# M3NetFlow: a novel multi-scale multi-hop multi-omics graph AI model for omics data integration and interpretation

Heming Zhang[1], S. Peter Goedegebuure[2,3], Li Ding[3,4], Williams Hawkins[2,3], David DeNardo[3,4], Ryan C. Fields[2,3], Yixin Chen[6], Philip Payne[1], Fuhai Li[1,5#]

[1]Institute for Informatics, Data Science and Biostatistics (I2DB), [2]Department of Surgery, [3]Siteman Cancer Center, [4]Department of medicine, [5]Department of Pediatrics, Washington University School of Medicine, [6]Department of Computer Science and Engineering, Washington University in St. Louis, St. Louis, MO, USA.
[#]Correspondence: Fuhai.Li@wustl.edu

**Abstract:** The integration and interpretation of multi-omics data play a crucial role in systems biology for prioritizing vital molecular targets and deciphering core signaling pathways of complex diseases, such as cancer, COVID-19 and Alzheimer's disease. However, it remains a challenge that has not been adequately addressed. Graph neural networks (GNN) have emerged as powerful artificial intelligence models for analyzing data with a graphical structure. Nevertheless, GNN models have not been sufficiently designed for integrative and interpretable multi-omics data analysis. In this study, we propose a novel multi-scale multi-hop multi-omics GNN model, ***M3NetFlow***, to integrate and interpret multi-omics data to rank key targets and infer core signaling pathways. Specifically, we applied the ***M3NetFlow*** model to infer cell-line-specific core signaling networks explaining drug combination response. The evaluation and comparison results on drug combination prediction showed that the ***M3NetFlow*** model achieved significantly higher prediction accuracy than existing GNN models. Furthermore, ***M3NetFlow*** can predict key targets and infer essential signaling networks regulating drug combination response. It is critical for guiding the development of personalized precision medicine for patients with drug resistance. This model can be applied to general multi-omics data-driven research. Aside from that, we developed the visualization tool, ***NetFlowVis***, the better analysis of targets and signaling pathways of drugs and drug combinations.

## 1. Introduction

In diseases and treatments with complex natures, such as cancer and Alzheimer's disease (AD), it is widely accepted that intricate and fundamental molecular signaling networks play a significant role in disease progression and the emergence of drug resistance. To gain a comprehensive understanding of the core disease signaling pathways and drug responses, extensive multi-omics datasets encompassing genetic, epigenetic, transcriptomic, and proteomic information have been generated[1–5]. And multi-omics data characterize the complex dysfunctional disease signaling pathways/biological processes from different perspectives. For example, non-mutated cancer genes can be dysregulated through cellular pathways, such as DNA hypermethylation around gene promoters, which can deactivate tumor suppressors, like the MLH1 gene[6]. Conversely, hypomethylation at oncogene promoters can activate them, leading to tumorigenesis[7]. Aside from that, combining genetic and pharmacologic profiling, the cell perturbations might reveal the potential drug targets for specified cancer cell lines. Deeper insights can be revealed by adding the proteomic data to genomic and transcriptomic data. For example, special high-grade serous ovarian cancer analysis was made by adding proteomic data sets to genomic and transcriptomic data[8]. Also, the chromosome instability (CIN[9]) index was leveraged to identify the proteins associated with chromosomal structural abnormality, such as copy number alterations (CAN[8]). What's more, co-expression patterns were analyzed to differentiate the HRD from non-HRD patients, pointing out a subnetwork of ~30 proteins. In addition, statistical analysis was made by applying a two-sided t-test to distinguish short- and long-term survival patients. This analysis showed that proteomics data can be integrated with genomic data to identify multiple pathways and processes that are driving ovarian cancer and help stratify patients for better therapeutic outcomes[10]. Hence, integrating multi-omics data from different levels can better describe the molecular mechanism of cellular processes.

It remains an open problem or challenging task for integrative and interpretable omics data analysis, though a set of computational models has been proposed. A comprehensive review of existing multi-omics data integration analysis models was reported[10]. Specifically, these models were clustered into a few categories, like similarity, correlation, Bayesian, multivariate, fusion and network-based models. The PAthway Representation and Analysis by Direct Inference on Graphical Models (PARADIGM[11]) is one of the most widely used methods among these traditional computational methods. For the scRNA-seq data analysis, some existing models, like NicheNet[12], Cytotalk[13], have been developed to infer the intra- and inter-cellular dysfunctional signaling pathways. In the current models, only traditional random walk-based and Steiner-tree-based approaches were used, which are limited to uncovering the core signaling network. For example, in the NicheNet model, only the potential ligands-target (of transcription factors) was scored using personalized PageRank (a type of network random-walk centrality score), and there is no function to uncover the core signaling network. In the CytoTalk model, the Steiner-tree-based approach was used, which can only identify a minimum set of interactions to link a set of priorly selected differentially expressed genes, and many essential signaling pathways will be missed. Despite identifying numerous novel molecular targets and medications through the integration analysis of multi-omics data, our comprehension of the fundamental signaling pathways in complex diseases is still constrained. The integration and analysis of multi-omics data to uncover the crucial molecular targets and core signaling pathways continue to be an ongoing and arduous endeavor[10,14,15].

Recently, Graph Neural Networks (GNNs) have gained prominence due to their capability to model relationships within graph-structured data[16–19]. And numerous studies have applied the Graph Neural Network (GNN) with the integration of the multi-omics data. MOGONET[20] initially creates similarity graphs among samples by leveraging each omics data, then employs a Graph Convolutional Network (GCN[16]) to learn a label distribution from each omics data independently. Subsequently, a Cross-omics discovery tensor is implemented to refine the prediction by learning the dependency among multi-omics data. MoGCN[21] adopts a similar approach by constructing a patient similarity network using multi-omics data and then using GCN to predict the cancer subtype of patients. GCN-SC[22] utilizes a GCN to combine single-cell multi-omics data derived from varying sequencing methodologies. MOGCL[23] takes this further by exploiting the potency of graph contrastive learning to pretrain the GCN on the multi-omics dataset, thereby achieving impressive results in downstream tasks with fine-tuning. Nevertheless, none of the aforementioned techniques contemplate incorporating structured signaling data like KEGG into the model. Aside from that, those traditional GNN models are limited by the issue of over-smoothing, which hampers their ability to incorporate a large number of layers. Consequently, these models can only consider a limited amount of multi-hop/$\boldsymbol{K}$-hop information ($K$ usually smaller than 4). The over-smoothing problem was firstly mentioned by extending the propagation layers in GCN[24]. Moreover, theoretical papers using Dirichlet energy showed diminished discriminative power by increasing the propagation layers[25].

In this research, we present a novel solution called ***M3NetFlow (Multi-scale, Multi-Hop, Multi-omics NETwork Flow inference)*** to address the challenges mentioned above. Using an attention mechanism, our approach tackles these challenges by first incorporating local $K$-hop information within each subgraph. Subsequently, we establish global-level bi-directional propagation to facilitate message passing between genes/proteins. In the interpretation phase, we leverage the attention mechanism to aggregate the weights of all connected paths for genes, and a reweighting process inspired by (Term Frequency – Inverse Document Frequency) TF-IDF[26] is employed to redistribute the importance scores of genes across different cell lines. Afterward, a visualization tool, ***NetFlowVis,***

was developed for the better analysis of targets and signaling pathways of drugs and drug combinations. To assess and demonstrate the effectiveness of our proposed model, ***M3NetFlow***, we applied it to extract cell-line-specific signaling networks that explain drug effects by integrating multi-omics data. Comparative results indicated the superior performance of the ***M3NetFlow*** model compared to state-of-the-art models. Moreover, the ***M3NetFlow*** model has the capability to infer personalized signaling targets and signaling networks, enabling the inference of critical targets and core signaling networks and guiding the advancement of personalized precision medicine.

