Bioinformatics, 2025, **41(1)**, btae750
https://doi.org/10.1093/bioinformatics/btae750
Advance Access Publication Date: 15 January 2025
Original Paper

OXFORD

Data and text mining

Heterogeneous entity representation for medicinal synergy prediction

Jiawei Wu[1,†], Jun Wen[2,†], Mingyuan Yan[1,†], Anqi Dong[3], Shuai Gao[4], Ren Wang[5], Can Chen[6,7,*]

[1]School of Medicine, National University of Singapore, Singapore 119077, Singapore
[2]Harvard Medical School, Harvard University, Boston, MA 02115, United States
[3]Division of Decision and Control Systems and Department of Mathematics, KTH Royal Institute of Technology, Stockholm SE-100 44, Sweden
[4]Harvey Cushing Neuro-Oncology Laboratories, Department of Neurosurgery, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, United States
[5]Department of Electrical and Computer Engineering, Illinois Institute of Technology, Chicago, IL 60616, United States
[6]School of Data Science and Society, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, United States
[7]Department of Mathematics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, United States

*Corresponding author. School of Data Science and Society, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, United States. E-mail: canc@unc.edu.

[†]Equal contribution.

Associate Editor: Zhiyong Lu

Abstract

Motivation: Forecasting the synergistic effects of drug combinations facilitates drug discovery and development, especially regarding cancer therapeutics. While numerous computational methods have emerged, most of them fall short in fully modeling the relationships among clinical entities including drugs, cell lines, and diseases, which hampers their ability to generalize to drug combinations involving unseen drugs. These relationships are complex and multidimensional, requiring sophisticated modeling to capture nuanced interplay that can significantly influence therapeutic efficacy.

Results: We present a novel deep hypergraph learning method named Heterogeneous Entity Representation for MEdicinal Synergy (HERMES) prediction to predict the synergistic effects of anti-cancer drugs. Heterogeneous data sources, including drug chemical structures, gene expression profiles, and disease clinical semantics, are integrated into hypergraph neural networks equipped with a gated residual mechanism to enhance high-order relationship modeling. HERMES demonstrates state-of-the-art performance on two benchmark datasets, significantly outperforming existing methods in predicting the synergistic effects of drug combinations, particularly in cases involving unseen drugs.

Availability and implementation: The source code is available at https://github.com/Christina327/HERMES.

1 Introduction

Exploring drug combinations leads to a promising avenue for enhancing cancer treatment efficacy while minimizing toxicity and adverse reactions in modern medicine (Jia *et al.* 2009, Wang *et al.* 2012, Csermely *et al.* 2013, Foucquier and Guedj 2015). Combination therapies, involving multiple drugs administered as a single treatment regimen, offer potential benefits over traditional single-drug approaches, particularly under cancer and tumor treatment contexts (Chou 2006, O'Neil *et al.* 2016). Not only do they hold the promise of greater therapeutic efficacy, but also present an opportunity to mitigate host toxicity and unwanted side effects, as the doses of drug combinations are often sub-mutagenic compared to individual drug doses. However, optimizing drug combinations can be challenging, as poorly chosen combinations may lead to adverse effects and sub-optimal outcomes (Hecht *et al.* 2009, Tol *et al.* 2009). Thus, there is a critical need to identify precise synergistic drug pairs tailored to different cancer types.

Historically, the identification of effective combination drugs relied on clinical experience, a process that is not only time-consuming and resource-intensive but also prone to trial and error. In contrast, high-throughput screening has emerged as an affordable and efficient strategy for identifying synergistic drug pairs, leading to the generation of extensive datasets (O'Neil *et al.* 2016, Holbeck *et al.* 2017, Jaaks *et al.* 2022). However, certain limitations persist, such as the inability of cancer cell lines to accurately represent *in vivo* states and the impracticality of exhaustively testing all members of the full combinatorial space with high-throughput screening (Ferreira *et al.* 2013, Goswami *et al.* 2015, Morris *et al.* 2016).

Recently, numerous computational methods for predicting drug synergy have been proposed. Pioneering methods, including DeepSynergy (Preuer *et al.* 2018) and Matchmaker (Kuru *et al.* 2022), utilize deep neural networks with both the chemical properties of drugs and the gene expression profiles of cell lines. The deep tensor factorization model

Received: 20 June 2024; **Revised:** 14 November 2024; **Editorial Decision:** 17 December 2024; **Accepted:** 13 January 2025

(Sun *et al.* 2020) combines tensor decomposition with neural networks to forecast the synergistic effects of drug combinations. Additionally, TransSynergy (Liu and Xie 2021) adopts a transformer network model using drug–target and protein–protein interaction data. DeepDDS (Wang *et al.* 2022) employs graph convolutional networks (GCNs) (Wu *et al.* 2021) and multi-layer perceptrons (MLP) for synergy prediction. The current state-of-the-art method HypergraphSynergy (Liu *et al.* 2022) made strides in this direction by incorporating hypergraph neural networks (HGNNs) (Feng *et al.* 2019, Bai *et al.* 2021) to model these dynamics in a more interconnected and multifaceted manner. Hypergraphs generalize graphs by allowing hyperedges to connect more than two nodes (Chen and Rajapakse 2020, Chen *et al.* 2021, Pickard *et al.* 2023), and encode multidimensional (or high-order) correlations and connections. However, these existing methods often fall short in effectively predicting drug synergy due to their neglect of higher-order drug interactions or lack of important biomedical knowledge. DeepDDS focuses solely on pairwise drug interactions using GCNs, and HypergraphSynergy, although it considers higher-order drug-drug-cell line combinations, fails to incorporate additional biomedical interactions, which hinders its generalization, particularly involving unseen drug molecules.

