Computational and Structural Biotechnology Journal 21 (2023) 5099–5110

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Method article

Using graph-based model to identify cell specific synthetic lethal effects

Mengchen Pu [a,*,1,2], Kaiyang Cheng [a,b,1], Xiaorong Li [a,c,1], Yucui Xin [a,1], Lanying Wei [a], Sutong Jin [a,d], Weisheng Zheng [a], Gongxin Peng [a], Qihong Tang [a,e], Jielong Zhou [a], Yingsheng Zhang [a,*,3]

[a] StoneWise, AI, Ltd., Beijing, China
[b] Nanjing University of Chinese Medicine, Shanghai, China
[c] Minzu University of China, Beijing, China
[d] Harbin Institute of Technology, Weihai, China
[e] Guilin University of Electronic Science and Technology, Guangxi, China

ARTICLE INFO

Keywords:
Synthetic lethality
GNN
Deep learning
Cell specific target identification
Multi-omics

ABSTRACT

Synthetic lethal (SL) pairs are pairs of genes whose simultaneous loss-of-function results in cell death, while a damaging mutation of either gene alone does not affect the cell's survival. This makes SL pairs attractive targets for precision cancer therapies, as targeting the unimpaired gene of the SL pair can selectively kill cancer cells that already harbor the impaired gene. Limited by the difficulty of finding true SL pairs, especially on specific cell types, current computational approaches provide only limited insights because of overlooking the crucial aspects of cellular context dependency and mechanistic understanding of SL pairs. As a result, the identification of SL targets still relies on expensive, time-consuming experimental approaches. In this work, we applied cell-line specific multi-omics data to a specially designed deep learning model to predict cell-line specific SL pairs. Through incorporating multiple types of cell-specific omics data with a self-attention module, we represent gene relationships as graphs. Our approach achieves the prediction of SL pairs in a cell-specific manner and demonstrates the potential to facilitate the discovery of cell-specific SL targets for cancer therapeutics, providing a tool to unearth mechanisms underlying the origin of SL in cancer biology. The code and data of our approach can be found at https://github.com/promethiume/SLwise

1. Introduction

Over the past decade, precision medicine has gained widespread acceptance as a concept for developing targeted therapies based on individual biological background. Identification of molecular biomarkers is now a common practice in clinical studies, especially in the field of cancer therapy. Synthetic lethality (SL), where simultaneous inactivation of a gene pair causes cell death, is considered to be of significant importance in cancer treatment. Cancer cells [1] often have a large number of damaging mutations and gene replication errors that are not present in normal cells. If the corresponding SL gene pair of a cancer cell is found as a target, it is possible to precisely kill tumors with the specific mutation without damaging healthy cells. The SL mechanism [2] has the potential to be utilized in precision, anti-cancer drug development. It can be exploited for therapeutic purposes, as targeting SL interactions can selectively kill cancer cells bearing multiple genetic alterations [3]. PARP inhibitors, such as olaparib, niraparib, rucaparib, and talazoparib, have been approved for various cancers based on the SL interaction between PARP and BRCA, providing significant benefits to patients in clinics [4–8]. The success of these inhibitors provides sufficient encouragement for the therapeutic application of the SL interactions. Besides, inhibitors of ATR, WEE1, CHK1, and mTOR, the SL partners of tumor suppressor gene TP53, have shown efficacy in clinical development [9–11]. Inhibitors of PRMT5 and MAT2A, the SL partners of MTAP [12] are also in clinical trials (https://clinicaltrials.gov/). The availability of high-throughput genomics data and therapeutic agents makes cancer an ideal field for the study of precision medicine, which matches a patient's genetic background with the selection of target-oriented

* Corresponding authors.
E-mail addresses: pumengchen@stonewise.cn (M. Pu), zhangyingsheng@stonewise.cn (Y. Zhang).
[1] These authors contributed equally to this work.
[2] ORCID: 0000-0001-6282-2454
[3] ORCID: 0000-0003-2520-3923

https://doi.org/10.1016/j.csbj.2023.10.011
Received 5 July 2023; Received in revised form 5 October 2023; Accepted 6 October 2023
Available online 9 October 2023

drugs. However, identifying SL pairs with large scaled in vitro experimental screening is time-consuming and labor-intensive. Thus, an accurate *in silico* predictor for SL pairs is deemed necessary [13].