## 2. Methodology

**Table 1** provided the links from which we collected data on drug combination effects, multi-omics data, gene-gene interactions, and drug-gene interactions. These data were preprocessed and represented in a diagram depicted in **Figure 1**.

**Table 1**. Four sources and file types of databases leveraged in the deep learning model

| Database Type | Database Name | Website Link |
|---|---|---|
| Drug Combination Effect Data | NCI ALMANAC[27] | https://wiki.nci.nih.gov/display/NCIDTPdata/NCI-ALMANAC |
| Multi-omics Data | Cell Model Passports[28] RNA-Seq | https://cog.sanger.ac.uk/cmp/download/rnaseq_20191101.zip |
| | Cell Model Passports[28] CNV | https://cog.sanger.ac.uk/cmp/download/cnv_20191101.zip |
| | CCLE[29] Methylation | https://data.broadinstitute.org/ccle/CCLE_RRBS_TSS1kb_20181022.txt.gz |
| | CCLE[29] Gene Amplification | https://data.broadinstitute.org/ccle_legacy_data/binary_calls_for_copy_number_and_mutation_data/CCLE_MUT_CNA_AMP_DEL_binary_Revealer.gct |
| | CCLE[29] Gene Deletion | https://data.broadinstitute.org/ccle_legacy_data/binary_calls_for_copy_number_and_mutation_data/CCLE_MUT_CNA_AMP_DEL_binary_Revealer.gct |
| Gene-Gene Interaction Data | KEGG[30] Gene Interactions Network | https://www.genome.jp/kegg/ |
| Drug-Target Interaction Data | DrugBank[31] | https://go.drugbank.com/ |

### 2.1 Drug combination screening datasets

In this paper, we utilized the NCI ALMANAC and O'Neil[32] drug screening datasets to train the deep learning models. The NCI ALMANAC dataset consists of combo scores for various permutations of 104 FDA-approved drugs, representing their impact on tumor growth in NCI60 human tumor cell lines. For evaluating the synergy score of two drugs on a specific tumor cell line, we utilized the average combo-score of two drugs at different doses, employing a 4-element tuple: $<D_A, D_B, C_C, S_{ABC}>$. On the other hand, the O'Neil dataset, obtained from the DrugComb[33] platform, served as another drug combination screening dataset. We extracted the processed dataset from this platform, which provided average synergy Loewe scores. These scores assessed the synergy score of two drugs on a given tumor cell line, employing a 4-element tuple: $<D_A, D_B, C_C, S_{ABC}>$.

### 2.2 Multi-omics data of cancer cell lines

In this study, multi-omics data comprising RNA-seq, copy number variation, gene methylation, and gene mutation data for a total of 1,489 genes were incorporated into the model. These data were obtained from the Cell Model Passports[28] and CCLE[29] databases. Through the identification of overlapping cell lines from these databases, we identified 42 cell lines that were present in both the Cell Model Passports database and the NCI ALMANAC dataset (refer to **Table 2**). Additionally, we found 24 cell lines that were present in both the Cell Model Passports database and the O'Neil dataset (refer to **Table 3**).

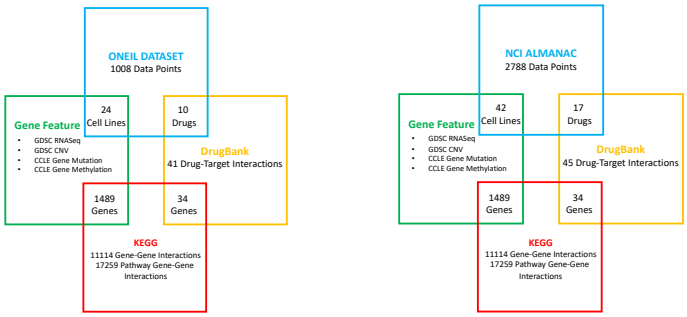


**Figure 1.** Preprocessed data of raw files collected from various databases

**Table 2.** 42 cell lines in preprocessed NCI ALMANAC and Multi-omics datasets

| | | | | |
|---|---|---|---|---|
| A498 | A549/ATCC | ACHN | BT-549 | CAKI-1 |
| DU-145 | EKVX | HCT-116 | HCT-15 | HOP-62 |
| HOP-92 | HS 578T | IGROV1 | K-562 | KM12 |
| LOX IMVI | MCF7 | MDA-MB-231/ATCC | MDA-MB-468 | NCI-H23 |
| NCI-H460 | NCI-H522 | OVCAR-3 | OVCAR-4 | OVCAR-8 |
| PC-3 | RPMI-8226 | SF-268 | SF-295 | SF-539 |
| SK-MEL-2 | SK-MEL-28 | SK-MEL-5 | SK-OV-3 | SNB-75 |
| SR | SW-620 | T-47D | U251 | UACC-257 |
| UACC-62 | UO-31 | | | |

**Table 3.** 24 human tumor cell lines included in Cell Model Passports database intersected with ONEIL

| | | | | |
|---|---|---|---|---|
| A2058 | A2780 | A375 | CAOV3 | HCT116 |
| HT144 | LOVO | MDAMB436 | NCI-H460 | NCIH1650 |
| NCIH2122 | NCIH23 | OV90 | OVCAR3 | RKO |
| RPMI7951 | SK-OV-3 | SKMEL30 | SKMES1 | SW-620 |
| SW837 | T-47D | UACC62 | VCAP | |

### 2.3 KEGG database

About 59,241 gene-gene interactions for over 8,000 genes across different signaling pathways were collected from KEGG[30] database. And 48 signaling pathways in the KEGG dataset were selected as the ground truth of subgraphs in the gene-gene interactions, which were AGE-RAGE signaling pathway in diabetic complications, AMPK signaling pathway, Adipocytokine signaling pathway, Apelin signaling pathway, B cell receptor signaling pathway, C-type lectin receptor signaling pathway, Calcium signaling pathway, Chemokine signaling pathway, ErbB signaling pathway, Estrogen signaling pathway, Fc epsilon RI signaling pathway, FoxO signaling pathway, Glucagon signaling pathway, GnRH signaling pathway, HIF-1 signaling pathway, Hedgehog signaling pathway, Hippo signaling pathway, Hippo signaling pathway - multiple species, IL-17 signaling pathway, Insulin signaling pathway, JAK-STAT signaling pathway, MAPK signaling pathway, NF-kappa B signaling pathway, NOD-like receptor signaling pathway, Neurotrophin signaling pathway, Notch signaling pathway, Oxytocin signaling pathway, PI3K-Akt signaling pathway, PPAR signaling pathway, Phospholipase D signaling pathway, Prolactin signaling pathway, RIG-I-like receptor signaling pathway, Rap1 signaling pathway, Ras signaling pathway, Relaxin signaling pathway, Signaling pathways regulating pluripotency of stem cells, Sphingolipid signaling pathway, T cell receptor signaling pathway, TGF-beta signaling pathway, TNF signaling pathway, Thyroid hormone signaling pathway, Toll-like receptor signaling pathway, VEGF signaling pathway, Wnt signaling pathway, cAMP signaling pathway, cGMP-PKG signaling pathway, mTOR signaling pathway, p53 signaling pathway. After preprocessing the dataset, 17,259 gene-gene interactions for 1,489 genes were obtained.

### 2.4 Drug-Target interactions from DrugBank database

Drug-target information was extracted from the DrugBank[31] database (version 5.1.5, released 2020-01-03). In total, 15,263 drug-target interactions were obtained for 5435 drugs/investigational agents and 2775 targets. Further, 17 drugs with known targets to the 1489 genes previously identified were selected for use in our model for NCI ALMANAC[27], specifically: Celecoxib, Cladribine, Dasatinib, Docetaxel, Everolimus, Fulvestrant, Gefitinib, Lenalidomide, Megestrol acetate, Mitotane, Nilotinib, Paclitaxel, Romidepsin, Sirolimus, Thalidomide, Tretinoin, Vorinostat. And 10 drugs with known targets to the 1489 genes previously identified were selected for use in our model, specifically: Dasatinib, Erlotinib, Lapatinib, Sorafenib, Sunitinib, Vorinostat, geldanamycin, metformin, paclitaxel, vinblastine.