In this article, we propose a novel deep hypergraph learning method—Heterogeneous Entity Representation for MEdicinal Synergy (HERMES) prediction—to enhance drug synergy prediction. HERMES distinguishes itself with its innovative strategy for integrating a variety of data sources, including drug chemical properties, cell line gene expressions, and interactions between drugs and indications. This integration is achieved through a heterogeneous hypergraph structure, which aids in assimilating extensive prior knowledge and enhances the model's ability to generalize in novel contexts. Incorporating information about drug indications is particularly important, as some drugs may function through similar molecular mechanisms in different diseases (e.g. Bevacizumab is effective against various cancers, including colorectal cancer and non-small cell lung cancer). In contrast, drugs like glucocorticoids demonstrate different mechanisms across diseases due to distinct biological pathways, such as their anti-inflammatory action in rheumatoid arthritis versus their apoptosis-inducing role in leukemia. This approach allows for a more comprehensive capture of the varied biological contexts in which drug combinations achieve synergetic effects. Another notable breakthrough of HERMES is addressing the widespread issue of over-smoothing in message-passing networks by implementing a gated residual mechanism (Li *et al.* 2018), which not only retains more information but also notably enhances the expressiveness of the network.

Our empirical results highlight HERMES's effectiveness, especially in novel scenarios, surpassing HypergraphSynergy and establishing it as a leading solution. The key contributions of this article are (i) integration of varied knowledge sources (including drugs, cancer cell lines, and disease indications) into a scalable and heterogeneous model architecture; (ii) enhancement of message passing over hypergraphs with gated residual mechanisms for augmented network expressiveness; (iii) empirical evidence showcasing superior performance compared to previous methods, highlighting robust generalization in novel contexts, with an average improvement of 5% in evaluation metrics over the state-of-the-art methods. This article is structured into four sections. The main architecture of HERMES is detailed in Section 2. We assess the performance of HERMES (along with other representative drug synergy prediction methods) including a comprehensive ablation study in Section 3. Finally, we conclude by discussing future research directions in Section 4.

2 Materials and methods

2.1 Overview of HERMES

Our framework conceptualizes drug synergy prediction under a hypergraph framework, treating it as a hyperedge prediction problem, where drugs, cell lines, and diseases are represented by nodes while synergistic drug–drug–cell line triplets and drug-disease pairwise relations are represented by hyperedges. Hyperedge prediction is a generalization of edge prediction on graphs (Zhang *et al.* 2018, Kumar *et al.* 2020, Yadati *et al.* 2020, Sharma *et al.* 2021, Chen and Liu 2023, Chen *et al.* 2023). This innovative approach unfolds through three interconnected phases: initialization, refinement, and consolidation (Fig. 1). Each phase plays a pivotal role in processing and integrating varied data types, thereby synthesizing comprehensive information crucial for accurate drug synergy predictions. Drawing inspiration from HypergraphSynergy, our model introduces novel methodologies and integrates cutting-edge techniques to significantly enhance prediction accuracy and reliability at each phase of analysis.

2.2 Feature initialization

The initialization phase is crucial in our synergy prediction methodology, as it involves the acquisition of initial features for drugs, cell lines, and diseases. This phase performs modality-specific representation learning to transform the raw features of different modalities into a common feature space. This alignment process is critical for the subsequent hypergraph learning as it ensures consistency in the representations, allowing for effective integration of features across modalities.

2.2.1 Drug features

We generate drug features by utilizing molecular graphs, which are built from the Simplified Molecular Input Line Entry System (SMILES) representations of drugs (Kim *et al.* 2019). By employing the ConvMolFeaturizer method from the DeepChem library (Duvenaud *et al.* 2015), we construct a molecular graph for each drug, with atoms represented as nodes and bonds as edges. Subsequently, we employ advanced graph transformer networks (GTNs) (Yun *et al.* 2019) on molecular graphs to enhance the representations of atoms and drugs. Denote the feature vector of atom i as $\mathbf{a}_i$. The updated feature (denoted as $\mathbf{a}'_i$) through a GTN is then computed as

$$\mathbf{a}'_i = \sigma\Big(\mathbf{W}_1\mathbf{x}_i + \sum_{j\in\mathcal{N}(i)} \alpha_{ij}\mathbf{W}_2\mathbf{a}_j\Big), \tag{1}$$

where $\mathcal{N}(i)$ denotes the neighboring atoms of atom i, $\mathbf{W}_1$ and $\mathbf{W}_2$ are learnable weight matrices, and σ is a nonlinear activation function. Here, α_{ij} is the multi-head attention coefficient defined as

$$\alpha_{ij} = \mathrm{Softmax}\left(\frac{(\mathbf{W}_3\mathbf{a}_i)^\top(\mathbf{W}_4\mathbf{a}_j)}{\sqrt{d}}\right),$$

where d is the latent size of each head, and $\mathbf{W}_3$ and $\mathbf{W}_4$ are learnable weight matrices. GTNs can flexibly capture the

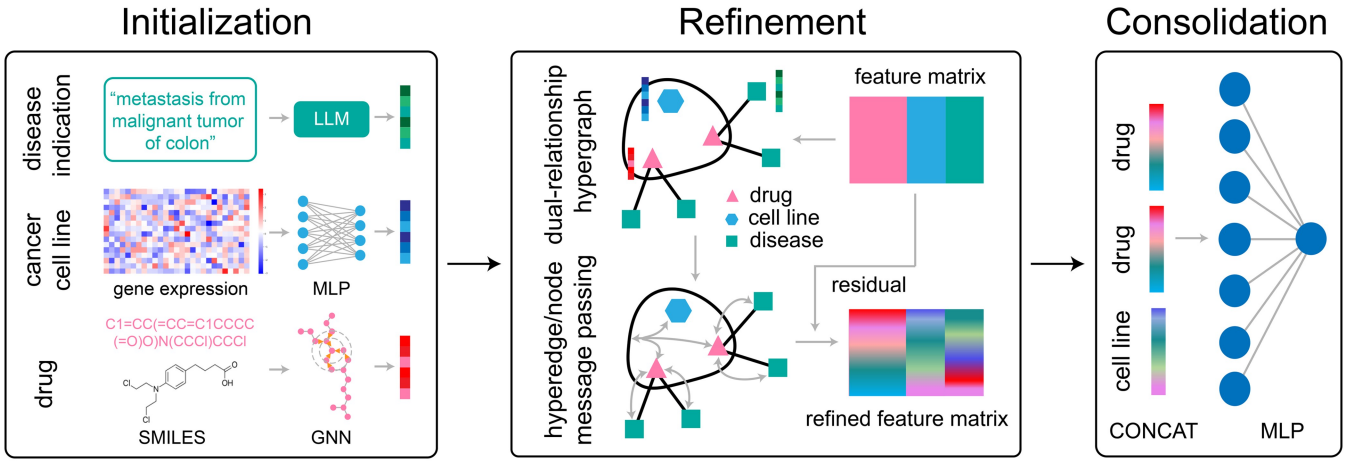

Figure 1. Overview of HERMES framework. HERMES has three key phases: (i) initialization: acquiring and transforming initial features of drugs, cell lines, and disease indications to a uniform dimensionality; (ii) refinement: enhancing feature representations through the construction of a dual-relationship hypergraph and the application of hypergraph neural networks with gated residual connections; (iii) consolidation: integrating refined features using a binary classification model to predict drug synergies with high accuracy. Each phase is integral to the framework's ability to process diverse data types and generate precise drug synergy predictions.

