (16)$$

Equations (5) to (8) can be further expanded by considering all samples and rewritten as Equations (9) to (12). To avoid overfitting issues in the system identification methods for the random models, the number of elements in the parameters of the PPIN and GRN in Equations (13) to (16) (i.e., Δ_i, Δ_j, Δ_k, and Δ_h cannot exceed half of the dataset's samples (N/2). Therefore, we determine the values of the parameter vectors Δ_i, Δ_j, Δ_k, and Δ_h by solving constrained linear least squares parameter estimation methods to ensure the negative regulatory role of miRNA in Equations (17) to (20).

$$\hat{\Delta}_i = \arg\Delta_i \min \frac{1}{2} \left\| \Omega_i \cdot \Delta_i - P_i \right\|_2^2 \quad (17)$$

$$\hat{\Delta}_j = \arg\Delta_j \min \frac{1}{2} \left\| \Omega_j \cdot \Delta_j - G_j \right\|_2^2$$

$$\text{subject to} \left[\begin{array}{c|c|c|c} 0\cdots0 & 0\cdots0 & 1\cdots0 & 0 \\ \vdots\ddots\vdots & \vdots\ddots\vdots & \vdots\ddots\vdots & \vdots \\ \underbrace{0\cdots0}_{R_j} & \underbrace{0\cdots0}_{S_j} & \underbrace{0\cdots1}_{T_j} & 0 \end{array} \right] \Delta_j \le \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \quad (18)$$

$$\hat{\Delta}_k = \arg\Delta_k \min \frac{1}{2} \left\| \Omega_k \cdot \Delta_k - L_k \right\|_2^2$$

$$\text{subject to} \left[\begin{array}{c|c|c|c} 0\cdots0 & 0\cdots0 & 1\cdots0 & 0 \\ \vdots\ddots\vdots & \vdots\ddots\vdots & \vdots\ddots\vdots & \vdots \\ \underbrace{0\cdots0}_{R_k} & \underbrace{0\cdots0}_{S_k} & \underbrace{0\cdots1}_{T_k} & 0 \end{array} \right] \Delta_k \le \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \quad (19)$$

$$\hat{\Delta}_h = \arg\Delta_h \min \frac{1}{2} \left\| \Omega_h \cdot \Delta_h - M_h \right\|_2^2$$

$$\text{subject to} \left[\begin{array}{c|c|c|c} 0\cdots0 & 0\cdots0 & 1\cdots0 & 0 \\ \vdots\ddots\vdots & \vdots\ddots\vdots & \vdots\ddots\vdots & \vdots \\ \underbrace{0\cdots0}_{Rh} & \underbrace{0\cdots0}_{S_h} & \underbrace{0\cdots1}_{T_h} & 0 \end{array} \right] \Delta_h \le \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \quad (20)$$

Through Equations (17) to (20), we can obtain the optimal solutions for the parameter vectors $\hat{\Delta}_i$, Δ_j, Δ_k, and Δ_h using the least squares method. This part is accomplished with the help of the MATLAB Optimization Toolbox. Additionally, the constraints in Equations (18) to (20) ensure that the transcriptional regulatory effects of miRNA on genes, lncRNA, and miRNA are negative.

After addressing the constrained least square parameter estimation problems from the respective genome-wide microarray data, we obtained the interaction strengths among proteins in the candidate GWGEN for FTD and healthy control, as well as the regulatory abilities for genes, lncRNA, and miRNA. However, due to potential false positive errors in the data from various databases caused by different experimental conditions, we used the Akaike Information Criterion (AIC) to eliminate false positives and perform system order detection for each protein, gene, lncRNA, and miRNA. Therefore, we deleted the false positives from their system orders in the candidate GWGEN to obtain the real GWGEN, as shown in Figure S1.

Based on the system order detection model, we provided four different AIC methods for each protein, gene, lncRNA, and miRNA, respectively, as follows:

$$\mathrm{AIC}(W_i) = \log(\tilde{\Theta}_i^2) + \frac{2(W_i + 1)}{N}$$
$$\text{where } \tilde{\Theta}_i = \sqrt{\frac{(P_i - \Omega_i \cdot \hat{\Delta}_i)^T (P_i - \Omega_i \cdot \Delta_i)}{N}} \quad (21)$$

$$\mathrm{AIC}(X_j, Y_j, Z_j) = \log(\tilde{\Theta}_j^2) + \frac{2(X_j + Y_j + Z_j + 1)}{N}$$
$$\text{where } \tilde{\Theta}_j = \sqrt{\frac{(G_j - \Omega_i \cdot \hat{\Delta}_i)^T (G_j - \Omega_i \cdot \Delta_i)}{N}} \quad (22)$$

$$\mathrm{AIC}(X_k, Y_k, Z_k) = \log(\tilde{\Theta}_k^2) + \frac{2(X_k + Y_k + Z_k + 1)}{N}$$
$$\text{where } \tilde{\Theta}_k = \sqrt{\frac{(L_k - \Omega_k \cdot \hat{\Delta}_k)^T (L_k - \Omega_k \cdot \Delta_k)}{N}} \quad (23)$$

$$\mathrm{AIC}(X_h, Y_h, Z_h) = \log(\tilde{\Theta}_h^2) + \frac{2(X_h + Y_h + Z_h + 1)}{N}$$
$$\text{where } \tilde{\Theta}_h = \sqrt{\frac{(M_h - \Omega_h \cdot \hat{\Delta}_h)^T (M_h - \Omega_h \cdot \Delta_h)}{N}} \quad (24)$$

The parameters $\tilde{\Theta}_i$, $\tilde{\Theta}_j$, $\tilde{\Theta}_k$, and $\tilde{\Theta}_h$ represent the estimated residuals between $\hat{\Delta}_i$, Δ_j, Δ_k, and Δ_h and the actual expression levels of the i-th protein, j-th gene, k-th lncRNA,

Table 1. The number of nodes in candidate GWGEN and real GWGEN of FTD and healthy control

Nodes	Candidate GWGEN	Real GWGEN of FTD	Real GWGEN of healthy control
Receptor	1,859	1,859	1,859
TF	1,132	1,132	1,132
Protein	11,775	11,771	11,774
miRNA	150	150	150
LncRNA	196	187	189
Total	15,112	15,099	15,104

FTD, Frontotemporal dementia; GWGEN, genome-wide genetic and epigenetic networks; TF, transcription factor.

and h-th miRNA, respectively.

The AIC method is a statistical measure used for model order selection. A lower AIC value indicates a better system model order, suggesting that the system model has achieved a good fit with fewer parameters. Therefore, we minimize the four system order detection models for each protein, gene, lncRNA, and miRNA in the candidate GWGEN using the following AIC Equations (25) to (28):

$$W_i^* = \arg Q_i \min AIC(W_i) \quad \text{for } i = 1, 2, \cdots, I \tag{25}$$

$$X_j^*, Y_j^*, Z_j^* = \arg X_j, Y_j, Z_j \min AIC(X_j, Y_j, Z_j) \quad \text{for } j = 1, 2, \cdots, J \tag{26}$$

$$X_k^*, Y_k^*, Z_k^* = \arg X_k, Y_k, Z_k \min AIC(X_k, Y_k, Z_k) \quad \text{for } k = 1, 2, \cdots, K \tag{27}$$

$$X_h^*, Y_h^*, Z_h^* = \arg X_h, Y_h, Z_h \min AIC(X_h, Y_h, Z_h) \quad \text{for } h = 1, 2, \cdots, H \tag{28}$$

where W_i^* represents the actual total number of interactions with the i-th protein. X_j^*, Y_j^*, Z_j^* represent the actual total numbers of TFs, lncRNAs, and miRNAs interacting with the j-th gene. X_k^*, Y_k^*, Z_k^* represent the actual total numbers of TFs, lncRNAs, and miRNAs interacting with the k-th lncRNA. X_h^*, Y_h^*, Z_h^* represent the actual total numbers of TFs, lncRNAs, and miRNAs interacting with the h-th miRNA. Using the AIC method, we successfully eliminated false positives in the PPI and GRN out of the actual interactions and regulations (i.e., their real system order), identifying the real GWGENs of FTD and healthy control from their genome-wide microarray data, as shown in Tables 1 and 2.

(III) Extraction of the core GWGEN by PNP method

After applying the system identification and system order detection methods, we obtained the real GWGEN for FTD and healthy control. However, the real GWGEN is still too complex to study directly. Additionally, to understand the pathogenic mechanisms of FTD, we need to use KEGG pathways for annotation. However, KEGG pathways can currently annotate only GWGENs with up to 6,000 nodes. Therefore, we use the PNP method to extract the top 6,000 nodes from the real GWGEN to form the core GWGEN for both FTD and healthy control. The PNP method involves performing singular value decomposition (SVD) on the real GWGEN. To

Table 2. The number of edges in candidate GWGEN and real GWGEN of FTD and healthy control

Nodes	Candidate GWGEN	Real GWGEN of FTD	Real GWGEN of healthy control
PPIs	3,134,515	1,854,695	1,891,770
TF-Receptor	9,728	2,361	2,103
TF-TF	7,900	1,778	1,639
TF-Protein	57,586	14,085	12,594
TF-miRNA	450	82	77
TF-LncRNA	273	124	128
miRNA-Receptor	6,718	1,307	1,214
miRNA-TF	5,763	1,168	1,057
miRNA-Protein	40,346	8,141	7,764
miRNA-miRNA	4	3	4
miRNA-LncRNA	149	44	36
LncRNA-Receptor	163	34	47
LncRNA-TF	161	34	45
LncRNA-Protein	1,299	368	423
LncRNA-miRNA	0	0	0
LncRNA-LncRNA	3	0	1
Total	3,265,058	1,884,224	1,918,902

FTD, Frontotemporal dementia; GWGEN, genome-wide genetic and epigenetic networks; TF, transcription factor.