### 2.5 Architecture design of M3NetFlow

**Figure 2** shows the schematic architecture of the proposed ***M3NetFlow*** model. The model input parameters are: $\mathcal{X} = \{(X^{(1)}, T^{(1)}), (X^{(2)}, T^{(2)}), \dots, (X^{(m)}, T^{(m)}), \dots, (X^{(M)}, T^{(M)})\}$ $(X^{(m)} \in \mathbb{R}^{n\times d}, T \in \mathbb{R}^{n\times 2}), A \in \mathbb{R}^{n\times n}, \mathcal{S} = \{S_1, S_2, \dots, S_p, \dots, S_P\}(S_p \in \mathbb{R}^{n_p\times n_p})$, $D_{in} \in \mathbb{R}^{n\times n}$, $D_{out} \in \mathbb{R}^{n\times n}$, where $M$ represents number of data points in the drug screening dataset. To predict the drug combination effect scores $Y$ $(Y \in \mathbb{R}^{M\times 1})$, we would like to build up the machine learning model $f(\cdot)$ with $f(\mathcal{X}, A, \mathcal{S}, D_{in}, D_{out}) = Y$, where $\mathcal{X}$ denotes all of the data points in the dataset and $(X^{(m)}, T^{(m)})$ is $m$-th data points in the dataset, where $X^{(m)}$ denotes the node features matrix with $n$ nodes of $d$ features and $T^{(m)}$ denotes the one-hot encoding of the number of drugs targeted on those $n$ nodes. The matrix $A$ is the adjacency matrix that demonstrates the node-node interactions, and the element in adjacency matrix $A$ such as $a_{ij}$ indicates an edge from $i$ to $j$, and $\mathcal{S}$ is a set of subgraphs that partition the whole graph adjacent matrix $A$ into multiple subgraphs with $S_p \in \mathbb{R}^{n_p\times n_p}$ of nodes interactions between its internal $n_p$ nodes and each subgraph has its own corresponding subgraph node feature matrix $X_p \in \mathbb{R}^{n_p\times d}$. $D_{in}$ is an in-degree diagonal matrix for nodes in directed graph, and $D_{out}$ is an out-degree diagonal matrix for nodes in directed graph. In this certain model, we have 1,489 genes plus the 17 drugs, with $n = 1506$ nodes in the NCI ALMANAC dataset and we have 1,489 genes plus the 10 drugs with $n = 1499$ nodes in the O'Neil data set, where each node has $d = 8$ features.

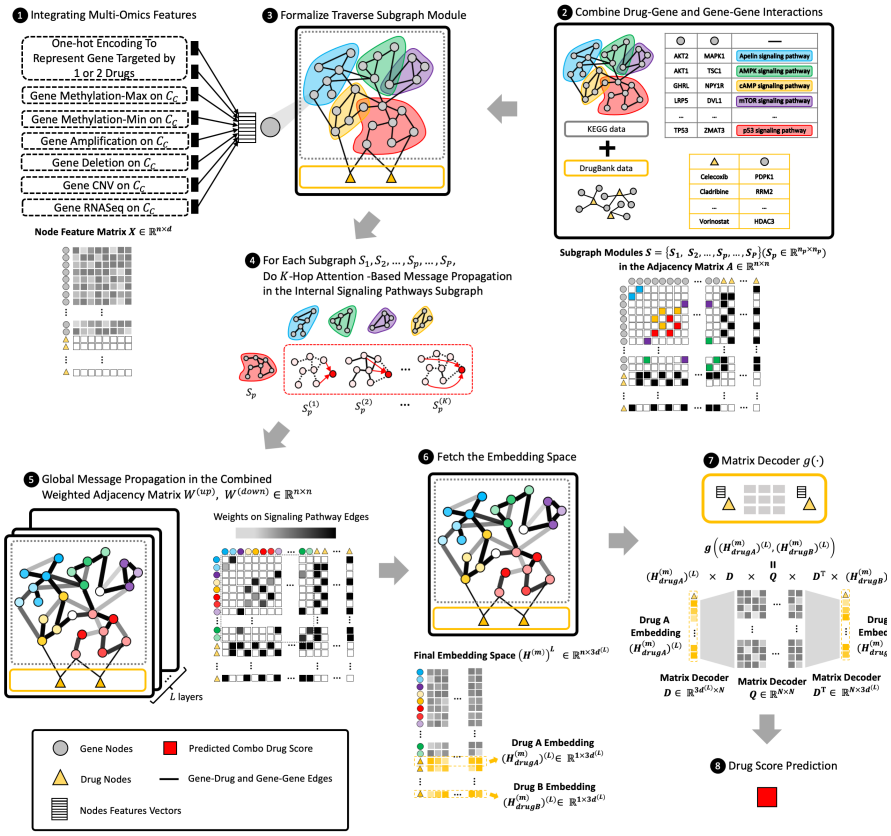


**Figure 2.** Architecture of ***M3NetFlow***

***Network Modular and Multi-hop Message Propagation*** In the graph message passing stages of our architecture (see **Figure 2** step3 and step4), we introduced the multi-scale design, i.e., the local network module/subgraph module message passing stage for each signaling pathway, and the global message passing stage. The multi-scale design will ensure the message fully interacts in the internal subgraph. Moreover, we opted for a $K$-hop attention-based graph neural network because it further allows for the consideration of longer distance information, which can be interpreted as signaling flow in a single message propagation layer. With those initial embedding features $X^{(m)} \in \mathbb{R}^{n \times d}, (m = 1, 2, \ldots, M)$, the $K$-hop attention-based graph neural network was built to incorporate the long-distance information for each subgraph with

$$\left(\alpha_{ij}^{(m)}\right)_p^{(k)} = \frac{\exp\left(\mathrm{LeakyReLU}(a[W(X_p^{(m)})_i || W(X_p^{(m)})_j])\right)}{\sum_{q \in \mathcal{N}_i^{(k)}} \exp\left(\mathrm{LeakyReLU}(a[W(X_p^{(m)})_i || W(X_p^{(m)})_q])\right)} \tag{1}$$

$$(H_p^{(m)})_i = \frac{1}{KQ_i} \sum_k^K \sum_{q \in \mathcal{N}_i^{(k)}} [(\alpha_{iq}^{(m)})_p^{(k)} W(X_p^{(m)})_q] \tag{2}$$

, where $\left(\alpha_{ij}^{(m)}\right)_p^{(k)}$ is the attention score between node $i$ and node $j$ in the $k$-th kop of the subgraph $S_p$ for the data point $m$ and $\mathcal{N}_i^{(k)}$ is the neighbor nodes calculated by the adjacency matrix in $k$-th hop $A^{(k)}$ for the node $i$ (See **Appendix A** for the algorithm of $k$-th hop adjacency matrix). The linear transformation vector $a \in \mathbb{R}^{2d'}$ was also defined. At the same time, the linear transformation for features of each node will be defined as $W \in \mathbb{R}^{d \times d'}$. And $(X_p^{(m)})_t, (H_p^{(m)})_t$ represent the feature and updated feature of node $t$ $(t = 1, 2, \ldots, n)$. Aside from that, $Q_t$ represents the number of signaling pathway the node $t$ $(t = 1, 2, \ldots, n)$ belongs to. And above formula shows the calculation for the 1-head attention. The number of head $h$ will be modified in the model, and the node embeddings for each subgraph will take the average of embeddings in every head attention.