intricate and long-range dependencies among drugs' atoms by adaptively learning attention weights among them, offering more comprehensive drug representations over traditional GCNs which are often limited by local receptive fields. Finally, we compute the feature vector of a drug by aggregating all its updated atom features using maximum pooling. The choice of maximum pooling is motivated by the fact that drug molecular graphs may vary in the number of atom nodes. The resulting feature matrix for all drugs is denoted as $\mathbf{X}_{\text{drug}}$.

2.2.2 Cell line features

We utilize gene expression profiles to generate cell line features. Following a log2 transformation and z-score normalization of gene expression data, we employ an MLP to derive the features for cell lines, ensuring uniform dimensionality with drug features. Denote the feature vector for cell line i by $\mathbf{c}_i$, the updated feature (denoted as $\mathbf{c}'_i$) through an MLP is computed as

$$\mathbf{c}'_i = \sigma(\mathbf{W}_5\mathbf{c}_i + \mathbf{b}_5), \tag{2}$$

where $\mathbf{W}_5$ and $\mathbf{b}_5$ are learnable weight matrix and a bias vector, respectively. The resulting feature matrix for all cell lines is denoted as $\mathbf{X}_{\text{cell}}$.

2.2.3 Disease features

We create disease features by leveraging a novel embedding method called CODER, a knowledge-graph-guided large language model for cross-lingual medical term representation using contrastive learning (Yuan *et al.* 2022). CODER excels in providing close vector representations for various terms representing similar medical concepts in multiple languages, making it particularly beneficial for extracting enriched and contextually relevant disease features (Yuan *et al.* 2022). Similar to cell line features processing, we employ an MLP on the obtained disease features to align with the consistent dimensionality of drug and cell line features. The resulting feature matrix for all diseases is denoted as $\mathbf{X}_{\text{dis}}$. Notably, integrating CODER with disease information enhances our model's ability to comprehend the intricate relationships between drugs, cell lines, and diseases within the hypergraph framework.

2.3 Feature refinement

In the refinement phase, we focus on transforming the obtained drug, cell line, and disease features into contextually enriched representations. This process is achieved through the construction of a dual-relation hypergraph and the implementation of an advanced feature enhancement technique, namely HGNNs with gated residual connections (Li *et al.* 2018). These elements collectively elevate our model's ability to identify complex interrelationships, setting the stage for more precise synergy predictions.

2.3.1 Hypergraph construction

We construct a novel dual-relationship hypergraph involving relationships between drugs, cell lines, and diseases. The hypergraph contains two types of hyperedges: (i) drug–drug–cell line triplets (third-order interactions), which capture the potential synergistic effects between specific drugs and cell lines; (ii) drug–disease relationships (pairwise interactions or traditional edges), which represent associations between drugs and their corresponding disease indications. In this context, diseases refer specifically to drug indications rather than the broader cancer types represented by cell lines. Therefore, we do not explicitly construct hyperedges between cell lines and diseases. This dual-relationship hypergraph, which has never been considered in previous methods, is pivotal in enriching the feature representations by integrating the various features of drugs, cell lines, and diseases.

2.3.2 Hypergraph neural networks

We employ advanced HGNNs with gated residual mechanisms to refine drug, cell lines, and disease features. HGNNs excel in capturing the complex relationships and interactions within the hypergraph, making it a pivotal component in our model for drug synergy prediction (Feng *et al.* 2019). Denote the feature matrix as $\mathbf{X} = \text{vercat}(\mathbf{X}_{\text{drug}}, \mathbf{X}_{\text{cell}}, \mathbf{X}_{\text{dis}})$ where vercat is the vertical concatenation operation. The refined features (denoted as $\mathbf{X}'$) through an HGNN are computed as

$$\mathbf{X}' = \sigma(\mathbf{D}^{-1}\mathbf{H}\mathbf{E}^{-1}\mathbf{H}^{\top}\mathbf{X}\mathbf{W}_6), \quad (3)$$

where $\mathbf{D}$ and $\mathbf{E}$ are diagonal matrices of node and hyperedge degrees, respectively. $\mathbf{H}$ is the incidence matrix of the hypergraph, containing binary values to indicate the presence or absence of any node in any hyperedge, and $\mathbf{W}_6$ is the learnable weight matrix. Here, the superscripts -1 and $\top$ denote matrix inversion and transposition, respectively. To enhance the expressive capacity of our network and extract more intricate patterns from the hypergraph, we increase the number of layers in the HGNN. However, this approach leads to a well-known challenge in graph neural networks, i.e. over-smoothing (Chen *et al.* 2020).

The phenomenon of over-smoothing occurs when the network layers become too deep, leading to the homogenization of node features and a loss of valuable information. To mitigate this issue and further empower our network's expressive capabilities, we introduce a novel solution using gated residual connections (Li *et al.* 2018). The implementation of gated residual connections within an HGNN is defined as

$$\mathbf{X}' = \mathbf{X} + \sigma(\mathbf{W}_7\sigma(\mathbf{D}^{-1}\mathbf{H}\mathbf{E}^{-1}\mathbf{H}^{\top}\mathbf{X}\mathbf{W}_6) + \mathbf{b}_7)\mathbf{X}, \quad (4)$$

where $\mathbf{W}_7$ and $\mathbf{b}_7$ are learnable weight matrix and bias vector of the gate, respectively. The introduction of the gated residual connection in HGNNs allows our neural network to dynamically regulate the integration of original and convoluted features, which significantly reduces the risk of over-smoothing by providing a controlled blending of features. Additionally, we implement the equilibrium bias initialization (EBI) strategy (Wang *et al.* 2010), where we initialize $\mathbf{b}_7$ with a relatively negative value. The EBI strategy ensures that the gate functions close to an identity operation at the start of training, allowing for gradual and more effective feature integration. Our refinement approach utilizing HGNNs coupled with gated residual connections significantly enhances the model's ability to discern complex patterns and interactions between drugs, cell lines, and diseases, thereby creating more accurate representations for them compared to previous methods such as HypergraphSynergy.