Nat Cell Sci 2024;2(4):207–223
https://doi.org/10.61474/ncs.2024.00043

begin, we first construct a composite network matrix A for the real GWGEN, which includes all the estimated parameters of the real GWGEN, as follows:(29)

$$A=\begin{bmatrix} a_{protein\leftrightarrow protein} & 0 & 0 \\ a_{TF\rightarrow gene} & a_{\ln cRNA\rightarrow gene} & a_{miRNA\rightarrow gene} \\ a_{TF\rightarrow \ln cRNA} & a_{\ln cRNA\rightarrow \ln cRNA} & a_{miRNA\rightarrow \ln cRNA} \\ a_{TF\rightarrow miRNA} & a_{\ln cRNA\rightarrow miRNA} & a_{miRNA\rightarrow miRNA} \end{bmatrix} \quad (29)$$

where A is divided into several submatrices. The submatrix $a_{protein\leftrightarrow protein}$ represents the system matrix of all protein-protein interactions, and $a_{TF\rightarrow gene}$, $a_{lncRNA\rightarrow gene}$, and $a_{miRNA\rightarrow gene}$ represent the system matrices of the transcriptional regulatory capabilities of TFs, lncRNAs, and miRNAs on all genes, respectively. The submatrices $a_{TF\rightarrow lncRNA}$, $a_{lncRNA\rightarrow lncRNA}$, and $a_{miRNA\rightarrow lncRNA}$ represent the system matrices of the transcriptional regulatory capabilities of TFs, lncRNAs, and miRNAs on all lncRNAs, respectively. The submatrices $a_{TF\rightarrow miRNA}$, $a_{lncRNA\rightarrow miRNA}$, and $a_{miRNA\rightarrow miRNA}$ represent the system matrices of the transcriptional regulatory capabilities of TFs, lncRNAs, and miRNAs on all miRNAs, respectively. The double-headed arrows indicate that protein-protein interactions are bidirectional, while the single-headed arrows indicate that transcriptional regulatory capabilities are unidirectional.

The following network matrix A is the expanded form of these submatrices:

$$A=\begin{bmatrix} \hat{\sigma}_{11} & \cdots & \sigma_{1w} & \cdots & \sigma_{1w_j} & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \hat{\sigma}_{i1} & \cdots & \sigma_{iw} & \cdots & \sigma_{iw_j} & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \hat{\sigma}_{I1} & \cdots & \sigma_{Iw} & \cdots & \sigma_{Iw_j} & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & 0 \\ \hat{\varepsilon}_{11} & \cdots & \varepsilon_{1x} & \cdots & \varepsilon_{1X_j} & \hat{\theta}_{11} & \cdots & \theta_{1y} & \cdots & \theta_{1Y_j} & \hat{\mu}_{11} & \cdots & \mu_{1z} & \cdots & \mu_{1Z_j} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \hat{\varepsilon}_{j1} & \cdots & \varepsilon_{jx} & \cdots & \varepsilon_{jX_j} & \hat{\theta}_{j1} & \cdots & \theta_{jy} & \cdots & \theta_{jY_j} & \hat{\mu}_{j1} & \cdots & \hat{\mu}_{jz} & \cdots & \mu_{jZ_j} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \hat{\varepsilon}_{J1} & \cdots & \varepsilon_{Jx} & \cdots & \varepsilon_{JX_j} & \hat{\theta}_{J1} & \cdots & \theta_{Jy} & \cdots & \theta_{JY_j} & \hat{\mu}_{J1} & \cdots & \mu_{Jz} & \cdots & \mu_{JZj} \\ \ddots & & & & & & & & & & & & & & \\ \hat{\alpha}_{11} & \cdots & \alpha_{1x} & \cdots & \alpha_{1X_k} & \hat{\beta}_{11} & \cdots & \beta_{1y} & \cdots & \beta_{1Y_k} & \gamma_{11} & \cdots & \gamma_{1z} & \cdots & \gamma_{1Z_k} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \hat{\alpha}_{k1} & \cdots & \alpha_{kx} & \cdots & \alpha_{kX_k} & \hat{\beta}_{k1} & \cdots & \beta_{ky} & \cdots & \beta_{kY_k} & \gamma_{k1} & \cdots & \gamma_{kz} & \cdots & \gamma_{kZ_k} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \hat{\alpha}_{K1} & \cdots & \alpha_{Kx} & \cdots & \hat{\alpha}_{KX_k} & \hat{\beta}_{K1} & \cdots & \beta_{Ky} & \cdots & \beta_{KY_k} & \gamma_{K1} & \cdots & \gamma_{Kz} & \cdots & \gamma_{KZ_k} \\ \hat{\eta}_{11} & \cdots & \eta_{1x} & \cdots & \eta_{1X_k} & \hat{\pi}_{11} & \cdots & \pi_{1y} & \cdots & \pi_{1Y_k} & \hat{\xi}_{11} & \cdots & \xi_{1z} & \cdots & \xi_{1Z_k} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \hat{\eta}_{h1} & \cdots & \eta_{hx} & \cdots & \eta_{hX_k} & \hat{\pi}_{h1} & \cdots & \pi_{hy} & \cdots & \pi_{hY_k} & \hat{\xi}_{h1} & \cdots & \xi_{hz} & \cdots & \xi_{hZ_k} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \hat{\eta}_{H1} & \cdots & \eta_{Hx} & \cdots & \eta_{HX_k} & \hat{\pi}_{H1} & \cdots & \pi_{Hy} & \cdots & \pi_{HY_k} & \hat{\xi}_{H1} & \cdots & \xi_{Hz} & \cdots & \xi_{HZ_k} \end{bmatrix} \quad (30)$$

$$\in \mathbb{R}^{(I^*+J^*+K^*+H^*)\times(X^*+Y^*+Z^*)}$$

Each row in A represents the interaction abilities of each protein with other proteins, or the regulatory abilities of each gene, lncRNA, and miRNA by TFs, lncRNAs, and miRNAs. Next, we perform SVD on the matrix A as follows:

$$\begin{aligned} &\mathrm{A}=U\Sigma V^T \\ &U\in\mathbb{R}^{(I^*+J^*+K^*+H^*)\times(I^*+J^*+K^*+H^*)} \\ &\Sigma\in\mathbb{R}^{(I^*+J^*+K^*+H^*)\times(X^*+Y^*+Z^*)} \\ &V\in\mathbb{R}^{(X^*+Y^*+Z^*)\times(X^*+Y^*+Z^*)} \end{aligned} \quad (31)$$

$$\Sigma=\begin{bmatrix} \sigma_1 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_r & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & \sigma_{X^*+Y^*+Z^*} \\ 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \end{bmatrix} \in\mathbb{R}^{(I^*+J^*+K^*+H^*)\times(X^*+Y^*+Z^*)} \quad (32)$$

Based on the SVD calculation, σ_r represents the r-th singular value, and $\sigma_{X^*+Y^*+Z^*}$ represents the last singular value. We select the top R singular values, which together account for at least 85% of all singular values, i.e., at least 85% of the network matrix in (30) from the network energy perspective.

$$E_r=\frac{\sum_{r=1}^{R}\sigma_r^2}{\sum_{r=1}^{X^*+Y^*+Z^*}\sigma_r^2}\geq 0.85 \quad (33)$$

In Equation (33), E_r represents the proportion of the total network energy accounted for by the top r singular values. Based on this definition, we can determine the top R singular values that account for at least 85% of the total network energy.

We then project each node (i.e., w_j of each row in A) of the composite network matrix A onto the top R singular vectors (i.e., v_i, i = 1, …, R):

$$\begin{aligned} &P(w_j)=\sqrt{\sum_{i=1}^{R}proj(v_i,w_j)} \\ &\text{where } proj(v_i,w_j)=v_i\cdot {w_j}^T \\ &\text{for i}=1,\cdots,\text{R, j}=1,\cdots,(I^*+J^*+K^*+H^*) \end{aligned} \quad (34)$$

where w_j represents the j-th row of matrix A (i.e., the j-th node of the real GWGEN), v_i represents the i-th singular value vector of V. $proj(v_i,w_j)$ denotes the projection value of the j-th node of matrix A onto the i-th singular value vector. We then calculate the 2-norm projection value $P(w_j)$ for each row j of matrix A. A higher $P(w_j)$ indicates a greater influence of the corresponding j-th row (or j-th node of the real GWGEN) on the principal network structure.