***Global Bi-directional Message Propagation*** Following the message propagation in the multiple internal subgraphs, the global weighted bi-directional message propagation will be performed, where nodes-flow contains both 'upstream-to-downstream' (from up-stream signaling to drug targets) and 'downstream-to-upstream' (from drug targets to down-stream signaling) (see **Figure 2** step 5). Before the global level message propagation, the node feature for data point $m$ in each subgraph $H_p^{(m)}$ $(p = 1, 2, \dots, P)$ will be combined into a new unified node features matrix $\left(H^{(m)}\right)^{(0)}$ as the initial node features with

$$\left(H^{(m)}\right)^{(0)}{}_i = \frac{1}{\sum_{p=1}^{P} I[\mathcal{V}_i \in S_p]} \sum_{p=1}^{P} \left(H_p^{(m)}\right)_i \cdot I[\mathcal{V}_i \in S_p] \tag{3}$$

, where $V_i$ represents the node/vertex $i$ in the graph and $I\left[V_i \in S_p\right]$ is the indicator function, whose value will be one if $\mathcal{V}_i \in S_p$. In the end, the initial node features for global level propagation will be $\left(H^{(m)}\right)^{(0)} \in \mathbb{R}^{n \times d^{(0)}}, d^{(0)} = d' + d$. The weight matrices are denoted as: $W_{up}^{(l)} \in \mathbb{R}^{n \times n},\ W_{down}^{(l)} \in \mathbb{R}^{n \times n},\ (l = 1, 2, \dots, L)$ respectively for $L$ global message propagation layers. And then the weight adjacency matrices will be denoted as:

$$A'^{(l)}_{up} = W_{up}^{(l)} \cdot A \tag{4}$$

$$A'^{(l)}_{down} = W_{down}^{(l)} \cdot A^{\mathrm{T}} \tag{5}$$

, where for each $A'^{(l)}_{up} \in \mathbb{R}^{n \times n},\ A'^{(l)}_{down} \in \mathbb{R}^{n \times n},\ (l = 1,2, \dots, L)$, parameters for drug-gene relations values are the same with those in original adjacency matrix. Afterwards, the mean aggregation of each node neighbors' features from upstream to downstream is achieved by:

$$\left(H_{up_neigh}^{(m)}\right)^{(l)} = D_{in}^{-1} (A'^{(l)}_{up})^{\mathrm{T}} \left(H^{(m)}\right)^{(l-1)} Z_{up}^{(l)} \tag{6}$$

And the mean aggregation of each node neighbors' features from downstream to upstream is achieved by:

$$\left(H_{down_neigh}^{(m)}\right)^{(l)} = D_{out}^{-1} (A'^{(l)}_{down})^{\mathrm{T}} \left(H^{(m)}\right)^{(l-1)} Z_{down}^{(l)} \tag{7}$$

And biased transformation of nodes themselves and their features is achieved by:

$$\left(H_{self}^{(m)}\right)^{(l)} = \left(H^{(m)}\right)^{(l-1)} B^{(l)} \tag{8}$$

, where the $Z_{up}^{(l)} \in \mathbb{R}^{d^{(l-1)} \times d^{(l)}}, Z_{down}^{(k)} \in \mathbb{R}^{d^{(l-1)} \times d^{(l)}}$ are the linear transformation matrix for each layer $l$, $(l = 1, 2, \dots, L)$. For $l = 1, 2, \dots, L$, the model used normalization function $N: \mathbb{R}^{n \times 3d^{(l)}} \to \mathbb{R}^{n \times 3d^{(l)}}$ and LeakyReLU function with parameter $\alpha$ to map concatenated nodes features $\left(H_{concat}^{(m)}\right)^{(l)} = \left(\left(H_{up_neigh}^{(m)}\right)^{(l)} | \left(H_{down_neigh}^{(m)}\right)^{(l)} | \left(H_{self}^{(m)}\right)^{(l)}\right) \in \mathbb{R}^{n \times 3d^{(l)}}$ to $H^{(l)} \in \mathbb{R}^{n \times 3d^{(l)}}$ with the equation

$$(H^{(m)})^{(l)} = \mathrm{LeakyReLU}(N(\left(H_{concat}^{(m)}\right)^{(l)})) \tag{9}$$

where normalization function $N$ enable L2 normalization for demo matrix $V \in \mathbb{R}^{p \times q}$ in row axis with the equation

$$v'_{ij} = \frac{v_{ij}}{\sqrt{\sum_{j=1}^{q} v_{ij}^2}} \tag{10}$$

where $v'_{ij}$ is the element of new matrix $V' \in \mathbb{R}^{p \times q}$.

***Prediction by decoding drug node embeddings*** Therefore, the output of $l$-th layer for global message propagation is $(H^{(m)})^{(l)}$. After obtaining the embedded nodes features with $(H^{(m)})^{(L)} \in \mathbb{R}^{n \times 3d^{(L)}}$, the embedded nodes features of drug_A are obtained as: $(H_{drugA}^{(m)})^{(L)} \in \mathbb{R}^{1 \times 3d^{(L)}}$, and the drug_B embedded nodes features are obtained as: $(H_{drugB}^{(m)})^{(L)} \in \mathbb{R}^{1 \times 3d^{(L)}}$. Based on the idea of decagon decoder[34], the prediction of combo score will be calculated in the following equation:

$$g\left((H_{drugA}^{(m)})^{(L)}, (H_{drugB}^{(m)})^{(L)}\right) = (H_{drugA}^{(m)})^{(L)} D U D^{\mathrm{T}} ((H_{drugB}^{(m)})^{(L)})^{\mathrm{T}} \tag{11}$$

where $D \in \mathbb{R}^{3d^{(L)} \times E}$ and $U \in \mathbb{R}^{E \times E}$ are trainable decoder matrices (see **Figure 2** step 6 and step 7).

**2.6 Drug combination response interpretation using the subgraph attention**

As aforementioned, we added the $K$-hop subgraph message propagation with attention in each subgraph, and the attentions or weights on each edge can potentially indicate and interpret the signaling flow on the signaling network to affect the drug combination response. Specifically, putting the attentions/weights on each edge will use the trained parameters $a \in \mathbb{R}^{2d'}$ and $W \in \mathbb{R}^{d \times d'}$. And for cell lines set, the set $\mathcal{C} = \{C_1, C_2, \dots, C_r, \dots, C_R\}$ contains $R = 39$ cell lines, and we can choose the input node feature matrix within a specific cell line to form the cell-line specific attention matrix in $k$-hop by

$$(\alpha_{ij}^{(C_r)})^{(k)} = \sum_{p=1}^{P} \left(\alpha_{ij}^{(m)}\right)_p^{(k)}, \left(\forall X^{(m)} \in C_r\right) \tag{12}$$

, where $(\alpha_{ij}^{(C_r)})^{(k)}$ will be the element in $\left(\mathcal{A}^{(C_r)}\right)^{(k)}$, the $k$-hop attention matrix for cell line $C_r$. And above formula shows the calculation for the 1-head attention matrix. For the multi-head attention mechanism, the averaged value will be used for the attention matrix.

## 3. Results

### 3.1 Experimental Setup

In order to validate the proposed model, a 5-fold cross-validation was used. Specifically, <drug_A, drug_B, cell_line name, combo score>, a 4-element vector, represents a model input data point. In total, there are 2788 model input data points for the NCI ALMANAC dataset and 1008 model input data points for the O'Neil dataset. In each fold model, 4 folds of the data points will be used as the training dataset, and the rest 1-fold will be used as the testing dataset. For those 1489 genes, their 8 features indicate the RNA sequence number, copy number variation, gene amplification, gene deletion, gene methylation maximum and minimum value and whether they have a connection spanning drug_A and drug_B respectively. For each drug pair, their 8 features were initiated with zeros. Furthermore, to indicate connections between nodes, adjacency matrices were also created. Those matrices were formed from the KEGG, which contains gene pairs with sources and destinations. In this way, the graphs generated were directed graphs**.** For drug-gene edges, the connections are bidirectional, which means that in adjacency matrices, those elements are symmetric.