2.4 Feature consolidation

In the consolidation phase, our primary objective is to produce predictive insights by integrating the refined features obtained from the earlier phase. We employ a binary classification model that uses the features refined over the dual-relationship hypergraph to identify drug synergies. It classifies the interactions between drug combinations and cell lines into two categories: synergistic and non-synergistic, which provides a clear, binary output for each potential drug synergy scenario.

2.4.1 Predictive model

The essence of the consolidation phase is using a predictive model that leverages the features obtained in the refinement phase to evaluate the potential synergistic effects of drug combinations on specific cell lines and their implications for disease treatment. Suppose the final refined features of drug i, drug j, and cell line k are denoted by $\tilde{\mathbf{d}}_i$, $\tilde{\mathbf{d}}_j$, and $\tilde{\mathbf{c}}_k$, respectively. The synergy score (denoted as S) can be computed using an MLP, i.e.

$$S = \sigma(\mathbf{W}_8\text{vercat}(\tilde{\mathbf{d}}_i, \tilde{\mathbf{d}}_j, \tilde{\mathbf{c}}_k) + \mathbf{b}_8), \quad (5)$$

where $\mathbf{W}_8$ and $\mathbf{b}_8$ are a learnable weight matrix and a bias vector, respectively.

To address the unordered nature of drug pairs, a critical aspect in synergy prediction, we implement a pairwise symmetric permutation augmentation strategy (Zhou *et al.* 2024). This approach involves presenting both permutations of each drug pair in our dataset. For example, if a data sample includes the combination (drug1, drug2, cell line), we also introduce (drug2, drug1, cell line) as a distinct sample. This augmentation is essential to ensure that our model is agnostic to the order of drugs, enhancing its capability to uniformly recognize synergy.

2.4.2 Model training

Given the binary nature of our prediction task, we utilize the cross-entropy loss function (Mao *et al.* 2023), a standard and effective choice for binary classification models defined as

$$\text{Loss} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)], \quad (6)$$

where y_i and $\hat{y}_i$ represent the actual and predicted label (synergistic or non-synergistic) of the ith sample, respectively, and N is the total number of samples. The model is trained on a dataset consisting of drug pairs, cell lines, and their corresponding synergy labels. The training process involves adjusting the weights of the neural network to minimize the cross-entropy loss, enhancing the model's ability to accurately classify drug synergies.

2.5 HERMES and HypergraphSynergy

Although HERMES is inspired by HypergraphSynergy, it significantly diverges from HypergraphSynergy in the following aspects:

- HERMES integrates an additional knowledge source of diseases and constructs a dual-relationship hypergraph.
- HERMES leverages advanced GTNs to obtain drug features, while HypergraphSynergy only utilizes traditional GCNs.
- HERMES improves HGNNs during the refinement phase by incorporating gated residual mechanisms to address the issue of over-smoothing.
- HERMES exploits useful learning techniques such as EBI and pairwise symmetric permutation augmentation to enhance the model's learning ability.

To mitigate the risk of overfitting, HERMES also includes various regularization techniques, including dropout layers, weight decay, and early stopping mechanism. Below, we show that HERMES significantly outperforms HypergraphSynergy in two drug synergy datasets, especially under the context of predicting new drug combinations.

3 Results

3.1 Datasets

We collected four categories of data, including drug synergy data, molecular information for drugs, genomic characteristics of cancer cell lines, and indications for drug therapy in

diseases, from multiple publicly available databases. The details of these data are listed as follows:

- **Drug synergy datasets:** We gathered data on the synergy of anti-cancer drugs from two prominent large-scale tumor screening datasets—the O'Neil dataset (O'Neil *et al.* 2016) and the NCI-ALMANAC dataset (Holbeck *et al.* 2017). The O'Neil dataset comprises 23 062 samples involving 38 unique drugs and 39 distinct human cancer cell lines. Each sample measures Loewe synergy scores for two drugs in combination with a specific cell line. The NCI-ALMANAC dataset contains 304, 549 samples, including ComboScores for 104 FDA-approved drugs in pairings across the NCI-60 cell line panel.
- **Drug molecular structures:** Information on the SMILES of drugs is obtained from the PubChem database (Kim *et al.* 2019).
- **Gene expression in cancer cell lines:** Data on gene expression in cancer cell lines are sourced from the Cell Lines Project within the COSMIC database (Forbes *et al.* 2015). In this context, we specifically considered the expression values of 651 genes related to the COSMIC cancer gene census. These expression values are subjected to logarithmic (log2) transformation and z-score normalization. We also utilized the CCLE (Barretina *et al.* 2012) and GDSC (Yang *et al.* 2013) databases as alternative data sources (see details in Supplementary Data).
- **Drug indications:** The drug-indication annotations are sourced from PrimeKG (Chandak *et al.* 2023), a multimodal knowledge graph designed for precision medicine. PrimeKG integrates information from various high-quality biomedical resources, such as DisGeNET Piñero *et al.* (2020) and DrugBank Wishart *et al.* (2018), as well as data from different biological scales, including disease pathways and phenotypes. It provides detailed relationships between drugs and diseases, covering indications, contraindications, and off-label uses. This comprehensive knowledge base is particularly valuable for identifying potential new applications for drugs and optimizing treatment approaches based on the underlying molecular mechanisms of diseases.