Using the PNP method, we extract the top 6,000 nodes with the highest $P(w_j)$ values from the real GWGENs of both FTD and healthy control to form the core GWGENs for both conditions, as shown in Figure S2. This is the maximum number of nodes that KEGG pathways can annotate. After KEGG annotation, we can identify the core signaling pathways for FTD and healthy control. Based on the signal transmission paths of the core signaling pathways of FTD, along with their downstream target genes and cellular dysfunctions, we investigate the pathogenic mechanisms of FTD and select the most suitable genes or proteins as significant biomarkers for FTD pathogens and drug targets for FTD treatment.

(IV) Designing a multi-molecule drug for the treatment of FTD using a DNN-based DTI model and drug design specifications

After identifying the significant biomarkers of the pathogenic

mechanism of FTD, we use these biomarkers as drug targets. Next, we train a DNN-based DTI model using DTI databases to predict the interaction probabilities between drugs and targets.

First, to train the DNN-based DTI model, we integrated multiple drug-target interaction databases, including BindingDB, ChEMBL, DrugBank, PubChem, and UniProt. These databases provide information on the features of drugs and their targets, as well as the interactions between molecules. Drug features include molecular properties such as structure, topology, and geometric descriptors. Target features are described based on the physicochemical and structural properties of proteins and peptides in amino acid sequences. Using Python's PyBioMed package, we convert the drug and target features into feature vector representations. The expression of the converted drug-target feature vectors is shown as follows:

$$u_{drug\text{-}target} = [D,T] = [d_1,\cdots d_p,\cdots,d_P,t_1,\cdots,t_q\cdots,t_Q] \quad (35)$$

where $u_{drug\text{-}target}$ represents the drug-target feature vector, D and T represent the feature vectors of the drug and the target, respectively. The parameters d_p and t_q represent the p-th drug feature and the q-th target feature, respectively. P and Q represent the total number of features for the drug and the target, respectively.

Before using the drug-target feature vectors as training data for the DNN-based DTI model, we preprocess the feature vectors to avoid potential bias issues in the model. Specifically, the data of unverified drug-target interactions (negative class) is much larger than that of confirmed drug-target interactions (positive class). To address this, we randomly sample the unverified drug-target interactions to equalize the sample size with that of the confirmed interactions. Additionally, because the feature vector variables for drugs and targets use different units, we standardize them as follows:

$$d_p^* = \frac{d_p - \mu_p}{\sigma_p}, \forall p = 1,\cdots,P \quad (36)$$

$$t_q^* = \frac{t_q - \mu_q}{\sigma_q}, \forall q = 1,\cdots,Q \quad (37)$$

where d_p* and t_q* represent the standardized p-th drug feature and the q-th target feature, respectively. μ_p and μ_q represent the mean values of the drug features and target features, respectively. σ_p and σ_q represent the standard deviations of the drug features and target features, respectively.

Since the input layer of the DNN network used in the DTI model has only 996 nodes, while the standardized feature vectors in Equation (35) still exceed this number, we use principal component analysis to reduce the dimensionality of the drug-target feature vectors. After reducing the dimensionality of the drug-target feature vectors to 996, we use 75% of the drug-target feature vector data as training data and the remaining 25% as testing data. We then use Python's TensorFlow and Keras libraries for training and prediction. The architecture of the DNN includes four hidden layers, both in the input and output layers, with the hidden layers employing the rectified linear unit activation function. Dropout layers are added to each hidden layer to avoid overfitting. The output layer uses a sigmoid activation function to constrain the output between 0 and 1, representing the probability of interaction between the drug and the target. The neural network parameters are set with a learning rate of 0.001, 100 epochs, and a batch size of 100. Additionally, the Adam optimization algorithm is used for training the neural network. Since drug-target interaction is a binary classification problem (interaction or no interaction), we use binary cross-entropy as our loss function:

$$C_n(p_n,\hat{p}_n) = -[p_n \log p_n + (1-p_n)\log(1-p_n)] \quad (38)$$

$$L(w,b) = \frac{1}{N}\sum_{n=1}^{N} C_n(w,b) \quad (39)$$

for the n-th sample, the true interaction probability p_n and the predicted interaction probability $\hat{p}_n$ can be obtained from the above expression. The terms $(1-p_n)$ and $(1-\hat{p}_n)$ represent the true and predicted probabilities of no interaction, respectively. $C_n(p_n, \hat{p}_n)$ represents the loss function for the n-th sample. $L(w, b)$ denotes the average loss across all samples, where w represents the weight vector, b represents the bias vector of the DNN, and N represents the total number of training samples.

To optimize the weight vector w and bias vector b, we combine them into a vector θ and use the backpropagation algorithm to compute the gradient and obtain the optimal model parameter set θ*. The advantage of the backpropagation algorithm is that it can efficiently compute high-dimensional vectors and adjust the DTI model parameters to fit the drug-target interaction data for each iteration. The gradient iteration algorithm is as follows:

$$\theta = \begin{bmatrix} w \\ b \end{bmatrix} \quad (40)$$

$$\theta^* = \arg\min L(\theta) \quad (41)$$

$$\theta^y = \theta^{y-1} - \eta\nabla L(\theta^{y-1})$$
$$\text{where } \nabla L(\theta^{y-1}) = \begin{bmatrix} \frac{\partial L(\theta^{y-1})}{\partial w} \\ \frac{\partial L(\theta^{y-1})}{\partial b} \end{bmatrix} \quad (42)$$

where y represents the y-th iteration of the DNN training process, η represents the learning rate, and $\nabla L(\theta^{y-1})$ represents the gradient of the cost function $L(\theta^{y-1})$.

To evaluate the performance of our trained DTI model, we use the five-fold cross-validation method. We divide the training data into five equal parts and, in each iteration, use one part as the validation data and the remaining four parts as the training data. We average the evaluation results of the five validations to obtain the final performance metric of the model, as shown in Figures S3 and S4. Additionally, we use the area under the curve (AUC) of the receiver operating characteristic (ROC) curve as another reference metric. For binary classification problems, AUC is an important indicator for evaluating the performance of the model visually. The larger the AUC, the better the performance of the DNN-

Table 3. Core signaling pathways of FTD through annotation of KEGG pathways

Pathway	Gene number	*p*-value
MAPK signaling pathway	148	2.4E-10
WNT signaling pathway	76	1.1E-3
JAK-STAT signaling pathway	72	2.6E-3
PI3K-Akt signaling pathway	135	1.9E-2
Amyotrophic lateral sclerosis	166	1.3E-7
TNF signaling pathway	69	8.8E-9

KEGG, Kyoto Encyclopedia of Genes and Genomes; FTD, Frontotemporal dementia; TNF, tumor necrosis factor.

based DTI model. The AUC equations for the ROC curve are given as follows:

$$\text{TPR(True Positive Rate)} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{43}$$

$$\text{TNR(True Negative Rate)} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{44}$$

$$\text{FPR(False Positive Rate)} = \frac{\text{FP}}{\text{TN+FP}} = 1 - \text{TNR} \tag{45}$$

$$\text{FNR(False Negative Rate)} = \frac{\text{FN}}{\text{TP} + \text{FN}} = 1 - \text{TPR} \tag{46}$$

where TP represents the number of true positives, where the actual value is positive and the predicted result is also positive. TN represents the number of true negatives, where the actual value is negative and the predicted result is also negative. FP represents the number of false positives, where the actual value is negative but the predicted result is positive. FN represents the number of false negatives, where the actual value is positive but the predicted result is negative. The ROC curve is a graphical representation where the TPR is plotted on the vertical axis and the FPR is plotted on the horizontal axis, as shown in Figure S5. The AUC of the ROC curve indicates the performance of the model. An AUC value closer to 1 signifies better performance and a lower probability of false positives.

After predicting the interaction probabilities between drugs and targets, we also need to consider drug design specifications to further refine the candidate drugs and select a suitable multi-molecular drug for treating FTD. We use three drug design specifications for screening: regulatory capacity, sensitivity, and toxicity, as shown in Table S1. For regulatory capacity, we used the LINCS L1000 Level 5 database, where a regulatory ability >0 indicates upregulation of expression levels, and <0 indicates downregulation. Sensitivity was assessed using the PRISM repurposing dataset, representing the compound's interference with human cells, with values closer to 0 indicating lower interference. Toxicity was evaluated using the ADMETlab 2.0 tool, where the standardized value LC50 indicates toxicity, with higher values representing lower toxicity to the human body. Based on strong regulatory capacity, high sensitivity, and low toxicity, we propose potential molecular drug combinations as candidates for multi-molecular drugs to treat FTD.[12]

Table 4. Core signaling pathways of healthy control through annotation of KEGG pathways

Pathway	Gene number	*p*-value
Cell Cycle	86	1.2E-9
PI3K-Akt signaling pathway	141	7.0E-4
Amyotrophic lateral sclerosis	148	1.4E-4
TNF signaling pathway	56	2.7E-4
Apoptosis	65	4.6E-5
Nucleocytoplasmic transport	60	2.0E-7

TNF, tumor necrosis factor.