### 3.2 Hyperparameters

Subsequently, a model was developed by using pytorch and torch geometric. Using the Adam optimizer, the learning rate started at 0.002 and was reduced equally within each batch for a certain epoch stage. For epochs range [1, 10], [11, 20], [21, 30], [31, 60], the learning rate reduced from [0.002, 0.0008], [0.0008, 0.0006], [0.0006, 0.0005], [0.0005, 0.0001] separately in each range. And the epochs after 60 will keep the learning rate at 0.0001. Adam optimizer was chosen for optimization with eps=1e-7 and weight_decay=1e-20. We empirically set the ***K-hop Subgraph Message Propagation*** part with $K = 3$ and the ***Global Bi-directional Message Propagation*** part with $L = 3$, and the initial input feature dimension was: $d = 8$, $d' = 1$ with setting the number of head $h = 3$. Afterward, the feature dimensions will vary at the different layers and will be denoted by: $(d^{(1)}, d^{(2)}, \dots, d^{(l)}, \dots, d^{(L)})$. At the global message propagation part, the layer will concatenate biased node features and transformed node features of the previous layer for both upstream and downstream, generating concatenated dimensions for the output dims being $3 \times d^{(l)}$ in $l$-th layer. Output dims of the previous layer served as the input dims for the current layer, as follows: (1) First layer (input dims, output dims): $(d^{(0)}, 3d^{(1)})$; (2) Second layer (input dims, output dims): $(3d^{(1)}, 3d^{(2)})$; (3) Third layer (input dims, output dims): $(3d^{(2)}, 3d^{(3)})$. The final embedded drug node dims were $3d^{(3)}, (L = 3)$. Hence, for dimensions of node embedding in each layer, we have $d^{(0)} = 8$, $d^{(1)} = 8, d^{(2)} = 8, d^{(3)} = 24$ and decoder trainable transformation matrix dims: $D \in \mathbb{R}^{3d^{(L)} \times E}$ and $U \in \mathbb{R}^{E \times E}$ were used as trainable decoder matrices, with changeable parameters of $E$ to adapt model performance and the model used $E = 150$. As for the LeakyReLU function, the parameter $\alpha = 0.1$.

### 3.3 *M3NetFlow* significantly improves drug combination prediction accuracy.

To evaluate the model performance in terms of synergy score prediction for drug combinations, we conducted 5-fold cross-validation. As shown in **Table 4**, the average prediction (using the Pearson correlation coefficient), was about 61% Pearson correlation using the test data in the NCI ALMANAC dataset and was about 64% Pearson correlation using the test data in the O'Neil dataset. These prediction results are comparable with existing deep learning models[35,36]. **Figure S1-S5** shows the scatterplots and boxplots of the predicted and experimental synergy scores on the 5 training and test datasets. The results indicated the feasibility of drug combination synergy score prediction using a graph neural network.

**Table 4.** 5-fold Cross Validation of the proposed ***M3NetFlow*** model in the NCI ALMANAC dataset

| Dataset | Training MSE Loss | Training Pearson Correlation | Test MSE Loss | Test Pearson Correlation |
|---|---|---|---|---|
| 1st dataset | 24.910785 | 78.8935% | 42.420434 | 60.0966% |
| 2nd dataset | 34.32198 | 69.0788% | 42.366917 | 60.1075% |

| | | | | |
|---|---|---|---|---|
| 3rd dataset | 32.42659 | 71.7287% | 38.854818 | 61.7941% |
| 4th dataset | 34.696549 | 69.2506% | 40.94755 | 60.0938% |
| 5th dataset | 31.875111 | 70.3966% | 47.63544 | 61.5159% |
| **Average** | **31.646141** | **71.9670%** | **42.450617** | **60.5340%** |

**Table 5.** 5-fold Cross Validation of the proposed ***M3NetFlow*** model in the O'Neil dataset

| Dataset | Training MSE Loss | Training Pearson Correlation | Test MSE Loss | Test Pearson Correlation |
|---|---|---|---|---|
| 1st dataset | 168.836656 | 68.2072% | 134.640407 | 62.9468% |
| 2nd dataset | 144.522245 | 73.3936% | 144.548363 | 63.8356% |
| 3rd dataset | 69.703043 | 85.8798% | 283.912207 | 61.6114% |
| 4th dataset | 87.725931 | 83.6393% | 180.135809 | 69.0636% |
| 5th dataset | 151.858177 | 71.0429% | 183.728567 | 64.3278% |
| **Average** | **124.508876** | **76.3542%** | **185.393071** | **62.6482%** |

Moreover, we also compared our proposed model M3NetFlow with other deep learning models (see **Table 6** and **Figure 3a**), which included the GCN[37] and Graph Attention network[38] (GAT). Based on this GCN and GAT model, both models were implemented with the decoder matrices to predict the drug effects, which improved the performance of both models, named as GCN Decoder and GAT Decoder.

**Table 6.** Comparisons between GCN Decoder, GAT Decoder and M3NetFlow of average MSE loss and Pearson correlation of 5-fold cross-validation in NCI ALMANAC dataset.

| Dataset | Training MSE Loss | Training Pearson Correlation | Test MSE Loss | Test Pearson Correlation |
|---|---|---|---|---|
| GCN Decoder | 43.0061078 | **58.9086%** | 48.5379624 | **51.93114%** |
| GAT Decoder | 43.1162382 | **58.65168%** | 50.9773424 | **49.1606%** |
| M3NetFlow | 31.646203 | **71.86964%** | 42.4450318 | **60.72158%** |

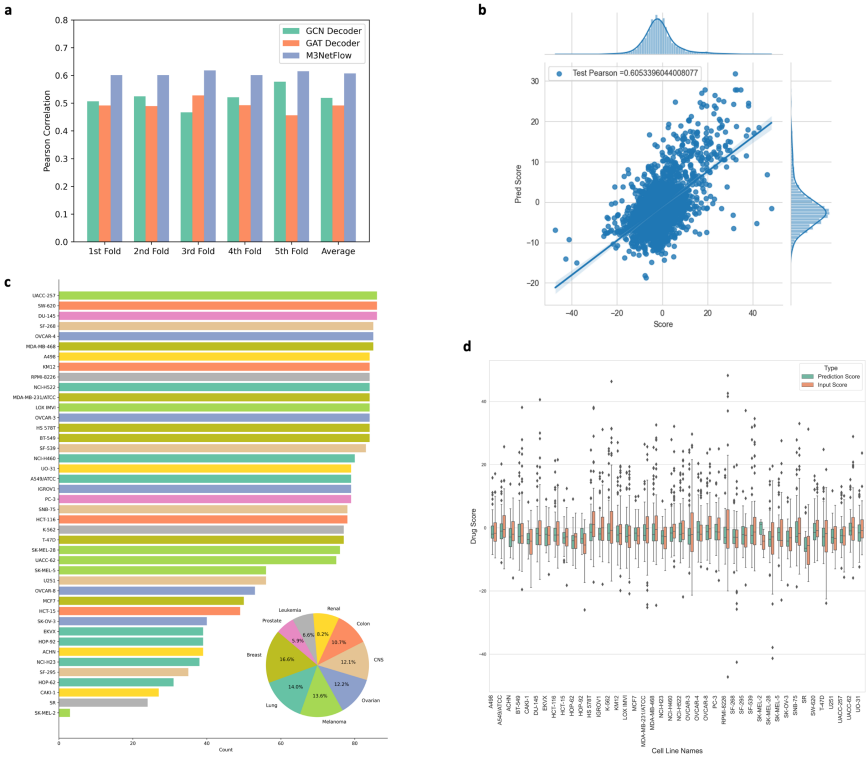


**Figure 3**. Model performance and overview of input dataset NCI ALMANAC **a.** Pearson correlation comparisons for GCN Decoder, GAT Decoder and ***M3NetFlow*** models in each split of 5-fold cross-validation for NCI ALMANAC dataset. **b.** Scatter plot of the model with data points in whole NCI ALMANAC dataset. **c.** Distributions of all cell lines in whole NCI ALMANAC dataset. **d.** Box plots across all cell lines in the NCI ALMANAC dataset

### 3.4 *M3NetFlow* ranks important targets via attention scores

To interpret the mechanism of drug effects, the top or most important genes in each cell line were selected as the potential drug targets from the NCI ALMANAC[27] dataset. Based on the attention matrix of each cell line from 1-hop neighbor nodes in each fold of cross validation, the average attention matrix in 1-hop was calculated. Therefore, the weighted degree of each node of each cell line will be calculated based on the attention with

$$\mathcal{D}_g^{(C_r)} = \sum_i^n \left(\bar{\mathcal{A}}_{ig}^{(C_r)}\right)^{(1)} + \sum_j^n \left(\bar{\mathcal{A}}_{gj}^{(C_r)}\right)^{(1)} \tag{13}$$

, where $\left(\bar{\mathcal{A}}_{ig}^{(C_r)}\right)^{(1)}$ is the element of 1-hop averaged 5-fold attention matrix from a cell line $C_r$ in the $i$-th row and $j$-th column and $\mathcal{D}_g^{(C_r)}$ is the node degree (importance score) for gene $g$ in the cell line $C_r$.