Considering the limitations of experiments and the complexity of the SL mechanism, researchers have undertaken efforts to develop computational models for SL prediction in a more efficient manner. Machine learning (ML) and deep learning (DL) methods can effectively integrate multi-dimensional biological data such as paralog data [14], mutation patterns [14,15], expression profiles [16] and protein-protein interaction networks (PPINs) [17] etc., and then perform feature learning through parameter fitting. These methods distill decisive correlations from the comprehensive information data for reliable SL prediction. Currently, there are two major approaches that use deep learning to predict synthetic lethality and both of them have delivered some promising results in predicting synthetic lethality. The first one focuses on identifying synthetic lethal relationships between gene pairs via analyzing the differences between positive and negative gene pairs. The GRSMF model [18] uses known SL interactions to learn the association representation and functional similarity from Gene Ontology (GO) to predict potential SL interactions in a pan-cancer cell manner. This is done by applying the information provided by known SL interactions and gene functional annotations. The SL2MF model [19] uses the SL interactions with additional GO similarity matrix and PPI similarity matrix as supporting information. It leverages the similarity in the network representation when two genes share similar functions. The NSF4SL model [20] includes a strategy for enhancing gene representation by incorporating both global and local information. The model uses two branches of a contrastive learning network to capture the complex characteristics of SL interactions. All the predicted SL pairs were ranked and compared with known SL pairs to evaluate the model's prediction power.

The second approach models the cell system as a graph network, using omics data and other relevant information to establish a network of gene relationships. Potential synthetic lethal pairs can be identified based on their network characteristics. In recent years, graph neural networks (GNNs) have been proven superior in link prediction. The KG4SL model [21] is the first one to integrate a knowledge graph and GNN for the prediction of SL pairs. The model introduces additional information besides genes, such as molecules, diseases and biological processes, to improve the prediction performance of the model. The GCATSL model [17] introduces the Graph Attention Network (GAT) [22] model to predict SL pairs. The GAT model captures the local and global features of each gene node, and uses additional feature data, such as PPI and GO terms, to weight and sum specific feature representations. This results in the reconstruction of a predicted probability matrix. The MGE4SL model [23] uses extra data from sources such as Corum, Reactome, KEGG, and STRING. They employ a Graph Convolutional Networks (GCN) [24] encoder to obtain feature representations of the gene nodes and their neighboring nodes. It combines the features of SL pairs and additional knowledge graph features to obtain a mixed matrix representation of all gene information. The SLMGAE model [25] treated SL pairs as the main graph and the other data sources (e.g., PPI, GO, etc.) as the support view. The implementation of the self-attention mechanism by assigning a randomly initialized and normalized weight matrix to each view, which makes each view adaptively learn the relationship between features. However, due to the uncertainty of the weight matrix, the association between features becomes scattered when more features are introduced, making it challenging to incorporate additional biological characteristics for genes or cells.

The context of species, tissue types, cell types and cellular conditions determines the SL interaction. This complex phenomenon known as the context-specific or context-dependent of SL pairs. Theoretically, by co-inactivating a cancer specific SL pair, a normal tissue can maintain its fitness and resist malignancy, as the cancer cells are selectively killed by the specific lethal effect. This might be dependent on certain intrinsic conditions such as the heterogeneity of different cell types, hypoxia, external disturbances which can result in specific genetic interaction networks [11,26]. Thus, the synthetic lethal effects will vary depending on the tumor cells or cell lines. Besides, targeting tumor-specific or cell-line specific SL pairs could help overcome the resistance to synthetic lethal drugs that target heterogeneous tumors as a whole, and has the potential for treating tumors with various complex conditions in future. Only few computational methods exist that are able to predict cell-specific SL pairs. EXP2SL [15] is a semi-supervised cell-specific SL pairs predictor utilizing L1000 gene expression profiles. Each gene is represented by a 978-dimensional z-score of the shRNA perturbation profile. The extracted features using MLP layers for given gene pairs are concatenated to predict the SL confidence score. In the inference phase, genes with perturbation profile data are the only ones that can be used to make predictions. MVGCN-iSL [27] applies multi-view GCN model to predict cell-specific SL pairs. Several cell-independent networks data, cell-dependent gene expression data and SL pairs information are utilized. The cell-specific relationship between genes is only provided by cell-specific SL labels, which has shown to be the most informative.

Despite having high prediction scores when using data from SL pairs database, the performance of state-of-the-art (SOTA) models may not generalize well on cell-line specific SL pairs. It remains challenging to identify new, robust cell-line specific SL pairs. Several studies rely on perturbation data, which is experimentally costly given the enormous number of gene sets and cell lines involved. It is also difficult to generalize the results of perturbation experiments to all genes for all cell lines. Additionally, variations in tumor cell mutation profiles exist among different clinical patients, it is deemed impractical to train a model for every single cell line in clinical practice. In this study, we have developed a computational tool to predict cell-specific SL pairs using cell-line specific omics data as input. Our approach employs a graph-based representation technique to represent the inter-relationship between genes comprehensively. Additionally, we incorporate a self-attention mechanism, which allows our model to identify the most relevant parts of the input data. We demonstrate that our predictor has the potential to be applied to diverse cancer cell-lines. Furthermore, we have evaluated our model on a cell-line transferable study, and demonstrate the model can generalize to new cell-lines and accurately predict SL pairs in unknown cell-lines.