The other detailed methods are provided in the Supplementary materials.

Results

Investigation of core signaling pathways using systems biology methods and prediction of candidate drugs using a trained DNN-based DTI model by DTI databases

Following KEGG pathway annotation, the core signaling pathways for FTD (as presented in Table 3) and healthy controls (as presented in Table 4) are illustrated in Figure 2. In the following subsection, we will analyze cytokines in the microenvironment, core signaling pathways, their downstream target genes, and associated cellular dysfunctions to explore the pathogenic mechanisms of FTD. We will then select biomarkers that play a key role in the pathogenesis of FTD, such as TAU, GSK-3β, STAT3, *ATG5*, WDR41, and RIPK1. By applying the DNN-based DTI model, trained on the DTI database, we predict potential drugs and select candidate molecular drugs based on design specifications to formulate a multi-molecule drug aimed at treating FTD by targeting these significant biomarkers.

The role of biomarker TAU in the MAPK signaling pathway

The growth factor (GF) family comprises proteins or peptides that control diverse cellular functions, including cell growth, differentiation, and survival. They play crucial roles in processes like development, tissue repair, immune response, and intercellular communication.[13] Receptor tyrosine kinases serve as receptors for GFs and are essential for neuronal function and development, as shown in Figure 2. For instance, neurotrophins and other GFs, which are expressed in very limited amounts, play a vital role in regulating neuronal development, plasticity, and survival.[14] Receptor tyrosine kinases, upon receiving signals from GFs, activate downstream signaling pathways, such as the GRB2/SOS/Ras signaling pathway. The activation of Ras is essential for processes such as cell proliferation, differentiation, and apoptosis.[15] RafA is a serine protein kinase that, upon receiving signals from Ras, can directly phosphorylate proteins or activate downstream MEK/ERK pathways to promote protein phosphorylation and regulate cell apoptosis.[16] ERK, a member of the MAPK family, is responsible for phosphorylating the transcription factor ELK1,

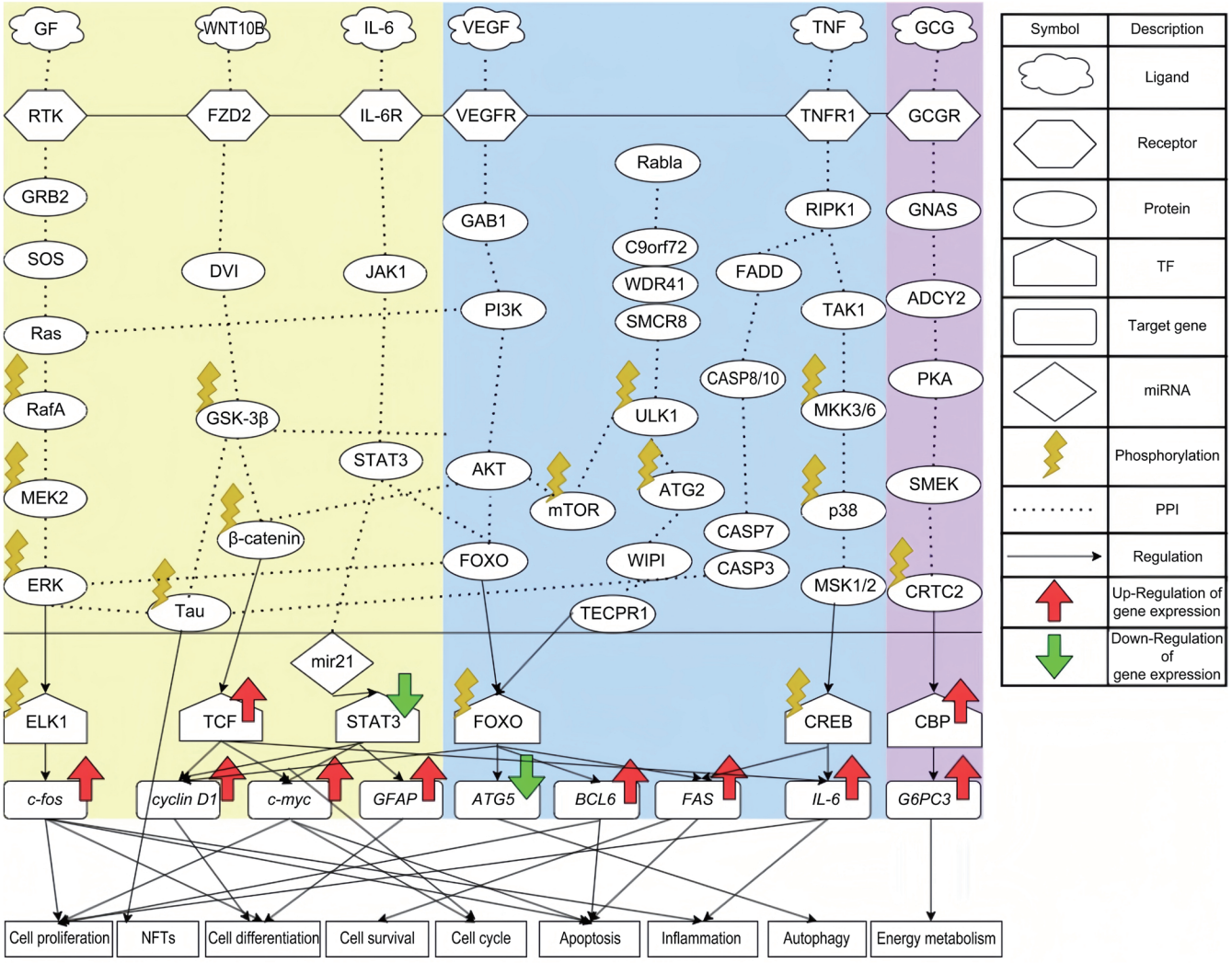

Fig. 2. Core signaling pathways of frontotemporal dementia (FTD) and healthy control. The left area of this picture presents the core signaling pathways of FTD. The middle area shows the common core signaling pathways between FTD and healthy control. The right area presents the core signaling pathways of healthy control. The green and red arrows represent abnormal downregulation and upregulation of selected significant biomarkers of FTD, respectively. TF, Transcription Factor; PPI, Protein-Protein Interaction; NFT, Neurofibrillary Tangle.

which upregulates *c-fos*. The ERK/ELK1/*c-fos* pathway is involved in inflammatory responses, cell differentiation, cell proliferation, and apoptosis.[17] Additionally, the persistent activation of ERK induces the phosphorylation of TAU protein. Various studies have indicated that TAU phosphorylation causes neurofibrillary tangles (NFTs),[18] one of the most common initial symptoms of FTD.

The role of biomarker GSK-3β in the WNT signaling pathway

Research has shown that the WNT signaling pathway can lead to mutations in progranulin, which are a significant pathogenic mechanism in FTD,[19] as shown in Figure 2. FZD2, serving as a receptor in the WNT signaling pathway, plays a crucial role in this process. A decrease in FZD2 levels results in increased cell apoptosis, while its upregulation promotes the survival of neurons *in vitro*.[20] After receiving the WNT signal, FZD2 activates GSK-3β, an important kinase for TAU. As discussed in the previous section, the phosphorylation of TAU plays a potential pathogenic role in FTD. In addition to causing TAU phosphorylation, GSK-3β is essential in the WNT pathway. It phosphorylates downstream β-catenin, leading to its degradation through the ubiquitin-proteasome pathway. β-catenin serves as an activator of T-cell factor-dependent transcription, resulting in the upregulation of various target genes, such as c-myc and cyclin D1.[21] The expression of c-myc is strongly linked to cell cycle progression and can also trigger apoptosis. As the cell cycle initiates, cyclin D1 moves through the entire cycle,[22] regulating cell proliferation and participating in cell differentiation.