We performed a comprehensive data preprocessing on two primary drug synergy datasets—NCI-ALMANAC and O'Neil. To ensure the quality and relevance of the data, we excluded cell lines lacking gene expression information and drugs without SMILES details. Following this preprocessing phase, the NCI-ALMANAC dataset comprises 74 139 measurement samples of ComboScores for 87 drugs across 55 cancer cell lines, while the O'Neil dataset encompasses 18 950 samples of Loewe synergy scores for 38 drugs and 39 cancer cell lines (Table 1). Subsequently, we removed drugs from the 'drug-disease dataset' that are not present in the aforementioned drug synergy datasets. This led to the extraction of indications for 82 diseases corresponding to 37 drugs within the NCI-ALMANAC dataset and 42 diseases corresponding to 12 drugs within the O'Neil dataset.

Table 1. Statistics of the two datasets.

Dataset	#Drugs	#CLine	#Samples
NCI-ALMANAC	87	55	74 139
O'Neil	38	39	18 950

3.2 Baselines

We conducted a comparative analysis of our approach with representative drug synergy prediction techniques. Below is a brief overview of each of the baseline methods:

- **DeepSynergy** (Preuer *et al.* 2018)**:** It utilizes a three-layer feedforward neural network to predict synergy scores, incorporating gene expression as cell line features and three types of chemical descriptors as drug features.
- **DTF** (Sun *et al.* 2020)**:** It extracts latent features from the drug synergy matrix through tensor factorization and employs them to train a deep neural network model for predicting drug synergy.
- **HypergraphSynergy** (Liu *et al.* 2022) **(the current state-of-the-art method):** It formulates synergistic drug combinations across cancer cell lines as a hypergraph. In this hypergraph, drugs and cell lines are represented by nodes, while synergistic drug–drug–cell line triplets are represented by hyperedges. It leverages the biochemical features of drugs and cell lines as node attributes. Additionally, a HGNN is employed to learn drug and cell line embeddings from the hypergraph and predict drug synergy.
- **NHP** (Yadati *et al.* 2020)**:** It is a GCN-based model specifically designed for hypergraphs to capture complex, higher-order relationships among multiple nodes. NHP employs hyperlink-aware GCN layers to transform hyperedges into clique expansions, which enables the modeling of multi-way interactions among drugs and cell lines in the drug synergy prediction task.

3.3 Experiment setup

In this study, we trained and validated the models independently using the NCI-ALMANAC and O'Neil datasets. For each dataset, it is initially divided into two distinct sets: a training set, accounting for 90% of the total data, and a test set comprising the remaining 10%. The training set undergoes a rigorous five-fold cross-validation process. This process is structured in three unique partitioning strategies to ensure comprehensive evaluation:

- **Random:** Samples are randomly divided, providing a baseline assessment of model performance.
- **CLine:** Samples are stratified by target cell line, ensuring each fold's validation set contains unique cell lines not present in its training set.
- **DrugComb:** Samples are stratified based on drug combinations. Each validation set included drug combinations not seen in the training set, although individual drugs might overlap.

The test set is used for final evaluation, whereby an unbiased assessment of the model's predictive power is ensured.

Additionally, for the classification task, synergy scores are converted to binary outcomes. Following established protocols (Preuer *et al.* 2018, Sun *et al.* 2020), a threshold of 30 is used. Scores above this threshold indicate a positive synergy, while scores below are deemed negative. Ultimately, the model's performance is evaluated using key metrics: area under the receiver operating characteristic curve (AUROC), area under the precision-recall curve (AUPRC), and F1-score. These three metrics together provide a comprehensive view of

the model's effectiveness in terms of precision, recall, and overall balance between precision and recall.

3.4 Hyperparameters

We conducted a grid search for optimized hyperparameters. Hyperparameters that yielded the highest AUROC in cross-validation were chosen for subsequent test experiments. Table 2 presents the range of values considered for each hyperparameter. The optimal values (in bold) are selected based on their performance during the training and validation phases.

3.5 Performance comparison and analysis

In the NCI-ALMANAC dataset, HERMES exhibits superior performance across all evaluation strategies compared to the baseline methods (Fig. 2A). Under the random partitioning strategy, HERMES achieves an AUROC of 85.91% (std: 0.0046), which is significantly higher than HypergraphSynergy, DTF, NHP, and DeepSynergy, which score 85.30%, 82.38%, 77.69%, and 83.50%, respectively (HERMES vs. HypergraphSynergy: P-value < 0.05, two-sample t-test). In the more challenging 'DrugComb' mode, HERMES achieves an AUROC of 79.75% (std: 0.0112), significantly outperforming the other methods, with HypergraphSynergy recording 77.98% (P-value < 0.01, two-sample t-test), NHP achieving 75.71%, and DTF and DeepSynergy achieving even lower scores. For the 'CLine' mode, HERMES also performs better than the rest of the methods. Additionally, the AUPRC and F1-score results reflect a similar trend (middle and right panels).

Table 2. Hyperparameter selection (selected hyperparameters are highlighted in bold).

Hyperparameter	Values
Learning rate	{1e-3, 5e-4, **2e-4**, 1e-4, 5e-5, 2e-5}
Weight decay	{1e-1, **1e-2**, 1e-3, 1e-4}
Attention heads	{2, **4**, 8}
Refinement layer	{2, **3**, 4}
Interaction weight	{0, **0.02**, 0.05, 0.1, 0.2, 0.5, 1.0}

The performance of HERMES is further validated using the O'NEIL dataset (Fig. 2B), where similar results are observed across the three modes in terms of AUROC, AUPRC, and F1-score. In the random stratification, HERMES attains an AUROC of 93.67% (std: 0.0032), significantly outperforming HypergraphSynergy, DTF, NHP, and DeepSynergy, which score 92.30%, 91.38%, 87.07%, and 90.60%, respectively (HERMES vs. HypergraphSynergy: P-value < 0.001, two-sample t-test). In the DrugComb mode, HERMES maintains a competitive edge with an AUROC of 88.34% (std: 0.0205), significantly surpassing HypergraphSynergy at 86.22% (P-value < 0.05, two-sample t-test) and outperforming NHP at 82.92%. HERMES also achieves better performance under the CLine mode compared to the other methods, further supporting its robustness across different datasets.