2. Materials and method

2.1. Data pre-processing

To predict cell-line specific SL pairs, we took individual cell-line multi-omics data as inputs. We also incorporated prior known SL mechanisms in our model to take advantage of their complementary strengths. Specifically, we leveraged paralog genes, which often perform similar functions but exhibit redundancy that can be exploited for targeted therapy when lost in tumors. We also utilized mutually exclusive mutation patterns, which suggest potential SL interactions between incompatible driver mutations. Moreover, integrating high-throughput CRISPR knockout screens with knowledge of low background gene expression levels provides another useful insight for identifying SL pairs. To capture the dynamic relationships between genes in specific cellular contexts, we incorporated data from the L1000 Connectivity Map, which tracks expression changes in response to perturbations to build comprehensive cellular interaction networks. By combining L1000 data, gene effect scores (ES) data, the exclusive mutation (EM) patterns and paralogs, our integrative models aim to reliably identify synthetic lethal gene relationships by accounting for cellular context as well as genomic and functional characteristics.

The 36600 paralog pairs were obtained directly from the supplementary table S8 of Kegel et.al. [16]. For the L1000 data, we utilized level 5 data directly downloaded from the LINCS L1000 project [28]. The original dataset was transformed into a 12328×238351 matrix, with each row representing a gene and each columns represented the

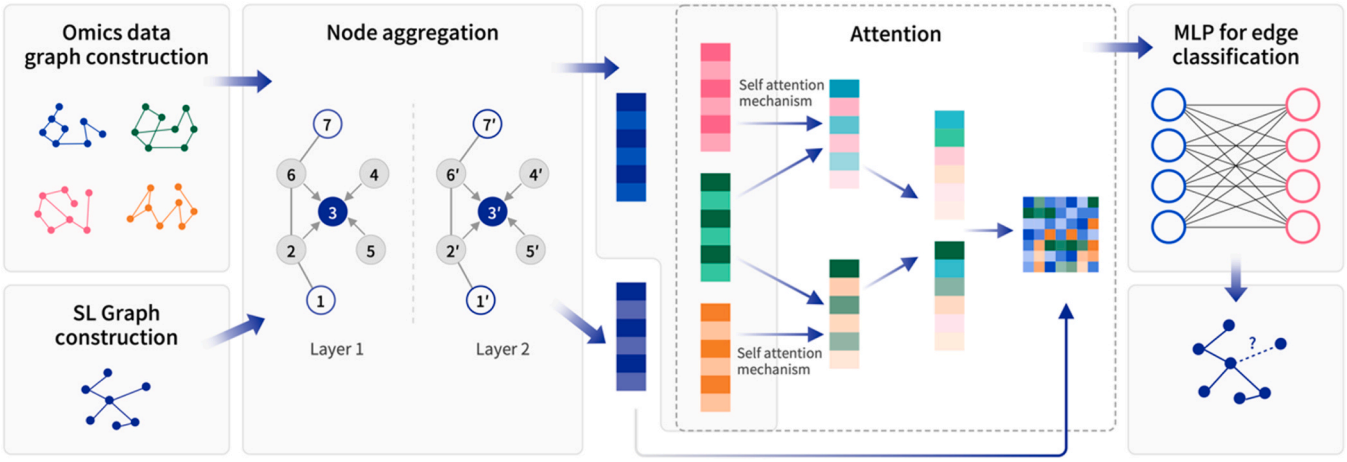

Fig. 1. The framework of our method. The SLwise model incorporates a multiple graph learning approach to integrate diverse cell-specific omics data. Individual graph neural networks are first applied to each omics data (e.g. ES, paralogs). Two layers of GraphSAGE networks are utilized to generate node embeddings by aggregating neighbor features. An attention mechanism then combines the omics embeddings into an integrated representation. These node embeddings are fed into a deep neural network to make SL predictions.

normalized expression fold change values for that gene after a perturbation compared to control, across various cell lines. Using this matrix along with the level 5 annotation data, we integrated CRISPR gene knockout data [29] and focus on the absolute fold change values exceeding 1.5. In our post-processed data, each columns represents the cell-lines, the target genes, the perturb genes, and the fold change values.

Our ES data combined the gene effect score and gene expression data. The gene expression profiles for the cell-lines were retrieved from the public 22Q2 dataset available through the DepMap portal [30]. The gene effect score originated from CRISPR knockout screens conducted by Broad's Project Achilles and Sanger's SCORE projects [31–33], which reflect the normalized impact of knocking out a specific gene on a certain cell-line. Negative scores indicate inhibition and/or death of cell growth following gene knockout, scores lower than − 0.5 indicating depletion on most cell-lines. For each cell-line, we computed z-scores for both gene expression data and gene ES using all genes. We then identified genes with low expression or low ES for each cell-line, defined as genes with expression values < 1 and expression z-scores < -1.28 (corresponding to the lowest 10% of the standard normal distribution), or with ES lower than − 0.5 and gene effect z-scores lower than − 1.28.