The role of biomarker STAT3 in the JAK-STAT signaling pathway

Interleukin-6 (IL-6) is produced by keratinocytes and white blood cells, and since its discovery, the IL-6 signaling pathway has become a core pathway involved in healthy immune

regulation and immune dysregulation in many diseases. Research has found that elevated levels of IL-6 can contribute to granulin precursor mutations, which are also a key pathogenic mechanism of FTD.[23] IL-6R receives IL-6 and activates JAK1 and STAT3, as shown in Figure 2. The JAK/STAT signaling pathway coordinates adaptive and innate immune mechanisms, ultimately limiting neuroinflammatory responses and acting as a key contributor to neuroinflammation in neurodegenerative diseases.[24] The downstream effector miR-21 of STAT3 is a miRNA associated with dysfunction in neuron-glial cell function. Some studies have shown it to be overexpressed in neurons derived from induced pluripotent stem cells of FTD patients with the PSEN1ΔE9 deletion (iNEU-PSEN).[25] Additionally, miR-21 is linked to the regulation of toxicity caused by amyloid-beta (Aβ) oligomers, which is generally considered one of the pathological mechanisms underlying neurodegenerative diseases in the brain. Finally, miR-21 activates the transcription factor STAT3, which upregulates many target genes, including *c-myc*, *cyclin D1*, and *GFAP*. The functions of *c-myc* and *cyclin D1* were introduced in previous sections. As for *GFAP*, it is a glial fibrillary acidic protein found in astrocytes in the central nervous system, where it plays a role in cell differentiation. Several studies have indicated that serum GFAP levels are significantly higher in FTD patients than in healthy controls.[26]

The role of biomarker ATG5 in the phosphoinositide 3-kinase (PI3K)-Akt signaling pathway

Vascular endothelial growth factor (VEGF) is involved in neurodevelopment, angiogenesis, and hematopoiesis, playing an essential role in maintaining homeostasis in the adult vascular system. A study reported that elevated levels of VEGF can cause hippocampal atrophy. Over time, hippocampal atrophy can lead to cognitive decline, which may contribute to the development of FTD,[27] as shown in Figure 2. VEGFR receives the VEGF signal and transmits it through GRB2-associated binding protein 1 to the downstream PI3K/AKT pathway.[28] The PI3K and protein kinase B (AKT) signaling pathway play roles in many important cellular functions. In the brain, the PI3K/AKT signaling pathway serves various functions, including regulating survival, cell proliferation, growth, differentiation, and other complex processes. It also plays a role in oxidative stress and autophagy during neuroinflammation.[29] The PI3K/AKT pathway phosphorylates downstream GSK-3β, which, as mentioned earlier, contributes to neurofibrillary tangle formation. In addition to its impact on GSK-3β, AKT can also phosphorylate downstream proteins mTOR and FOXO. Phosphorylation of FOXO by AKT inhibits FOXO's transcriptional function, promoting cell survival and proliferation. The transcription factor FOXO targets many genes, including *cyclin D1*, *ATG5*, *BCL6*, and *FAS*. FOXO induces apoptosis by upregulating mitochondrial-targeting proteins of the Bcl family.[30] *ATG5* is a gene involved in the autophagy process and also participates in regulating cell survival and metabolic balance, which is crucial for maintaining cellular function. However, FOXO downregulates *ATG5*, leading to changes in the autophagy process. A lack of autophagy can impair learning and memory, which is one of the important symptoms in patients with FTD.[31]

The role of biomarker WDR41 in the amyotrophic lateral sclerosis signaling pathway

The GGGGCC hexanucleotide repeat expansion in the C9orf72 gene is the leading genetic cause of amyotrophic lateral sclerosis and FTD.[32] In related studies, we have found that C9orf72, along with SMCR8 and WDR41, forms a stable complex through their interaction, which is involved in the regulation of macroautophagy. The C9orf72-SMCR8-WDR41 complex interacts with the autophagy initiation complex involving Rab1a and Unc-51-like kinase 1 (ULK1).[33] As an effector of Rab1a, the C9orf72-SMCR8-WDR41 complex regulates the initiation of autophagy by controlling Rab1a-dependent transport of the ULK1 autophagy initiation complex to phagophores. Within the C9orf72-SMCR8-WDR41 complex, WDR41 is a prominent C9orf72-interacting protein, playing a significant role in supporting the regulatory association of C9orf72 with lysosomes. In the complex, SMCR8 acts as an upstream component of ULK1. The interaction between ULK1 and mTOR is essential for autophagic function. After the fusion of the ULK1 autophagosome with the lysosome, mTOR can be reactivated. The activation of mTOR subsequently reduces ULK1 kinase activity by phosphorylating it at the Ser757 site, thereby suppressing autophagy.[34] The phosphorylation of ULK1 also leads to the phosphorylation of its downstream protein Atg2. Atg2, as a key protein in membrane expansion during the ULK1-initiated autophagy process, ensures the formation of autophagosomes.[35] By binding to WIPI proteins, Atg2 is localized and stabilized on the autophagosome membrane, promoting the transport of membrane lipids and membrane expansion.[36] Another role of WIPI is to activate TECPR1, which transmits signals to the transcription factor FOXO to downregulate *ATG5*. This downregulation results in the impairment of autophagic function, contributing to neurodegenerative diseases.

The role of biomarker RIPK1 in the tumor necrosis factor (TNF) signaling pathway.

The primary pro-inflammatory cytokine TNF has been demonstrated to regulate various signaling pathways, leading to a broad range of downstream effects.[37] These effects encompass the regulation of cell proliferation, differentiation, apoptosis, immune response, and the induction of inflammation. Due to such extensive cellular effects and complex signaling pathways, TNF is also associated with many age-related disease states. Upon receiving the TNF signal, the receptor TNFR1 begins to activate downstream pathways.[37] The RIPK1 is recognized as a key regulator of TNFR1 signal transduction. RIPK1 controls cell fate decisions, promoting either cell survival or death. The downstream FADD/CASP signaling pathway is responsible for regulating extrinsic apoptosis and necroptosis.[38] Caspase-3 (CASP3) in the CASP family can cleave TAU protein, leading to its phosphorylation and the formation of NFTs.[39] In addition to regulating cellular apoptosis, RIPK1 is also involved in mediating inflammatory responses in neurodegenerative diseases. The downstream TAK1/MKK/p38 signaling pathway plays an essential role in inflammation, as activation of p38 induces the expression of inflammatory mediators involved in tissue remodeling and oxidative regulation.[40] Downstream of p38, mitogen- and stress-activated protein kinases 1 and 2 act as epigenetic modifiers that activate genes related to cell proliferation, inflammation, and neuronal function, as well as phosphorylate the transcription factor CREB.[41] CREB has a vital role in the nervous system, participating in the formation of learning and memory, and is responsible for upregulating target gene

IL-6. *IL-6* is important for neuronal development, differentiation, and regeneration; therefore, its dysregulation is associated with neuroinflammation to some degree.[42]

The core signaling pathways of healthy control

In the healthy control tissues, we observed that the glucagon signaling pathway plays a crucial role. However, recent studies indicate that, besides its role in glucose metabolism, the glucagon signaling pathway also protects the nervous system by regulating neuronal metabolism, antioxidant stress response, and inflammation.[43] Upon binding of pancreatic glucagon and its receptor in the microenvironment, the GNAS/ADCY2/PKA/SMEK signaling pathway is activated, as shown in Figure 2. The actions of these kinases are also linked to metabolic regulation. Overexpression of SMEK, the downstream component, leads to phosphorylation of CRTC2.[44] Phosphorylated CRTC2 then interacts with CREB-binding protein, regulating the expression of the target gene *G6PC3*. The main function of *G6PC3* is to maintain energy homeostasis and mitochondrial function. In this study, we observed that the expression level of the G6PC3 gene is significantly elevated in normal tissues compared to those in FTD patients, which may explain the metabolic abnormalities in FTD patients due to *G6PC3* deficiency.[45]

Predicting potential drugs for treating FTD using biomarkers as drug targets and leveraging a deep neural network-based drug-target interaction model

After investigating the core signaling pathways involved in the pathogenic mechanism of FTD and identifying significant biomarkers TAU, GSK-3β, STAT3, *ATG5*, WDR41, and RIPK1 as drug targets, we began studying the interactions between these biomarkers and drugs, considering drug design specifications such as regulatory ability, sensitivity, and toxicity. Based on these significant drug properties, we selected potential drugs expected to reverse the expression levels of these biomarkers without causing excessive side effects. To study the interactions between biomarkers and drugs, we developed a DNN-DTI model, as shown in Figure 3. This model was pre-trained using the DTI database, enabling it to effectively predict the interaction probabilities between biomarkers and candidate drugs after DTI data training via the Adam learning algorithm.

However, in the DTI database, there were 80,291 confirmed drug-target interactions but 100,024 unconfirmed interactions. To address the prediction issues caused by the imbalanced class distribution, we randomly selected 80,291 unconfirmed drug-target interactions for prediction. Another potential issue arises from the need to observe multiple variables in drug-target interaction feature data, with many variables possibly correlated, increasing the complexity of the analysis. Therefore, we standardized the data and used principal component analysis to reduce the dimensionality of the feature vectors to 996, in order to meet the input requirements of the DNN for computational convenience. These 996 nodes were used as the input layer, with the DNN-based DTI model comprising four hidden layers with 512, 256, 128, and 64 neurons, respectively. The hidden layers employed rectified linear unit activation functions, and dropout layers were added to each hidden layer to avoid overfitting. The output layer consisted of a single neuron using a sigmoid activation function to constrain the output between 0 and 1, representing the interaction probability between the drug and the drug target.

Our DNN-based DTI model was evaluated using 5-fold cross-validation to assess its performance. The loss and accuracy curves during the training process are shown in Figures S3 and S4 of the supplementary materials, respectively, and the results of the five-fold cross-validation are presented in Table S2 of the supplementary materials. The average test accuracy was 93.2%, with a standard deviation of 0.118%. Additionally, we used ROC curves as a reference metric, where the AUC of the ROC ranges between 0 and 1, with higher values indicating better DTI model performance. Ultimately, our DNN-based DTI model achieved an AUC of 0.981, as shown in Figure S5 of the supplementary materials, outperforming a random prediction model (AUC = 0.5).