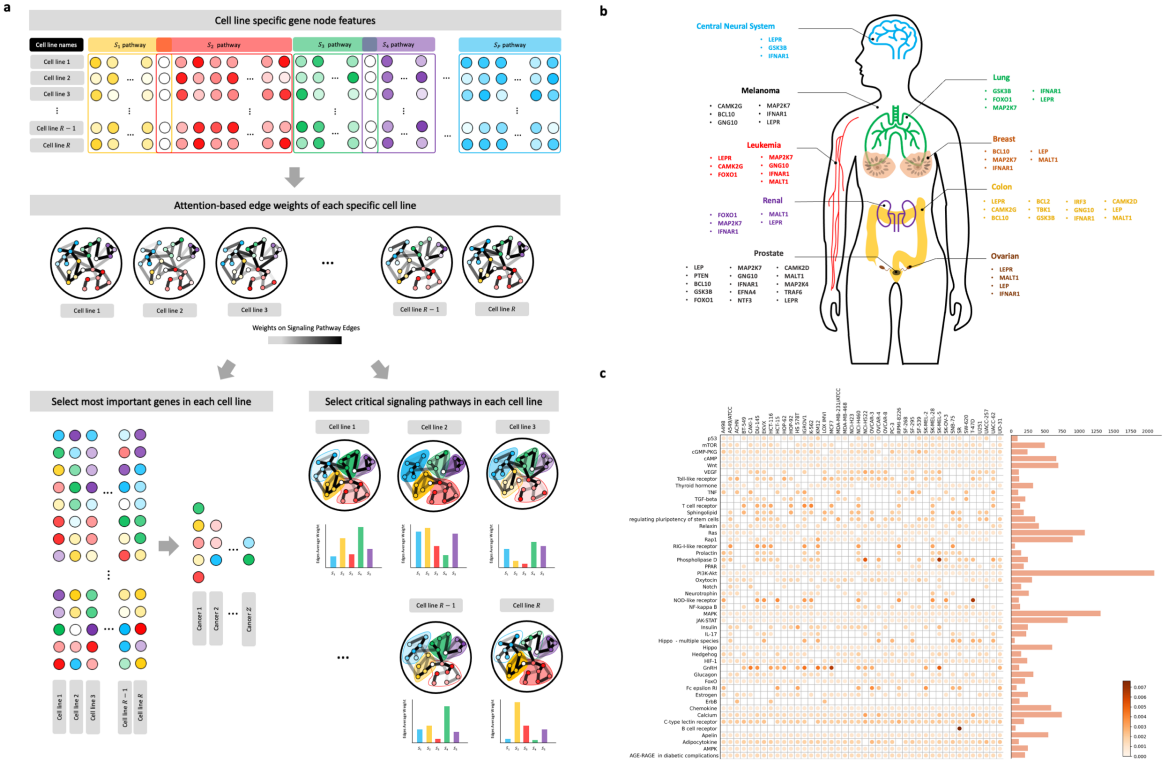


**Figure 4**. Downstream analysis of important genes and signaling pathways **a.** Procedures of identifying the important genes in different cell lines and cancers, and identifying the important signaling pathways in different cell lines **b.** Overlapped genes for the cell lines in each type of cancer **c.** Heatmap of sum of edge weights in different signaling pathways in whole NCI ALMANAC dataset

Combining all the node degree vectors from each cell line, the matrix $\mathcal{D} \in \mathbb{R}^{n \times R}$ will be generated. However, some of genes may show the importance of relatively higher scores in every cell line, which weakens the analysis of the cell line-specific drug targets. In this way, the idea of reweighting the gene importance score was created based on the TF-IDF to reweight the node degree in each cell line with

$$\mathcal{W}_g = Log\left(\frac{R+1}{S_g + 0.01}\right) \tag{14}$$

$$\mathcal{D}'_g = \mathcal{W}_g \cdot \mathcal{D}_g \tag{15}$$

, where $S_g$ is the number of cell lines with a higher node degree than the threshold $\mathcal{S}$ in all 39 cell lines. Here, the 95 percentile of node degree values in the whole matrix $\mathcal{D}$ was used for $\mathcal{S}$. Finally, the reweighted node degree (importance score) matrix $\mathcal{D}'$ will be generated. Based on the reweighted matrix $\mathcal{D}'$, the node degree (gene importance score) will be defined in each cell line (check **Figure 4a**). To validation our conclusions on the gene importance generated from the proposed model and reweighting process borrowed from TF-IDF, the following statistical analysis was made.

### 3.5 Effective drugs are statistically associated with the top-ranked key targets

For each cell line, the highest 5, and lowest 5 drug scores and their corresponding drugs and drug-targeted genes were collected (check **Figure 5a**). Comparing the targeted genes by drugs from the highest 5 drug scores and lowest drug scores, the differences are statistically significant with p values in more than half cell lines. Since the SK-MEL-2 cell line only has 3 drug screen data points and all of them are smaller than 0, this cell line was excluded from the statistical analysis.

In total, there are 27 out of 41 (~65%) cell lines have p values smaller than 0.1 (filled with deep orange color in **Table S1**) and 32 out of 41 (~78%) cell lines have p value smaller than 0.3 (filled with light orange color in **Table S1**). And 37 out of 41 (~90%) cell lines have higher gene importance scores on both mean and median values (filled with light green in **Table S1**).