To identify significant mutually exclusive mutation patterns for all tested gene pairs in each cell-line, we utilized somatic mutations data from the corresponding TCGA cohorts accessed through the cBioPortal database (https://www.cbioportal.org/). We employed a weighted sampling-based approach, called WeSME [34], to identify significant mutually exclusive gene pairs. After creating a binary gene-sample mutation matrix for the same tumor cohort by recording whether a sample had one or more non-synonymous mutations in a certain gene, we calculated the mutation frequencies of samples, and estimating a null distribution of the mutation profiles of a gene by conducting a simulation 1000 times based on the mutation frequencies of samples. We deemed gene pairs with p-values less than 0.05 to be mutually exclusive. As the final output, we generated a binary matrix to indicate whether each gene pair exhibited significant mutual exclusivity statistically. A value of 1 in this matrix signifies the gene pair is mutually exclusive, while 0 indicates the pair does not.

GEMINI, a variational Bayesian method [35], was applied to generate the ground truth for our model. It was utilized to identify lethal interactions from high-throughput CRISPR-based combinatorial perturbation experiment results [36–39]. The sensitive interaction score was generated from GEMINI for each gene pair and the false discovery rate (FDR), sensitivity scores, and p-values were applied with 0.05 as cutoff for the final ground truth of positive SL pairs. Those pairs with negative SL scores and among the bottom 50% were used to define the negative ground truth. To address the imbalance between the positive and negative SL pairs for each cell-lines, we randomly selected a subset of the majority class with an equal number of examples as the minority class to ensure both positive and negative pairs were equally sampled during training and model evaluation. We employed FDR, p-values, and sensitive scores to identify positive labels from GEMINI calculation and integrated all the results as positive samples. Finally, we obtained 173,

Table 1
List of technical terms, symbols and notations used in the paper.

Term or Symbol	Description / Definition
G^{em}	Graph representation of mutually exclusive genes
G^{es}	Graph representation of gene low expression or low effect scores
G^{par}	Graph representation of gene paralog
G^{fc}	Graph representation of L1000 perturbation fold change expression
G^{SL}	Graph representation of gene pair's SL label for input
$V = \{v_1, v_2, \ldots, v_{\|V\|}\}$	the set of feature vectors of genes (nodes)
$E = \{e_{ij} \mid (i,j) \in V \times V\}$	the set of feature vectors of gene pairs (edges)
h_v^0 or h_v^T	the genetic representation
α_l	The SL label of a gene pair
$\widehat{a}_l$	the predicted SL score of a gene pair
$A = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4, \ldots, \alpha_l, \ldots\}$	the batch of label of gene pairs
$\widehat{A} = \{\widehat{a}_1, \widehat{a}_2, \widehat{a}_3, \widehat{a}_4, \ldots, \widehat{a}_l, \ldots\}$	the batch of predicted score of gene pairs
k	the iteration number of graph convolutional network
L1000	L1000 is a part of the Library of Integrated Network-Based Cellular Signatures (LINCS) Program. The LINCS L1000 project has collected gene expression profiles for thousands of perturbagens at a variety of time points, doses, and cell lines.
Paralog	Paralogs are genes that are related to each other within the same genome, and they arise from a gene duplication event.
Gene effect score	A gene effect score is a measure of the effect size of knocking out a gene, normalized against the distributions of non-essential and pan-essential genes.
Exclusive mutation pattern	Also known as mutual exclusivity. It is when genetic changes in a group of genes don't usually happen in the same sample. This can indicate alternative tumor functions or adverse effects of co-occurrence.
Transformers	A deep learning architecture that is made up of layers based on attention mechanisms

130, and 1345 positive samples and 1446, 28, and 2820 negative samples in HT29, A375, and A549 respectively. We balanced the negative and positive samples and partitioned the dataset into a training set (80%) and a validation set (20%), and then used four groups for training and one group for testing. The metric for evaluating multi-omics data and model architectures is the average performance of all the folds. In addition, we used SL pairs identified in recent years (from 2018 to 2021) in the same cell-line as an external test set to validate the model's ability to make specific cell-line predictions.

2.2. Graph-based neural network

2.2.1. Overview

The prediction task of SL interactions can be represented as a matrix completion task, which aims to predict unobserved interactions. The interactions between and within the different omics levels, such as genomics, transcriptomics, proteomics, and metabolomics, form vast and complex networks, which are challenging to understand. The graph neural network, on the other hand, is well-suited to handle such type of non-Euclidean data. This makes it an ideal tool for analyzing and understanding the complex network of interactions within the omics levels. Here, we presented a graph-based model for SL prediction. The framework of our approach is illustrated in Fig. 1. It uses graphs generated from multiple biological data of gene and cell information, with the SL graph serving as the reference in training process. To generate relevant features, we employ various graph encoders to extract features that take different perspectives on the data, and utilize a self-attention mechanism to integrate all the reconstructed graphs. This is then followed by using a multi-layer perceptron (MLP) for SL pair prediction. In following sections, each module in our approach is detailed. All symbols and notations used in this paper have been summarized in Table 1, along with several technical terms and concepts used throughout the document, which provides a quick reference for understanding.