However, we aimed to design molecular drugs that not only effectively alleviate symptoms of FTD but also minimize adverse reactions. Generally, potent molecular drugs may be more irritating to the human body. Therefore, balancing drug efficacy and side effects is our next focus. We used three drug design specifications for screening the predicted molecular drugs: Regulatory capacity, sensitivity, and toxicity, referring to the LINCS L1000 Level 5, PRISM, and ADMETlab 2.0 databases. Based on strong regulatory capacity, high sensitivity, and low toxicity, we selected several potential molecular drugs for each biomarker of FTD, as shown in Table 5. Additionally, since Iodophenpropit, TTNPB, Probucol, and Clobenpropit serve as potential molecular drugs for these significant biomarkers, we combined these four molecular drugs to propose potential multi-molecular drugs for treating FTD.

Discussion

Iodophenpropit and Clobenpropit are both antagonists of H3 receptors commonly used to study the role of histamine in the nervous system. The histamine H3 receptor is a biogenic amine that plays a significant role in central nervous system activities, including learning and memory. Literature data indicate that elevated levels of H3 receptor expression can lead to cognitive impairment.[46] Given preclinical evidence that blocking the H3 receptor reduces impulsivity, improves attention, and enhances learning and memory, H3 receptor antagonists have been clinically used to treat various cognitive disorders. Iodophenpropit, as one of these H3 receptor antagonists, was the first compound successfully developed for labeling H3 receptors in rat brain membranes in previous studies.[46] In another study, Clobenpropit was found to improve memory impairment induced by lipopolysaccharides. Lipopolysaccharide triggers neuroinflammation by modulating cyclooxygenase activity and cytokine levels in the brain, leading to memory deficits.[47]

TTNPB is a synthetic analog of all-trans retinoic acid, belonging to the vitamin A derivatives (retinoid) family. Retinoids interact with retinoic acid receptors and retinoid X receptors, significantly impacting physiological and pathological signaling pathways in the brain. Impaired retinoic acid signaling is a crucial factor leading to neurodegenerative diseases.[48] Recent studies have shown that deprivation of retinoic acid in mice results in severe deficits in spatial learning and memory. Additionally, a critical pathological hallmark of retinoids is the production and deposition of Aβ, the forma-

Nat Cell Sci 2024;2(4):207–223
https://doi.org/10.61474/ncs.2024.00043

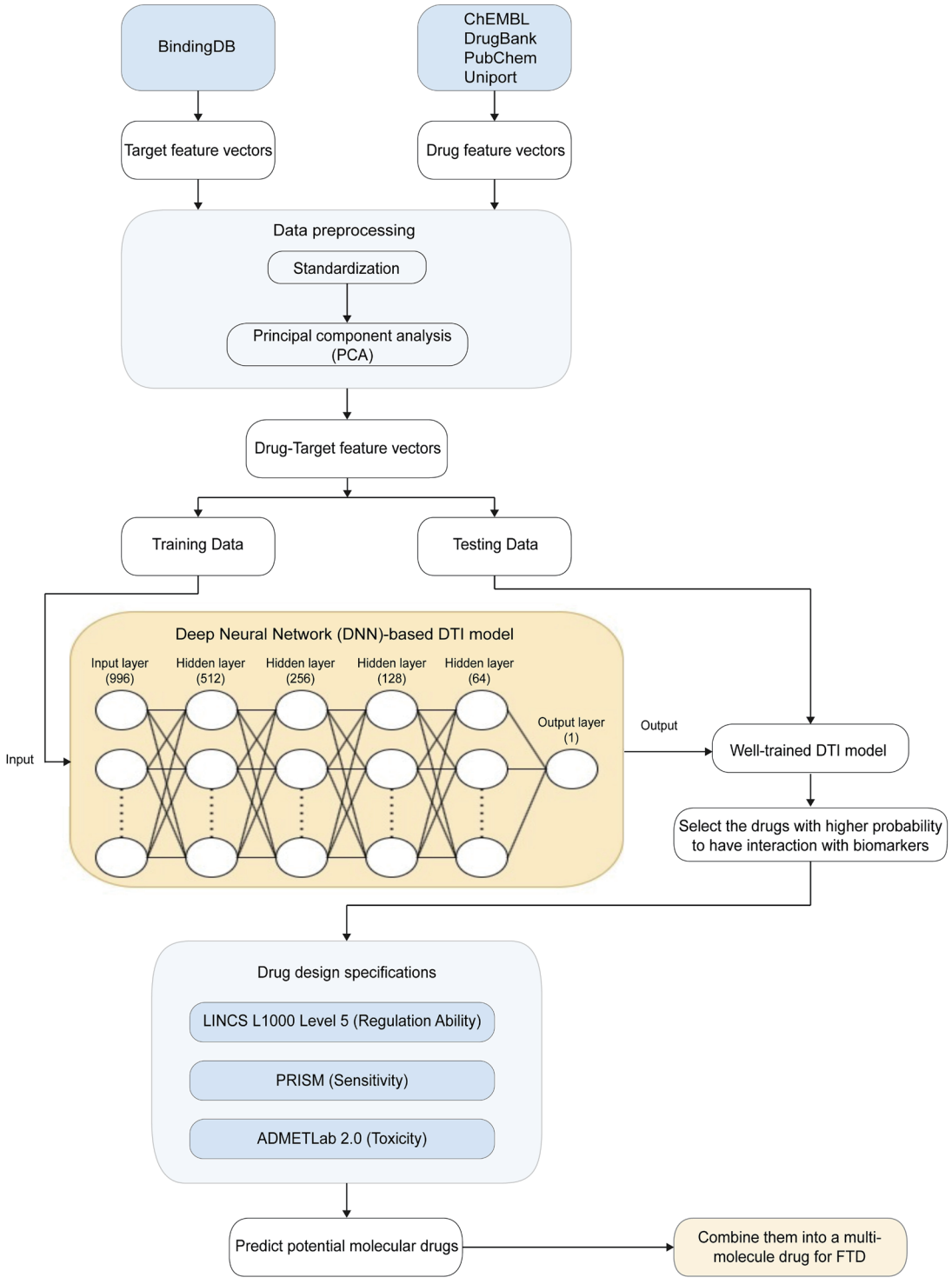

Fig. 3. The flowchart of training a deep neural network (DNN)-based drug-target interaction (DTI) model and designing a multi-molecule drug for the treatment of frontotemporal dementia (FTD). First, we preprocess the original data through drug-target interaction databases to generate the drug-target feature vectors. Seventy-five percent of the drug-target feature vector data was used as training data, and the remaining 25% was used for testing. After training the DTI model, we obtain the score for each molecular drug (i.e., the higher the score, the higher the probability of interaction with biomarkers). We then select candidate molecular drugs according to their scores. Finally, we use three drug design specifications to predict the most promising molecular drugs and combine them into a multi-molecule drug for treating FTD.

Table 5. Using drug design specifications to select potential molecule drugs

Drug	Target							
	TAU	GSK-3β	STAT3	*ATG5*	WDR41	RIPK1	Sensitivity (PRISM)	Toxicity (LC50)
Iodophenpropit				↑	↓	↓	−0.0032	4.94
Clobenpropit	↓	↓		↑			0.0416	4.223
TTNPB	↓		↓			↓	0.2480	5.823
Probucol	↓		↓			↓	−0.2931	7.007

↑ denotes up-regulation; ↓ denotes down-regulation. PRISM, Protein Interactions by Structural Matching; LC50, Lethal Concentration 50%.

tion and phosphorylation of NFT, and the inflammation and autoimmune response.[49] These pathologies are also significant mechanisms in the etiology of FTD. Currently, TTNPB has been identified in clinical settings as a potent retinoid receptor agonist with potential for treating neurodegenerative diseases due to its metabolic resistance and high affinity for retinoic acid receptors.[48]

Probucol is a historically established cholesterol-lowering drug, but recent studies have explored its potential as a treatment for dementia. This interest arises because Probucol has been shown to inhibit Aβ secretion in mouse models while maintaining blood-brain barrier function, suppressing neurovascular inflammation, and directly influencing neuroprotection and adaptability.[50] In mouse models with ischemia-induced blood-brain barrier dysfunction, Probucol preserved the proper localization of tight junction proteins in endothelial cells by attenuating sphingosine-1-phosphate signaling and inhibiting the expression of STAT3, thereby reducing the leakage of small molecules into the brain parenchyma. Furthermore, in *in vitro* models of brain endothelial dysfunction, Probucol was found to inhibit the expression of CASP3.[50] The cleavage of CASP3 is one of the causes of TAU phosphorylation, and inhibiting CASP3 can reduce the formation of NFT caused by TAU phosphorylation. Taken together, these findings suggest that Probucol may provide therapeutic benefits for FTD by effectively enhancing neuronal survival and plasticity, making it a promising candidate for FTD treatment.[51]

In summary, there is currently no drug on the market that can completely cure FTD. The four molecular drugs we have screened have not yet been practically applied to treat FTD in humans. However, based on our systematic research and analysis, by understanding the mechanisms of action of these molecular drugs, effectively regulating the expression of biomarkers, and improving the pathogenic mechanisms of FTD, we have ultimately selected these four molecular drugs as potential treatments for FTD. Compared to traditional drugs, these small-molecule compounds we have chosen as a molecular drug combination offer several advantages. Firstly, development costs: Traditional drugs and multi-molecular drugs are associated with high costs, time consumption, and low efficiency. Secondly, multi-target action: Multi-molecular drugs can act on multiple targets simultaneously, thereby increasing efficacy, especially for complex diseases like neurodegenerative disorders.