To further validate the performance of our model, we considered two new validation modes using the ALMANAC dataset (Fig. 3) to assess the model's generalization ability to previously unseen drugs. The two modes are defined as follows: (i) **DrugSingle**, where samples are stratified based on individual drugs, with one drug in each drug combination in the validation set being novel to the training set. This approach reduces the dataset size and challenges the model to predict synergies involving new drugs; (ii) **DrugDouble**, where samples are stratified such that each drug combination in the validation set contains two drugs, neither of which appears in the training set, further reducing the dataset size and increasing the difficulty. The results of 'DrugSingle' and 'DrugDouble' are shown in Fig. 3. In the 'DrugSingle' mode, HERMES achieves an AUROC of 72.13% (std: 0.0118),

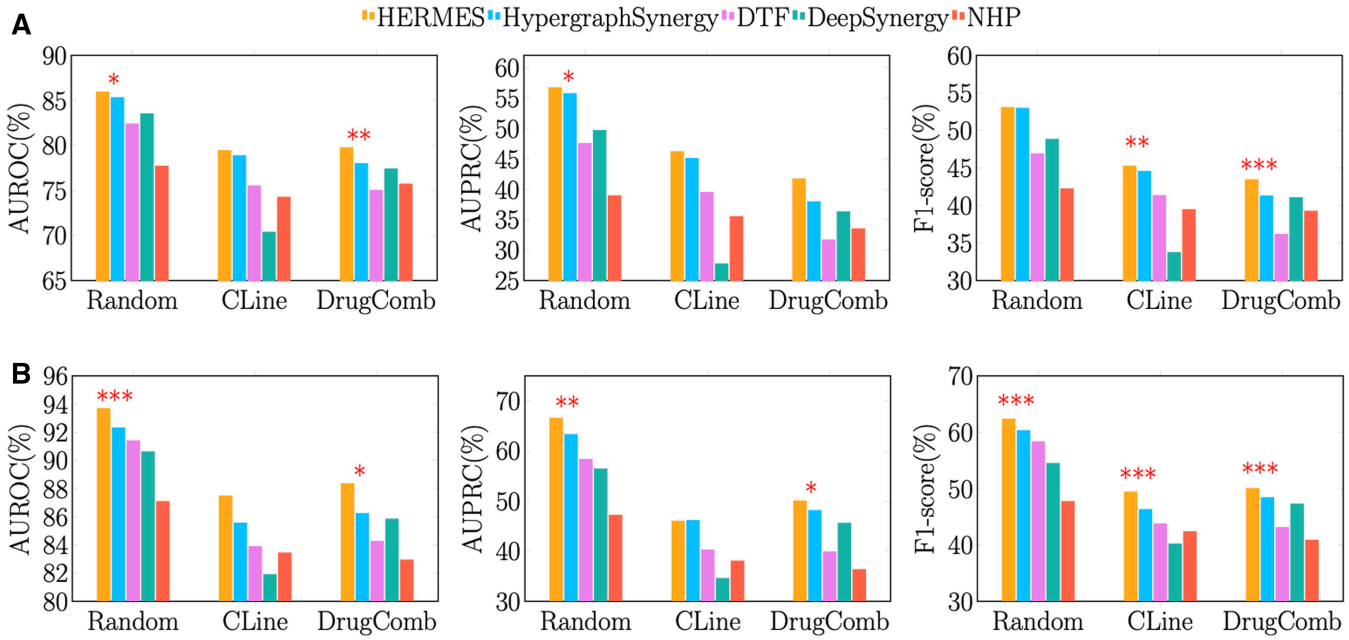

Figure 2. Model performance comparison for (A) ALMANAC dataset and (B) O'Neil Dataset. The left panels show the AUROC (%), the middle panels present the AUPRC (%), and the right panels display the F1-score (%) for different models across three validation modes (Random, CLine, DrugComb). Asterisks indicate statistical significance between HERMES and HypergraphSynergy (***P-value < 0.001; **P-value < 0.01; *P-value < 0.05; two-sample t-test).

which is significantly higher than HypergraphSynergy's 67.84% (P-value < 0.001, two-sample t-test). Similarly, in the 'DrugDouble' mode, HERMES scores an AUROC of 68.39% (std: 0.0468), significantly outperforming HypergraphSynergy (P-value < 0.001, two-sample t-test). The AUPRC metrics also show a consistent advantage for HERMES.

The comprehensive evaluation across different datasets and validation modes underscores the efficacy of HERMES in drug synergy prediction. Although statistical insignificance is observed in certain validation modes, indicating that the extent of improvement can vary based on data distribution, sample size, and task difficulty, HERMES consistently demonstrates competitive and robust performance compared to other methods. These results validate the superiority of HERMES in addressing diverse drug synergy prediction tasks and emphasize its potential for predicting untested drug synergies in clinical applications.

3.6 Ablation study

The ablation study, as detailed in Table 3, scrutinizes the contributions of various components within the HERMES model utilizing the NCI-ALMANAC dataset. The evaluation encompasses three distinct modes: 'Random', 'CLine', and 'DrugComb', with the primary metric being the average test AUROC for each configuration. The benchmark performance of the complete HERMES model registers an AUROC of 85.9% in the 'Random' mode, 79.4% in the 'CLine' mode, and 79.8% in the 'DrugComb' mode. These figures provide a baseline for evaluating the effects of systematically removing specific components or combinations of components.

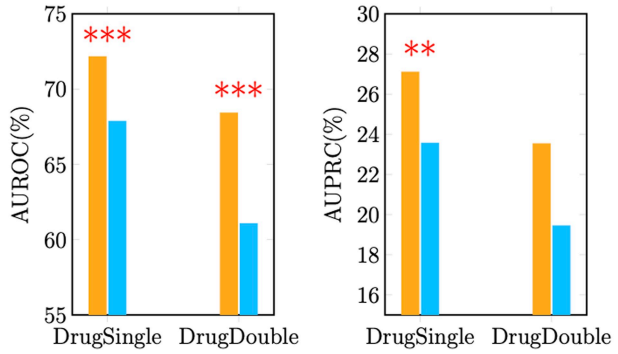

Figure 3. Performance comparison of HERMES and HypergraphSynergy in the DrugSingle and DrugDouble modes using the ALMANAC dataset. Asterisks indicate statistical significance between HERMES and HypergraphSynergy(*** P-value < 0.001; ** P-value < 0.01; two-sample t-test).