2.2.2. Graph neural network for cell specific feature extraction

To better understand the relationship between genes in distinguished cells, we applied multi-omics data that are converted to a graph representation G^{em}, G^{es}, G^{par}, and G^{fc} respectively. We first convert the omics data into a graph representation, in which $G = \{v_1, v_2, \ldots, v_{|V|}\}$ is the set of the feature vectors of genes, $E = \{e_{ij} | (i,j) \in V \times V\}$ is the set of feature vectors of a pair of genes, and $e_{ij} \in \{0,1\}^m$ where m is the number of cell-line genes and 1 indicates the whether the value of $gene_i$ and $gene_j$ in those data matrices is 1.

We built a two-layer GraphSAGE [40] module with a fixed number of sampled neighbors to aggregate information for each omics data, since two-hops neighborhood covers most connection information and is usually sufficient for large graph structure learning. The different dimensions of two graph convolution modules are 128 and 16, respectively, each operation is defined as:

$$H_i^{k+1} = \sigma\left(\sum_{j \in N(i)} \frac{1}{|N(i)|} w^{(k)} H_j^{(k)} + b^{(k)}\right)$$

Where $N(i)$ is the neighbor gene list, $w^{(k)}$ is the trainable parameter matrix of the k-th layer, $H_j^{(k)}$ is the representation matrix of gene j, $b^{(k)}$ is the bias term at layer k, and σ is the activation function. The rectified linear unit (ReLU) is applied, which is defined as follows:

$$ReLU(x) = \max(0, x)\#$$

To avoid overfitting, we added a dropout function after each convolutional block with a probability of 0.5.

The complexity and homogeneity of those omics data make it crucial to assign specific edge weights in order to accurately identify important genes. In this case, we assigned a weight of 1 to the EM data, the ES data and paralog data, due to their binary nature. For the L1000 data, which represent gene expression in treatment, we used the actual numbers from the data matrix as edge weights. To prevent over-smoothing in the graph neural network, we only considered edge weights with values greater than three or less than negative three, effectively eliminating less important nodes.

To generate a more informative embedding, we applied an aggregation operation, z_s, to ensemble cell specific representation of gene representations in multi-omics derived from different graph structures. the final embedding can be represented as

$$z_s = Agg\left(z^{EM}, z^{ES}, z^{par}, z^{fc}\right)\#$$

Here, we simply concatenate these latent features together for use in the next attention module.

2.2.3. Feature fusion module

We present a multi-head transformer cross-attention method that directs attention to three features such as EM data from both ES data and paralogs in two stages. The three omics data (EM, ES, paralog) are derived from encoders and then fed into the multi-head attention module [41], also known as the transformer block. The latent feature, combined with SL pairs feature representation, is passed through an MLP layer for SL interactions reconstruction. We use the attention mechanism to learn the weight distribution of different features, which helps to identify the important features for prediction. The multi-head attention is calculated by the following formulas

$$X_{HEAD} = Concat(head_1, head_2, \ldots, head_m) w^0\#$$

$$Head_i = softmax\left(\frac{Q_i * K_i^T}{\sqrt{d_k}}\right) V_i\#$$

$$Q_i = X \times W_i^Q, K_i = X \times K_i^Q, V_i = X \times V_i^Q, \#$$

where Q_i, K_i and V_i are the Q, K and V matrices derived from the linear transformation of those biological features are passed through the attention layer and the feed-forward network layer. For the L1000 data, we simply concatenated the GraphSAGE features of it with the output of the attention module.

The final representation for link prediction (positive SL pairs) is created by combining the relevant information with the graph representations of SL pairs, G^{SL}. Additionally, layer normalization is also applied to accelerate the convergence of the neural network and prevent the 'covariate-shift' and 'high-parameters' issues.

2.2.4. MLP for edge prediction

We implemented a fully connected neural network to predict the potential SL pairs. It takes in features from the fusion layer as input and has multiple hidden layers utilizing ReLU as activation function. The output layer contains a single node with a sigmoid activation function, which outputs a probability indicating the likelihood that a given edge corresponds to an SL pair in a specific cell-line. The Binary Cross Entropy loss function is applied.

2.2.5. Model training

The model was trained to minimize mean square error loss using the Adam optimizer with a learning rate of 0.001 and weight decay of 0.0005. The model was trained for up to 2000 epochs with early stopping after 30 epochs of no improvement in validation loss to prevent overfitting. The training process of our approach is illustrated in Algorithm 1.

Algorithm 1. The training process of our model.

Input: adacency matrix and edge weight matrix of multi-omics data φ, gene pairs label, maximum training epoches T, learning rates $\gamma = 2e^{-4}$, weight decay=0.01, parameter sets θ
Output: gene pairs with SL probability scores P

Initialize parameter θ randomly;
Initialize $bestLoss \leftarrow 10000$
Initialize $tempEp \leftarrow 0$
while $epoch \leq T$ **do**
  Compute graph representations for φ and G^{SL}
  Compute the BCE loss L
  $tempEp = tempEp + 1$;
  if $L < bestLoss$ **then**
    $bestLoss \leftarrow L$;
    Calculate $\partial H/\partial\theta$;
    use optimizer to update θ
    $tempEp \leftarrow 0$;
  end
  if $tempEp = 30$ **then**
    Return P ; /* early stop */
  end
end
return P

2.2.6. Ablation experiment

In our experiments, we evaluated the impact of specific combinations of input features and architectural components on our current model by conducting feature ablation and module ablation. In these experiments, we removed individual input features from multi-omics data in feature ablation and replaced different feature extraction modules or feature aggregation modules in module ablation.