Conclusions

In this study, we explored the pathogenic mechanism of FTD from a systems biology perspective and designed a multi-molecule drug for its treatment. To achieve this goal, we began by constructing candidate GWGENs for FTD and healthy control through big data mining, including candidate PPINs and candidate GRNs. The next step was to identify the true GWGENs for FTD and healthy control using their microarray data and system identification methods. Subsequently, we applied the PNP method to extract the core GWGENs for both FTD and healthy control. By annotating these core GWGENs with KEGG pathways, we identified the core signaling pathways involved in FTD and healthy control and investigated the pathogenetic mechanisms to pinpoint significant biomarkers for FTD.

In the core pathogenic signaling pathways of FTD, we identified significant biomarkers, including TAU, GSK-3β, STAT3, *ATG5*, WDR41, and RIPK1, as potential drug targets. Based on the prediction of candidate molecular drugs from the DNN-based DTI model, we selected Iodophenpropit, TTNPB, Probucol, and Clobenpropit as a multi-molecule drug combination targeting multiple biomarkers to restore the pathogenic cellular functions of FTD to normal levels. With further clinical and experimental validation, we hope that the proposed multi-molecule drug will improve the cellular functions in FTD patients. The potential therapeutic efficacy of these molecular drug combinations is expected to offer new treatment options for FTD patients.

Acknowledgments

The authors thank Prof. Yung-Jen Chuang from the Institute of Bioinformatics and Structural Biology, National Tsing Hua University, Hsinchu, Taiwan for general discussions about this topic.

Funding

None.

Conflict of interest

The authors declare no conflict of interest.

Author contributions

Conceptualization (BSC), methodology (WLC), software (WLC), validation (WLC), formal analysis (WLC), investigation (WLC), data curation (WLC), writing—original draft preparation (WLC), writing—review and editing (BSC), visualiza-

tion (WLC), and supervision (BSC). All authors have read and agreed to the published version of the manuscript.

Ethical statement

Ethical approval is not applicable due to the use of the publicly available dataset GSE140830 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE140830).

Data sharing statement

The whole blood data for FTD and healthy controls is from GSE140830 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE140830) (accessed on November 13, 2020). Drug sensitivity data is from the DepMap portal (https://depmap.org/portal/download/) (accessed on May 19, 2021). Data related to this research can be found in the Supplementary file.

References

[1] Wilfong L, Edwards NE, Yehle KS, Ross K. Frontotemporal Dementia: Identification and Management. The Journal for Nurse Practitioners 2016;12(4):277–282. doi:10.1016/j.nurpra.2015.08.006.
[2] Silveri MC. Frontotemporal dementia to Alzheimer's disease. Dialogues Clin Neurosci 2007;9(2):153–160. doi:10.31887/DCNS.2007.9.2/msilveri, PMID:17726914.
[3] Wang IF, Guo BS, Liu YC, Wu CC, Yang CH, Tsai KJ, *et al*. Autophagy activators rescue and alleviate pathogenesis of a mouse model with proteinopathies of the TAR DNA-binding protein 43. Proc Natl Acad Sci USA 2012;109(37):15024–15209. doi:10.1073/pnas.1206362109, PMID:22932872.
[4] Wang IF, Tsai KJ, Shen CK. Autophagy activation ameliorates neuronal pathogenesis of FTLD-U mice: a new light for treatment of TARDBP/TDP-43 proteinopathies. Autophagy 2013;9(2):239–240. doi:10.4161/auto.22526, PMID:23108236.
[5] Blagosklonny MV. Rapamycin for longevity: opinion article. Aging (Albany NY) 2019;11(19):8048–8067. doi:10.18632/aging.102355, PMID:31586989.
[6] Orgeta V, Leung P, Del-Pino-Casado R, Qazi A, Orrell M, Spector AE, *et al*. Psychological treatments for depression and anxiety in dementia and mild cognitive impairment. Cochrane Database Syst Rev 2022;4(4):CD009125. doi:10.1002/14651858.CD009125.pub3, PMID:35466396.
[7] Zhou SF, Zhong WZ. Drug Design and Discovery: Principles and Applications. Molecules 2017;22(2):279. doi:10.3390/molecules22020279, PMID:28208821.
[8] Yamaguchi S, Kaneko M, Narukawa M. Approval success rates of drug candidates based on target, action, modality, application, and their combinations. Clin Transl Sci 2021;14(3):1113–1122. doi:10.1111/cts.12980, PMID:33831276.
[9] Abdul Raheem AK, Dhannoon BN. Comprehensive Review on Drug-target Interaction Prediction - Latest Developments and Overview. Curr Drug Discov Technol 2024;21(2):e010923220652. doi:10.2174/1570163820666230901160043, PMID:37680152.
[10] Chen R, Liu X, Jin S, Lin J, Liu J. Machine Learning for Drug-Target Interaction Prediction. Molecules 2018;23(9):2208. doi:10.3390/molecules23092208, PMID:30200333.
[11] Zhan YP, Chen BS. Drug Target Identification and Drug Repurposing in Psoriasis through Systems Biology Approach, DNN-Based DTI Model and Genome-Wide Microarray Data. Int J Mol Sci 2023;24(12):10033. doi:10.3390/ijms241210033, PMID:37373186.
[12] Su PW, Chen BS. Systems Drug Design for Muscle Invasive Bladder Cancer and Advanced Bladder Cancer by Genome-Wide Microarray Data and Deep Learning Method with Drug Design Specifications. Int J Mol Sci 2022;23(22):13869. doi:10.3390/ijms232213869, PMID:36430344.
[13] Stone WL, Leavitt L, Varacallo M. Physiology, Growth Factor. Treasure Island, FL: StatPearls Publishing; 2024.
[14] Sengupta P, Das R, Majumder P, Mukhopadhyay D. Connecting the ends: signaling via receptor tyrosine kinases and cytoskeletal degradation in neurodegeneration. Explor Neurosci 2024;3(1):1–26. doi:10.37349/en.2024.00033.
[15] Seiler C, Stainthorp AK, Ketchen S, Jones CM, Marks K, Quirke P, *et al*. The Grb2 splice variant, Grb3-3, is a negative regulator of RAS activation. Commun Biol 2022;5(1):1029. doi:10.1038/s42003-022-03985-7, PMID:36171279.
[16] Li L, Zhao GD, Shi Z, Qi LL, Zhou LY, Fu ZX. The Ras/Raf/MEK/ERK signaling pathway and its role in the occurrence and development of HCC. Oncol Lett 2016;12(5):3045–3050. doi:10.3892/ol.2016.5110, PMID:27899961.
[17] Lara Aparicio SY, Laureani Fierro ÁJ, Aranda Abreu GE, Toledo Cárdenas R, García Hernández LI, Coria Ávila GA, *et al*. Current Opinion on the Use of c-Fos in Neuroscience. NeuroSci 2022;3(4):687–702. doi:10.3390/neurosci3040050, PMID:39483772.
[18] Guise S, Braguer D, Carles G, Delacourte A, Briand C. Hyperphosphorylation of tau is mediated by ERK activation during anticancer drug-induced apoptosis in neuroblastoma cells. J Neurosci Res 2001;63(3):257–267. doi:10.1002/1097-4547(20010201)63:3<257::AID-JNR1019>3.0.CO;2-T, PMID:11170175.
[19] Korade Z, Mirnics K. Wnt signaling as a potential therapeutic target for frontotemporal dementia. Neuron 2011;71(6):955–957. doi:10.1016/j.neuron.2011.09.002, PMID:21943593.
[20] Rosen EY, Wexler EM, Versano R, Coppola G, Gao F, Winden KD, *et al*. Functional genomic analyses identify pathways dysregulated by progranulin deficiency, implicating Wnt signaling. Neuron 2011;71(6):1030–1042. doi:10.1016/j.neuron.2011.07.021, PMID:21943601.
[21] Kolligs FT, Bommer G, Göke B. Wnt/beta-catenin/tcf signaling: a critical pathway in gastrointestinal tumorigenesis. Digestion 2002;66(3):131–144. doi:10.1159/000066755, PMID:12481159.
[22] Yang K, Hitomi M, Stacey DW. Variations in cyclin D1 levels through the cell cycle determine the proliferative fate of a cell. Cell Div 2006;1:32. doi:10.1186/1747-1028-1-32, PMID:17176475.
[23] Gibbons L, Rollinson S, Thompson JC, Robinson A, Davidson YS, Richardson A, *et al*. Plasma levels of progranulin and interleukin-6 in frontotemporal lobar degeneration. Neurobiol Aging 2015;36(3):1603.e1–1603.e4. doi:10.1016/j.neurobiolaging.2014.10.023, PMID:25435337.
[24] Rusek M, Smith J, El-Khatib K, Aikins K, Czuczwar SJ, Pluta R. The Role of the JAK/STAT Signaling Pathway in the Pathogenesis of Alzheimer's Disease: New Potential Treatment Target. Int J Mol Sci 2023;24(1):864. doi:10.3390/ijms24010864, PMID:36614305.
[25] Garcia G, Pinto S, Ferreira S, Lopes D, Serrador MJ, Fernandes A, *et al*. Emerging Role of miR-21-5p in Neuron-Glia Dysregulation and Exosome Transfer Using Multiple Models of Alzheimer's Disease. Cells 2022;11(21):3377. doi:10.3390/cells11213377, PMID:36359774.
[26] Benussi A, Ashton NJ, Karikari TK, Gazzina S, Premi E, Benussi L, *et al*. Serum Glial Fibrillary Acidic Protein (GFAP) Is a Marker of Disease Severity in Frontotemporal Lobar Degeneration. J Alzheimers Dis 2020;77(3):1129–1141. doi:10.3233/JAD-200608, PMID:32804092.
[27] Hohman TJ, Bell SP, Jefferson AL, Alzheimer's Disease Neuroimaging Initiative. The role of vascular endothelial growth factor in neurodegeneration and cognitive decline: exploring interactions with biomarkers of Alzheimer disease. JAMA Neurol 2015;72(5):520–529. doi:10.1001/jamaneurol.2014.4761, PMID:25751166.
[28] Kiyatkin A, Aksamitiene E, Markevich NI, Borisov NM, Hoek JB, Kholodenko BN. Scaffolding protein Grb2-associated binder 1 sustains epidermal growth factor-induced mitogenic and survival signaling by multiple positive feedback loops. J Biol Chem 2006;281(29):19925–19938. doi:10.1074/jbc.M600482200, PMID:16687399.
[29] Razani E, Pourbagheri-Sigaroodi A, Safaroghli-Azar A, Zoghi A, Shanaki-Bavarsad M, Bashash D. The PI3K/Akt signaling axis in Alzheimer's disease: a valuable target to stimulate or suppress? Cell Stress Chaperones 2021;26(6):871–887. doi:10.1007/s12192-021-01231-3, PMID:34386944.
[30] Zhang X, Tang N, Hadden TJ, Rishi AK. Akt, FoxO and regulation