3.6.1 Contribution of transformer architecture

Excluding the transformer component results in significant performance drops across all three modes. These findings highlight the indispensable role of the transformer in capturing intricate relationships within the data.

3.6.2 Impact of disease knowledge

The exclusion of disease (drug-indication relation) demonstrates a slight yet notable impact on the model's performance. These findings suggest that incorporating disease knowledge provides a performance advantage, but the margin remains relatively narrow. This may stem from the single-modal nature of the current disease data and the limitations of the classical CODER model. Exploring the integration of multimodal disease knowledge and adapting more advanced large language models could potentially enhance this performance benefit.

3.6.3 Role of gated residuals

Omitting the gated residual connections results in a notable performance degradation, particularly in the 'Random' mode. This emphasizes the significance of gated residuals in maintaining high predictive accuracy, especially in scenarios with higher data variability.

3.6.4 Importance of gating mechanisms

The removal of the gating mechanism induces the most pronounced decline in performance across all modes. This underscores the critical function of gating mechanisms in modulating information flow and enhancing the model's resilience.

3.6.5 Effect of weighted hyperedges

The absence of weighted hyperedges also leads to considerable performance decline across all three modes. This underscores the necessity of weighted hyperedges for accurately modeling complex interactions, particularly in the diverse Random mode.

In summary, the ablation study clearly demonstrates that each component of the HERMES model contributes uniquely to its overall efficacy. Transformer architecture, molecule indication relation, gated residuals, gating mechanisms, and weighted hyperedges collectively enhance the model's predictive performance. These results affirm the sophisticated design of HERMES, showcasing the synergistic effect of its components in achieving superior predictive accuracy across various validation strategies.

Table 3. Ablation study results.

Strategy	Random	CLine	DrugComb
HERMES	85.9% ± 0.005	79.4% ± 0.017	79.8% ± 0.011
w/o transformer	82.8%±0.007	78.5%±0.013	78.3%±0.011
w/o disease	85.7%±0.006	79.2%±0.016	79.5%±0.007
w/o gated residual	84.0%±0.060	79.0%±0.014	78.8%±0.009
w/o gate	82.5%±0.006	78.7%±0.013	78.5%±0.009
w/o transformer + w/o disease	82.1%±0.005	77.8%±0.015	77.3%±0.021
w/o transformer + w/o gate	82.0%±0.004	77.7%±0.014	77.2%±0.021
w/o transformer + w/o disease + w/o gate	80.4%±0.006	76.5%±0.016	77.8%±0.024

Average test AUROC (with standard derivation) of HERMES with and without key components.

4 Conclusion

In this article, we introduced HERMES, a novel deep hypergraph learning framework for drug synergy prediction that integrates heterogeneous biomedical data, including drug molecular structures, gene expression profiles of cell lines, and disease indications. Our results demonstrate that HERMES consistently achieves superior performance across two large-scale benchmark datasets, excelling particularly in challenging scenarios with previously untested drug combinations. The model's gated residual mechanism mitigates over-smoothing in message-passing, enabling it to capture intricate high-order relationships among drugs, cell lines, and diseases with enhanced precision.

While HERMES establishes a flexible and general framework for drug synergy prediction, there are areas for improvement. First, the model currently incorporates only drug, cell line, and disease information, limiting its ability to leverage other relevant biomedical knowledge, such as protein–protein interactions, drug–target interactions, and disease–gene interactions. Integrating such information would further enrich the hypergraph structure and improve the model's generalizability and predictive accuracy. Additionally, HERMES's computational demands, particularly in terms of CPU and GPU usage, could present challenges when scaling to extremely large datasets or incorporating more data sources. To address this, innovative graph/hypergraph sampling techniques and optimized memory management strategies should be explored. Finally, HERMES presently models interactions between only two drugs. Extending the framework to accommodate multi-drug combinations could open promising new avenues for exploring complex treatment regimens and identifying synergies in multi-drug therapies.

In summary, HERMES presents a scalable and adaptable approach to drug synergy prediction, with potential applications extending into other biomedical and clinical research domains, such as drug discovery, personalized medicine, and the development of more effective and efficient therapeutic strategies. It enhances the ease of laboratory implementation while also providing cost-effective solutions for assessing drug synergy. The framework's flexibility opens further opportunities to incorporate additional data sources and adapt to diverse predictive tasks, making it a foundational step towards more accurate and generalizable drug synergy models.

Supplementary data

Supplementary data are available at *Bioinformatics* online.

Conflict of interest: None declared.

Funding

None declared.

Data availability

The data underlying this article are available in https://github.com/Christina327/HERMES.

References

Bai S, Zhang F, Torr PH. Hypergraph convolution and hypergraph attention. *Pattern Recogn* 2021;**110**:107637.

Barretina J, Caponigro G, Stransky N *et al.* The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 2012;**483**:603–7.

Chandak P, Huang K, Zitnik M. Building a knowledge graph to enable precision medicine. *Sci Data* 2023;**10**:67.

Chen C, Liao C, Liu Y-Y. Teasing out missing reactions in genome-scale metabolic networks through hypergraph learning. *Nat Commun* 2023;**14**:2375.

Chen C, Liu Y-Y. A survey on hyperlink prediction. IEEE Trans Neural Netw Learn Syst 2024;**35**:15034–50.

Chen C, Rajapakse I. Tensor entropy for uniform hypergraphs. *IEEE Trans Netw Sci Eng* 2020;**7**:2889–900.

Chen C, Surana A, Bloch AM *et al.* Controllability of hypergraphs. *IEEE Trans Netw Sci Eng* 2021;**8**:1646–57.

Chen D, Lin Y, Li W *et al.* Measuring and relieving the over-smoothing problem for graph neural networks from the topological view. *AAAI* 2020;**34**:3438–45.

Chou T-C. Theoretical basis, experimental design, and computerized simulation of synergism and antagonism in drug combination studies. *Pharmacol Rev* 2006;**58**:621–81.

Csermely P, Korcsmáros T, Kiss HJ *et al.* Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review. *Pharmacol Ther* 2013;**138**:333–408.

Duvenaud DK, Maclaurin D, Iparraguirre J *et al.* Convolutional networks on graphs for learning molecular fingerprints. *Adv Neural Inform Proc Syst* 2015;**28**.