Besides GraphSAGE, other graph based encoders such as GCN, GAT, and Topology adaptive graph convolutional networks (TAG) [42] are also designed for processing graph-structured data through the use of aggregation function, combination function, and readout function. GraphSAGE learns node features by aggregating information from a fixed number of sampled neighbors. GCN use convolutional operations to aggregate information from neighboring nodes, GATs use attention mechanisms to weight the importance of the neighboring nodes while aggregating the node's feature information. while TAG is a variant of GCN that adapts its convolutional filter based on the graph's topology. These three architectures are utilized in the ablation study of feature extraction module.

The final model incorporated an attention mechanism into the omics feature aggregation modules. Due to the time and computational complexity of the transformer module, we conducted an additional ablation experiment in which we substituted the transformer module with a much simpler operation that concatenated the output of GraphSAGE modules. This experiment demonstrated the impact of the transformer module on the model performance.

2.2.7. Performance evaluation

To evaluate the performance of our approach, we use four metrics: recall, precision, the area under the precision-recall curve (AUPR) and the area under the receiver operating characteristic curve (AUROC). We compare our predictor with EXP2SL, which is a cell-line specific predictor that serves as a baseline. All the metrics were calculated using Python (Scikit Learn package). The sparsity of the SL labels may lead to overfitting of the model. To address this issue, we conducted three types of evaluations with different ways of splitting the dataset. In evaluation study 1 (CV1), the dataset was partitioned by gene pairs, such that both genes in a test set might also appear in the training set. In evaluation study 2 (CV2), we divided the dataset by genes, ensuring that only one gene in a test pair was also present in the training set. In evaluation study 3 (CV3), we separated the dataset by genes, excluding both genes in a test set from the training set.

2.3. Analyses of Cellular Context-specific SL Mechanisms

The top-ranked SL pairs from SLWise model are compared with SOTA model: EXP2SL, NSF4SL, and MGE4SL. The predicted SL pairs from our approach were then used to decipher the mechanisms underlying the cellular specificity. Each gene of these SL pairs was used as the candidate perturbation gene. L1000 data include many other gene expression data in the given cell line used for candidate gene perturbation. The significantly perturbed gene sets were identified by setting the log2FC > 1 and adjusted p-value < 0.05 in expression compared to the control from L1000 data. For the candidate gene pair including gene A and gene B, we filtered out the gene_set A that was significantly downregulated by gene A, the gene_set B that was significantly down-regulated by gene B, and gene_set C, the overlapping portion of the gene_set A and gene_set B, which can be down-regulated by either gene A or B. The genes in gene set C that both classified as tumor driver genes and had a CERES score less than -0.5, were retained as the essential genes. Then, to localize potential cellular damage, we performed GSEA using clusterProfiler (version 4.4.4) [69] on the subset of genes in the gene set C with CERES scores less than -0.5. The enrichment items that achieved statistical significance ($p < 0.01$) were taken into consideration.

3. Results And Discussion

3.1. Overall performance

We evaluated our approach using five-fold cross-validation with specific settings on the aforementioned datasets for the three cell-lines and then compared the performance of our model to the SOTA EXP2SL model as baseline. We started by splicing the ground truth data into training and testing sets, and assessing the model's performance within a single cell-line using AUC, AUPR, recall and precision, which presented in Fig. 2. Detailed information can be found in the supplementary table S1. Compared to the baseline model EXP2SL, our model demonstrated a relatively stable performance with only a slight decline from CV1 to CV3 test, whereas the performance of the EXP2SL model dropped significantly.

Fig. 2 summarizes the performance of our approach and EXP2SL on three different cell-lines: A375, A549, and HT29. Both methods achieved similar results under CV1, where they both performed best on A375 cell-line. However, our approach showed a clear advantage over EXP2SL under CV2 and CV3, where the gene sets for training and testing are more independent. Under CV2 datasets, our model improved by 12.3% and 1% on AUC, 51.5% and 33.9% on AUPR compared to EXP2SL in A549 and HT29 respectively. In particular, under CV3, EXP2SL suffers a significant decline in performance, while our approach maintained a high level performance across all three cell-lines. Our model improved by 51%, and 30.7% on Recall, and by 26.3%, and 9.7% on AUPR in the A549, and HT29 respectively.

These results demonstrate that our model is able to effectively capture the cell features from multi omics data and can stratify SL pairs in the dataset for different cell lines. It is noted that our approach also struggled a little bit in identifying SL pairs within the cell-line A549. However, the overall performance of this study suggests that this approach is promising for identifying SL pairs through cell-specific genetic interaction data.