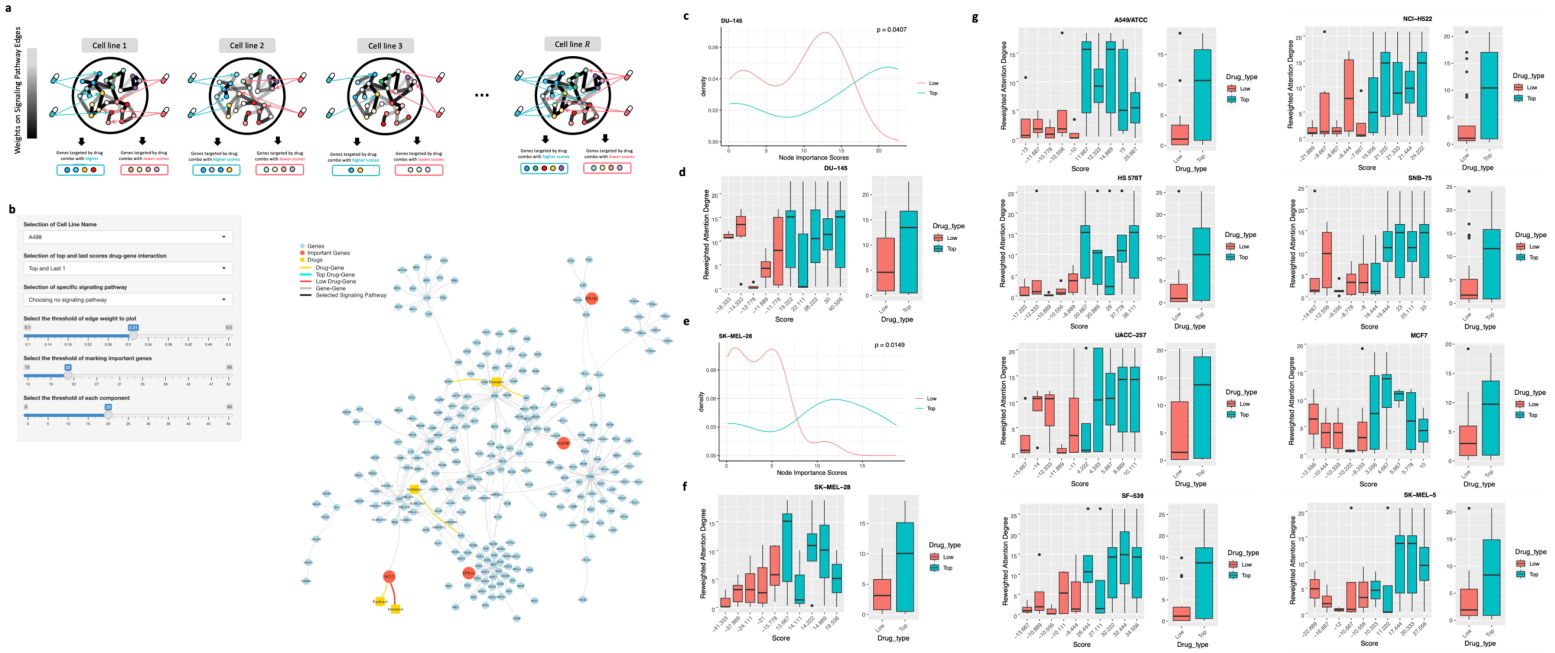


**Figure 5.** Validation and visualization through important node analysis **a.** Procedures of analyzing the important genes targeted by by the highest 5 and lowest 5 drug combinations **b.** Visualization tool ***NetFlowVis*** for core signaling network interactions of cell line DU-145 **c-f.** Density and box plots by comparing the genes targeted by the highest 5 and lowest 5 drug combinations in DU-145 and SK-MEL-28 cell lines g. Selected box plots by comparing the genes targeted by the highest 5 and lowest 5 drug combinations in cell lines with 8 lowest p-values (check Table S1 for details of the p-values by comparing the gene important score lists targeted by the highest 5 and lowest 5 drug combinations

**Figure S6-S8** provided more explicit evidence of a trend where genes targeted by drugs with higher scores exhibited higher node degrees (gene importance scores). In each cell line, the left boxplots compared the node degree distribution of genes targeted by the top 5 drugs with the highest scores against those targeted by the bottom 5 drugs with the lowest scores. The right boxplots in each cell line represented the overall comparisons between the top 5 and bottom 5 drug scores. Consequently, we could assign prior targets to genes with the highest node degree (gene importance scores) using our calculations. This allowed us to rank the genes in each cell line according to their node degree (importance scores), which can be found in **Table S2**. We generated a list of the top 20 genes in each cell line, available in **Table S3**. Subsequently, we identified overlapping genes among the cell lines of the same cancer type based on the top 20 genes in each line, as depicted in **Figure 4b**.

### 3.6 *NetFlowVis*: a tool to visualize the analysis results

In order to visualize the results and gain a better understanding of the underlying mechanism of drug effects, we generated a core network of signaling interactions by applying a threshold. To enhance the interactions, we utilized the RShiny package and developed a visualization tool called ***NetFlowVis*** (The website link has been provided in the code availability part). This tool allows users to control the edge threshold and set a minimum number of nodes in each network component. Users have the option to select the desired cell line for visualization and mark specific signaling pathways for detailed analysis. As an example, **Figure 5b** demonstrates the visualization of the cell line DU-145. Furthermore, by referring to the top 20 prior gene targets and their corresponding gene degree (node importance scores) for cell line DU-145, we observed that the majority of gene targets were included in the filtered signaling network interactions depicted in **Figure 5b**.

By utilizing the ***NetFlowVis*** visualization tool, users could take control of the size of the core signaling network. In the case of the DU-145 cell line, an edge threshold of 0.31 was selected, which corresponded to approximately the 98.5th percentile for edge weight among all edges. Additionally, a threshold of 23 was chosen for marking important nodes/genes in red, representing ~99.9th percentile for node importance scores across all nodes. To gain a deeper understanding of the underlying mechanism, the interactions between drugs and genes were analyzed, with the highest and lowest drug-gene interactions being represented by the colors purple and green, respectively. The analysis revealed

that the drug ***Paclitaxel*** targeted the important gene ***BCL2***, while the drug ***Celecoxib*** targeted less essential genes (see **Figure 5b**). This finding demonstrated that targeting the important genes selected by the ***M3NetFlow*** model led to higher drug scores, as depicted in **Figure 5c-f**. Moreover, by choosing a specific signaling pathway within the core signaling network, it was observed that the ***IL-17 signaling pathway*** was downstream of the critical gene ***TRAF6***. Interestingly, experimental validation has shown the involvement of the ***IL-17 signaling pathway*** in promoting migration and invasion of ***DU-145*** cell lines through the upregulation of ***MTA1*** expression in prostate cancer[39].

## 4. Discussion and conclusion

Along with the advancement of next-generation sequencing (NGS) technology, multi-omics data, like gene expression, copy number variation, methylation, genetic mutation, and microRNA, are being generated to characterize the dysfunctional molecules and complex signaling pathways in cancer patients. The valuable multi-omics data provides the basis for personalized medicine or precision medicine prediction. Drug combinations are being investigated to reduce the drug resistance influenced by the complex signaling networks in cancer. However, it remains an open problem and challenging because of the mysterious, heterogeneous molecular (should change the description) of drug combinations.

Deep learning models have been reported to predict drug combinations by simply concatenating a large number of multi-omics data and chemical structure information using the deep neural network (DNN) model. One limitation of the existing DNN models is the lack of complex and biologically meaningful gene regulatory relationships. Thus, it is hard to interpret these models to support the actual clinical use, which requires clinical experts to understand the molecular mechanisms used in AI prediction.

In this study, we aim to explore the feasibility of building an interpretable AI model using the proposed ***M3NetFlow*** model. The model is built on a core cancer signaling network consisting of 1489 genes, with gene expression and copy number data derived from 48 core cancer signaling pathways. To incorporate a longer range of node information and form the cell-line-specific network, the ***(multi-)K-hop Subgraph Message Propagation*** was created. Aside from that, ***Global Bi-directional Message Propagation*** was designed to build the novel up-stream signaling flow (from up-stream signaling to drug targets), and the down-stream signaling flow (from drug targets to down-stream signaling), was designed using trainable weights of network edges. In the end, the most important genes for each cell line were calculated based on the attention weight from their connected edges in both upstream and downstream directions and a reweighting process was also used to reorder the important genes. From the statistical perspective, our model can sort the important genes aligning with drug scores.

Aside from that, the visualization tool ***NetFlowVis*** was developed for a better analysis of drug effects. By selecting the specific cell lines or drug pairs in certain cell lines, the critical paths with edge threshold and (should change the description) could be easily analyzed. In conclusion, the $\boldsymbol{K}$-hop subgraph propagation in the internal signaling pathways and the up-stream and down-stream signaling flow design for global propagation is novel and can potentially interpret the drug effects.