Feng Y, You H, Zhang Z *et al.* Hypergraph neural networks. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33(1), pp. 3558–3565. Honolulu, Hawaii, USA, 2019.

Ferreira D, Adega F, Chaves R. The importance of cancer cell lines as in vitro models in cancer methylome analysis and anticancer drugs testing. *Oncogenomics and Cancer Proteomics-Novel Approaches in Biomarkers Discovery and Therapeutic Targets in Cancer*, vol. **1**. 2013, 139–66.

Forbes SA, Beare D, Gunasekaran P *et al.* Cosmic: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res* 2015;**43**:D805–11.

Foucquier J, Guedj M. Analysis of drug combinations: current methodological landscape. *Pharmacol Res Perspect* 2015;**3**:e00149.

Goswami CP, Cheng L, Alexander P *et al.* A new drug combinatory effect prediction algorithm on the cancer cell based on gene expression and dose–response curve. *CPT Pharmacometrics Syst Pharmacol* 2015;**4**:e9.

Hecht JR, Mitchell E, Chidiac T *et al.* A randomized phase IIIB trial of chemotherapy, bevacizumab, and panitumumab compared with chemotherapy and bevacizumab alone for metastatic colorectal cancer. *J Clin Oncol* 2009;**27**:672–80.

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**:3564–76.

Jaaks P, Coker EA, Vis DJ *et al.* Effective drug combinations in breast, colon and pancreatic cancer cells. *Nature* 2022;**603**:166–73.

Jia J, Zhu F, Ma X *et al.* Mechanisms of drug combinations: interaction and network perspectives. *Nat Rev Drug Discov* 2009;**8**:111–28.

Kim S, Chen J, Cheng T *et al.* PubChem 2019 update: improved access to chemical data. *Nucleic Acids Res* 2019;**47**:D1102–9.

Kumar T, Darwin K, Parthasarathy S *et al.* HPRA: hyperedge prediction using resource allocation. In: *Proceedings of the 12th ACM Conference on Web Science*, pp. 135–143. Online, 2020.

Kuru HI, Tastan O, Cicek AE. Matchmaker: a deep learning framework for drug synergy prediction. *IEEE/ACM Trans Comput Biol Bioinform* 2022;**19**:2334–44.

Li Q, Han Z, Wu X-M. Deeper insights into graph convolutional networks for semi-supervised learning. In: *Proceedings of the AAAI*

Conference on Artificial Intelligence, vol. 32(1). New Orleans, Louisiana, USA, 2018.
Liu Q, Xie L. Transynergy: mechanism-driven interpretable deep neural network for the synergistic prediction and pathway deconvolution of drug combinations. *PLoS Comput Biol* 2021;**17**:e1008653.
Liu X, Song C, Liu S *et al.* Multi-way relation-enhanced hypergraph representation learning for anti-cancer drug synergy prediction. *Bioinformatics* 2022;**38**:4782–9.
Mao A, Mohri M, Zhong Y. Cross-entropy loss functions: theoretical analysis and applications. In: *International Conference on Machine Learning*, pp. 23803–23828. Honolulu, Hawaii, USA: PMLR, 2023.
Morris MK, Clarke DC, Osimiri LC *et al.* Systematic analysis of quantitative logic model ensembles predicts drug combination effects on cell signaling networks. *CPT Pharmacometrics Syst Pharmacol* 2016;**5**:544–53.
O'Neil J, Benita Y, Feldman I *et al.* An unbiased oncology compound screen to identify novel combination strategies. *Mol Cancer Ther* 2016;**15**:1155–62.
Pickard J, Chen C, Salman R *et al.* HAT: hypergraph analysis toolbox. *PLoS Comput Biol* 2023;**19**:e1011190.
Piñero J, Ramírez-Anguita JM, Saüch-Pitarch J *et al.* The disgenet knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res* 2020;**48**:D845–D855.
Preuer K, Lewis RP, Hochreiter S *et al.* Deepsynergy: predicting anti-cancer drug synergy with deep learning. *Bioinformatics* 2018; **34**:1538–46.
Sharma G, Patil P, Murty MN. C3mm: clique-closure based hyperlink prediction. In: *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, pp. 3364–70. Yokohama, Japan, 2021.
Sun Z, Huang S, Jiang P *et al.* Dtf: deep tensor factorization for predicting anticancer drug synergy. *Bioinformatics* 2020;**36**:4483–9.
Tol J, Koopman M, Cats A *et al.* Chemotherapy, bevacizumab, and cetuximab in metastatic colorectal cancer. *N Engl J Med* 2009; **360**:563–72.
Wang H-H, Klessen RS, Dullemond CP *et al.* Equilibrium initialization and stability of three-dimensional gas discs. *Mon Not R Astronom Soc* 2010;**407**:705–20.
Wang J, Liu X, Shen S *et al.* Deepdds: deep graph neural network with attention mechanism to predict synergistic drug combinations. *Brief Bioinform* 2022;**23**:bbab390.
Wang Y-Y, Xu K-J, Song J *et al.* Exploring drug combinations in genetic interaction network. *BMC Bioinform* 2012;**13**, 1–7.
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**:D1074–82.
Wu Z, Pan S, Chen F *et al.* A comprehensive survey on graph neural networks. *IEEE Trans Neural Netw Learn Syst* 2021;**32**:4–24.
Yadati N, Nitin V, Nimishakavi M *et al.* NHP: neural hypergraph link prediction. In: *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pp. 1705–1714. Online, 2020.
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**:D955–61.
Yuan Z, Zhao Z, Sun H *et al.* Coder: knowledge-infused cross-lingual medical term embedding for term normalization. *J Biomed Inform* 2022;**126**:103983.
Yun S, Jeong M, Kim R *et al.* Graph transformer networks. *Adv Neural Inform Proc Syst* 2019;**32**.
Zhang M, Cui Z, Jiang S *et al.* Beyond link prediction: predicting hyperlinks in adjacency space. *AAAI* 2018;**32**:4430–7.
Zhou A, Yang K, Burns K *et al.* Permutation equivariant neural functionals. *Adv Neural Inform Proc Syst* 2024;**36**.

Bioinformatics, 2025, 41, 1–9
https://doi.org/10.1093/bioinformatics/btae750
Original Paper