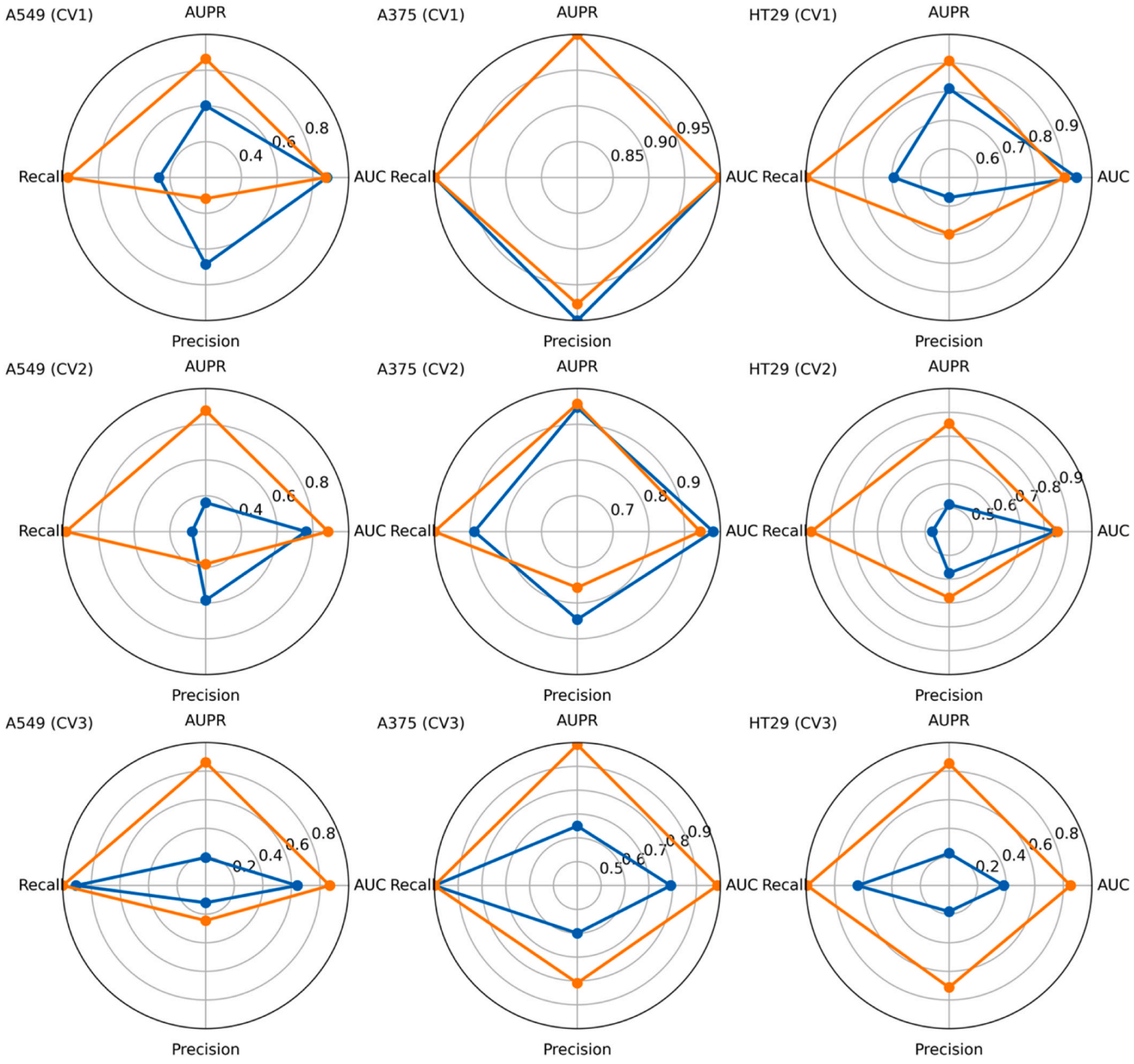

Fig. 2. The performance of evaluation in three different cell lines under different split test set.

3.2. Ablation result

The ablation study helps to verify the rationality of our current features and model architecture, and provide valuable insights into how to improve the model in the future. In this paper, we investigate the impact of the encoder module, the attention module and input omics data on the accuracy of predicting SL interactions. The results are shown in Supplementary Table S2-S4.