This is still an exploratory study to predict and investigate drug combination response using deep neural graph network models. There are some limitations to be further investigated. For example, more signaling pathways and protein-protein interactions should be considered to include more drugs and (should change the description) should be examined in detail. Moreover, the signaling pathways and signaling cascades of a large set of cell lines might be informative to further identify the important pan-cancer targets and cancer-specific targets. Whereas, it is not a trivial task to identify the statistically significant targets and signaling cascades. Also, the multi-omics data represents complex and multi-scale signaling cascades in cellular functions. Thus, it remains an open problem, and novel improved computational models are needed to model and uncover the roles of multi-omic and multi-gene interactions in disease development and drug response.

## Code Availability

Check the ***M3NetFlow*** code on the website: **https://github.com/heming-zhang/M3NetFlow**
Check the ***NetFlowVis*** app on the website: **https://m3netflow.shinyapps.io/NetFlowVis/**

## Author contributions

FL conceived this study. FL, HZ designed the model and prepared the manuscript. HZ implemented the model and analyzed the data and results. FL, HZ, PG, LD, WK, DD, RF, YC, PP discussed the method and results.

**References**


1. TCGA. *https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga*.
2. Barretina J, Caponigro G, Stransky N, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*. 2012;483(7391):603-607. doi:10.1038/nature11003
3. Yang W, Soares J, Greninger P, et al. Genomics of Drug Sensitivity in Cancer (GDSC): A resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res*. 2013;41(D1):955-961. doi:10.1093/nar/gks1111
4. Li F, Eteleeb A, Buchser W, et al. Weakly activated core inflammation pathways were identified as a central signaling mechanism contributing to the chronic neurodegeneration in Alzheimer's disease. *bioRxiv*. Published online 2021.
5. Li F, Oh I, Kumar S, et al. Loss of estrogen unleashing neuro-inflammation increases the risk of Alzheimer's disease in women. *bioRxiv*. Published online 2022:2022-2029.
6. Gazzoli I, Loda M, Garber J, Syngal S, Kolodner RD. A hereditary nonpolyposis colorectal carcinoma case associated with hypermethylation of the MLH1 gene in normal tissue and loss of heterozygosity of the unmethylated allele in the resulting microsatellite instability-high tumor. *Cancer Res*. 2002;62(14):3925-3928.
7. Baylin SB, Jones PA. Epigenetic determinants of cancer. *Cold Spring Harb Perspect Biol*. 2016;8(9):a019505.
8. Zhang H, Liu T, Zhang Z, et al. Integrated proteogenomic characterization of human high-grade serous ovarian cancer. *Cell*. 2016;166(3):755-765.
9. Stuart T, Butler A, Hoffman P, et al. Comprehensive Integration of Single-Cell Data. *Cell*. 2019;177(7):1888-1902.e21. doi:10.1016/j.cell.2019.05.031
10. Subramanian I, Verma S, Kumar S, Jere A, Anamika K. Multi-omics data integration, interpretation, and its application. *Bioinform Biol Insights*. 2020;14:1177932219899051.
11. Vaske CJ, Benz SC, Sanborn JZ, et al. Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM. *Bioinformatics*. 2010;26(12):i237-i245.
12. Browaeys R, Saelens W, Saeys Y. NicheNet: modeling intercellular communication by linking ligands to target genes. *Nat Methods*. 2020;17(2):159-162.
13. Hu Y, Peng T, Gao L, Tan K. CytoTalk: De novo construction of signal transduction networks using single-cell transcriptomic data. *Sci Adv*. 2021;7(16):eabf1356.
14. Zhou G, Li S, Xia J. Network-based approaches for multi-omics integration. *Computational methods and data analysis for metabolomics*. Published online 2020:469-487.
15. Zhang T, Zhang L, Payne P, Li F. Synergistic Drug Combination Prediction by Integrating Multi-omics Data in Deep Learning Models. *arXiv preprint arXiv:181107054*. Published online 2018.
16. Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:160902907*. Published online 2016.
17. Hamilton WL, Ying R, Leskovec J. Inductive representation learning on large graphs. *Adv Neural Inf Process Syst*. 2017;2017-Decem(Nips):1025-1035.
18. Veličković P, Casanova A, Liò P, Cucurull G, Romero A, Bengio Y. Graph attention networks. *6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings*. Published online 2018:1-12.
19. Xu K, Hu W, Leskovec J, Jegelka S. How powerful are graph neural networks? *arXiv preprint arXiv:181000826*. Published online 2018.
20. Wang T, Shao W, Huang Z, et al. MOGONET integrates multi-omics data using graph convolutional networks allowing patient classification and biomarker identification. *Nat Commun*. 2021;12(1):3445.
21. Li X, Ma J, Leng L, et al. MoGCN: a multi-omics integration method based on graph convolutional network for cancer subtype analysis. *Front Genet*. 2022;13:806842.
22. Gao H, Zhang B, Liu L, Li S, Gao X, Yu B. A universal framework for single-cell multi-omics data integration with graph convolutional networks. *Brief Bioinform*. 2023;24(3):bbad081.
23. Rajadhyaksha N, Chitkara A. Graph Contrastive Learning for Multi-omics Data. *arXiv preprint arXiv:230102242*. Published online 2023.
24. Li G, Muller M, Thabet A, Ghanem B. DeepGCNs: Can GCNs go as deep as CNNs? *Proceedings of the IEEE International Conference on Computer Vision*. 2019;2019-Octob:9266-9275. doi:10.1109/ICCV.2019.00936
25. Cai C, Wang Y. A Note on Over-Smoothing for Graph Neural Networks. Published online 2020.
26. Salton G. Modern information retrieval. *(No Title)*. Published online 1983.
27. Holbeck SL, Camalier R, Crowell JA, et al. The National Cancer Institute ALMANAC: A comprehensive screening resource for the detection of anticancer drug pairs with enhanced therapeutic activity. *Cancer Res*. 2017;77(13):3564-3576. doi:10.1158/0008-5472.CAN-17-0489

28. Van Der Meer D, Barthorpe S, Yang W, et al. Cell Model Passports - a hub for clinical, genetic and functional datasets of preclinical cancer models. *Nucleic Acids Res*. 2019;47(D1):D923-D929. doi:10.1093/nar/gky872
29. Barretina J, Caponigro G, Stransky N, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*. 2012;483(7391):603-607. doi:10.1038/nature11003
30. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. Published online 1999:28. doi:10.1093/nar/27.1.29
31. Wishart DS, Feunang YD, Guo AC, et al. DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Res*. 2018;46(D1):D1074-D1082. doi:10.1093/nar/gkx1037
32. O'Neil J, Benita Y, Feldman I, et al. An unbiased oncology compound screen to identify novel combination strategies. *Mol Cancer Ther*. 2016;15(6):1155-1162. doi:10.1158/1535-7163.MCT-15-0843
33. Zagidullin B, Aldahdooh J, Zheng S, et al. DrugComb: An integrative cancer drug combination data portal. *Nucleic Acids Res*. 2019;47(W1):W43-W51. doi:10.1093/nar/gkz337
34. Zitnik M, Agrawal M, Leskovec J. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics*. 2018;34(13):i457-i466. doi:10.1093/bioinformatics/bty294
35. Preuer K, Lewis RPI, Hochreiter S, Bender A, Bulusu KC, Klambauer G. DeepSynergy: Predicting anti-cancer drug synergy with Deep Learning. *Bioinformatics*. 2018;34(9):1538-1546. doi:10.1093/bioinformatics/btx806
36. Zhang T, Zhang L, Payne P, Li F. Synergistic Drug Combination Prediction by Integrating Multi-omics Data in Deep Learning Models. *arXiv preprint arXiv:181107054*. Published online 2018.
37. Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. *5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings*. Published online 2017:1-14.
38. Veličković P, Casanova A, Liò P, Cucurull G, Romero A, Bengio Y. Graph attention networks. *6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings*. Published online 2018:1-12.
39. Guo N, Shen G, Zhang Y, Moustafa AA, Ge D, You Z. Interleukin-17 promotes migration and invasion of human cancer cells through upregulation of MTA1 expression. *Front Oncol*. 2019;9(JUN):1-11. doi:10.3389/fonc.2019.00546