of apoptosis. Biochim Biophys Acta 2011;1813(11):1978–1986. doi:10.1016/j.bbamcr.2011.03.010, PMID:21440011.
[31] Longobardi A, Catania M, Geviti A, Salvi E, Vecchi ER, Bellini S, *et al*. Autophagy Markers Are Altered in Alzheimer's Disease, Dementia with Lewy Bodies and Frontotemporal Dementia. Int J Mol Sci 2024;25(2):1125. doi:10.3390/ijms25021125, PMID:38256197.
[32] Webster CP, Smith EF, Bauer CS, Moller A, Hautbergue GM, Ferraiuolo L, *et al*. The C9orf72 protein interacts with Rab1a and the ULK1 complex to regulate initiation of autophagy. EMBO J 2016;35(15):1656–1676. doi:10.15252/embj.201694401, PMID:27334615.
[33] Tang D, Sheng J, Xu L, Zhan X, Liu J, Jiang H, *et al*. Cryo-EM structure of C9ORF72-SMCR8-WDR41 reveals the role as a GAP for Rab8a and Rab11a. Proc Natl Acad Sci U S A 2020;117(18):9876–9883. doi:10.1073/pnas.2002110117, PMID:32303654.
[34] Alers S, Löffler AS, Wesselborg S, Stork B. Role of AMPK-mTOR-Ulk1/2 in the regulation of autophagy: cross talk, shortcuts, and feedbacks. Mol Cell Biol 2012;32(1):2–11. doi:10.1128/MCB.06159-11, PMID:22025673.
[35] Osawa T, Noda NN. Atg2: A novel phospholipid transfer protein that mediates de novo autophagosome biogenesis. Protein Sci 2019;28(6):1005–1012. doi:10.1002/pro.3623, PMID:30993752.
[36] Graef M. Membrane tethering by the autophagy ATG2A-WIPI4 complex. Proc Natl Acad Sci U S A 2018;115(42):10540–10541. doi:10.1073/pnas.1814759115, PMID:30275332.
[37] Holbrook J, Lara-Reyna S, Jarosz-Griffiths H, McDermott M. Tumour necrosis factor signalling in health and disease. F1000Res 2019;8(F1000 Faculty Rev):111. doi:10.12688/f1000research.17023.1, PMID:30755793.
[38] Rodriguez DA, Tummers B, Shaw JJP, Quarato G, Weinlich R, Cripps J, *et al*. The interaction between RIPK1 and FADD controls perinatal lethality and inflammation. Cell Rep 2024;43(6):114335. doi:10.1016/j.celrep.2024.114335, PMID:38850531.
[39] Opland CK, Bryan MR, Harris B, McGillion-Moore J, Tian X, Chen Y, *et al*. Activity-dependent tau cleavage by caspase-3 promotes neuronal dysfunction and synaptotoxicity. iScience 2023;26(6):106905. doi:10.1016/j.isci.2023.106905, PMID:37305696.
[40] Han J, Wu J, Silke J. An overview of mammalian p38 mitogen-activated protein kinases, central regulators of cell stress and receptor signaling. F1000Res 2020;9(F1000 Faculty Rev):653. doi:10.12688/f1000research.22092.1, PMID:32612808.
[41] Sattarifard H, Safaei A, Khazeeva E, Rastegar M, Davie JR. Mitogen- and stress-activated protein kinase (MSK1/2) regulated gene expression in normal and disease states. Biochem Cell Biol 2023;101(3):204–219. doi:10.1139/bcb-2022-0371, PMID:36812480.
[42] Kummer KK, Zeidler M, Kalpachidou T, Kress M. Role of IL-6 in the regulation of neuronal development, survival and function. Cytokine 2021;144:155582. doi:10.1016/j.cyto.2021.155582, PMID:34058569.
[43] Grieco M, Giorgi A, Gentile MC, d'Erme M, Morano S, Maras B, *et al*. Glucagon-Like Peptide-1: A Focus on Neurodegenerative Diseases. Front Neurosci 2019;13:1112. doi:10.3389/fnins.2019.01112, PMID:31680842.
[44] Zheng HY, Wang YX, Zhou K, Xie HL, Ren Z, Liu HT, *et al*. Biological functions of CRTC2 and its role in metabolism-related diseases. J Cell Commun Signal 2023;17(3):495–506. doi:10.1007/s12079-023-00730-5, PMID:36856929.
[45] Garrett LR, Niccoli T. Frontotemporal Dementia and Glucose Metabolism. Front Neurosci 2022;16:812222. doi:10.3389/fnins.2022.812222, PMID:35281504.
[46] Chen PY, Tsai CT, Ou CY, Hsu WT, Jhuo MD, Wu CH, *et al*. Computational analysis of novel drugs designed for use as acetylcholinesterase inhibitors and histamine H3 receptor antagonists for Alzheimer's disease by docking, scoring and de novo evolution. Mol Med Rep 2012;5(4):1043–1048. doi:10.3892/mmr.2012.757, PMID:22267207.
[47] Mani V, Arfeen M, Ali HM, Abdel-Moneim AH, Aldubayan M, Alhowail A. Neuroprotective Effect of Clobenpropit against Lipopolysaccharide-Induced Cognitive Deficits via Attenuating Neuroinflammation and Enhancing Mitochondrial Functions in Mice. Brain Sci 2021;11(12):1617. doi:10.3390/brainsci11121617, PMID:34942919.
[48] Chakrabarti M, McDonald AJ, Will Reed J, Moss MA, Das BC, Ray SK. Molecular Signaling Mechanisms of Natural and Synthetic Retinoids for Inhibition of Pathogenesis in Alzheimer's Disease. J Alzheimers Dis 2016;50(2):335–352. doi:10.3233/JAD-150450, PMID:26682679.
[49] Alsharief M. How Do Retinoids Affect Alzheimer's Disease and Can They Be Novel Drug Candidates? Cureus 2024;16(4):e57548. doi:10.7759/cureus.57548, PMID:38572181.
[50] Sharif A, Mamo J, Lam V, Al-Salami H, Mooranian A, Watts GF, *et al*. The therapeutic potential of probucol and probucol analogues in neurodegenerative diseases. Transl Neurodegener 2024;13(1):6. doi:10.1186/s40035-024-00398-w, PMID:38247000.
[51] Santos DB, Colle D, Moreira ELG, Peres KC, Ribeiro RP, Dos Santos AA, de Oliveira J, Hort MA, de Bem AF, Farina M. Probucol mitigates streptozotocin-induced cognitive and biochemical changes in mice. Neuroscience 2015;284:590–600. doi:10.1016/j.neuroscience.2014.10.019, PMID:25453776.