We conducted an ablation study to evaluate the performance of different encoder modules in our model for predicting SL interactions. We compared three types of graph convolutional network encoders (GAT, GCN, and TAG) and validated our model in three cell lines, averaging the results and compared with our original selection, GraphSAGE. As shown in Table S2, the performance of the GraphSAGE encoder module was slightly inferior to GCN in terms of AUC and AUPR, but demonstrated a competitive advantage in other metrics, showing that mean aggregation from neighborhood features of each node from multi-omics graph is very effective for SL prediction. Moreover, the

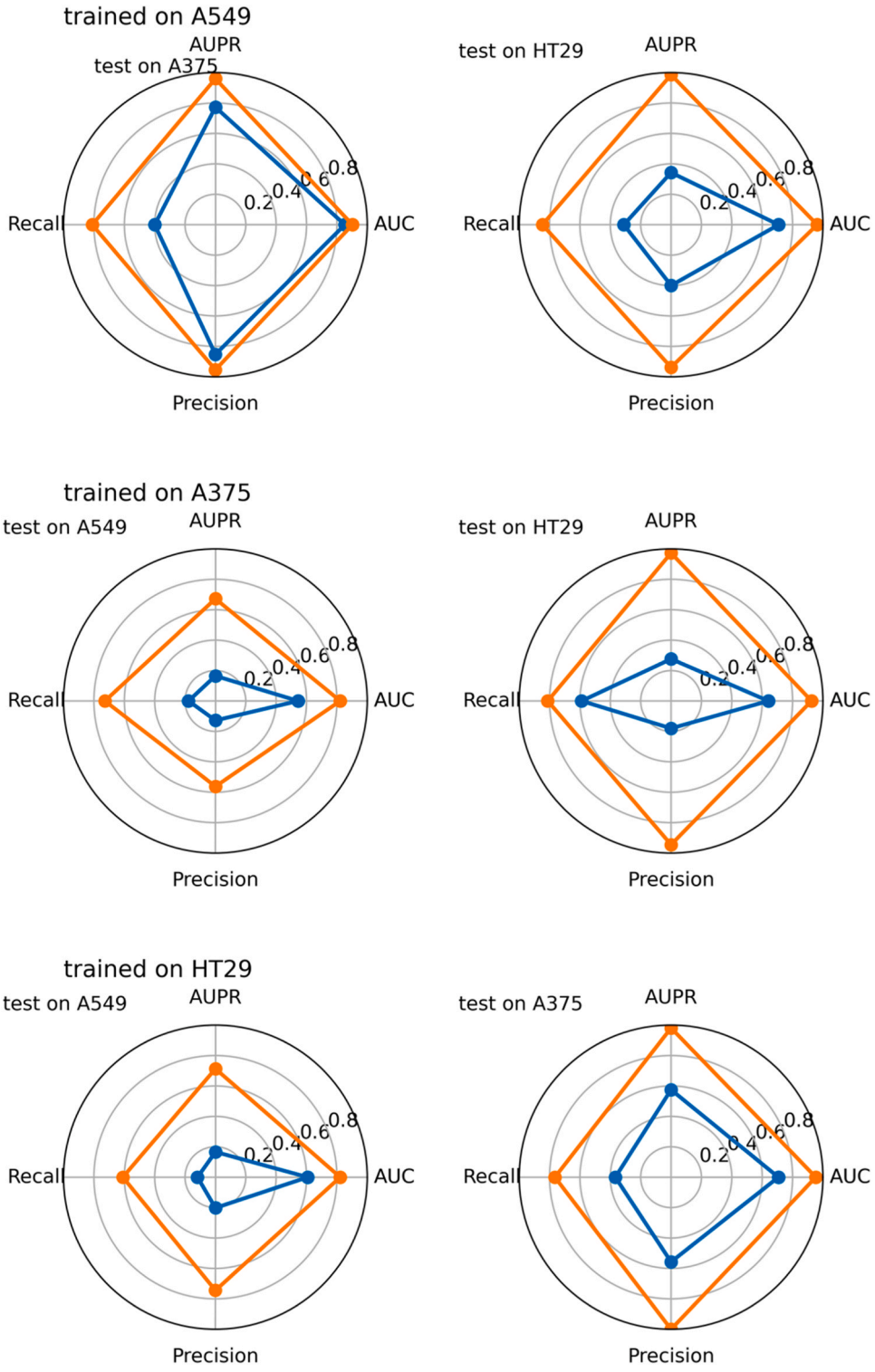

Fig. 3. The performance of transferable evaluation in three different cell lines.

fixed-number neighborhood sampling strategies enable GraphSAGE to aggregate a subset of each gene node, making it more scalable.

In addition, we also conducted an ablation study to evaluate the effect of attention mechanism on the feature aggregation module. The experimental results (Table S3) demonstrate that incorporating attention module can significantly improve the model performance compared with simply concatenating the extracted features.

Furthermore, we performed ablation study to examine the impact of different multi-omics data on effectiveness of our model. Details of the performance on each cell-line are shown in Table S4. We tested different combinations of input data by removing one of those different omics data types, and compared the result with those of the complete dataset. The ablation results demonstrate that using the complete dataset, the performance shows improvements on all metrices, indicating that each multi-omics data is beneficial to predict known positive SL samples in our approach. These experiments show that utilizing all multi-omics information to predict SL interactions can improve the performance of predictions. In the future, we may incorporate additional omics data or use data enhancement techniques to improve our predictions even further.

3.3. Cell-line transferable study

Finally, we evaluated the model's ability to predict SL pairs in unknown cell-lines, where the training and testing data came from different cell-lines, mimicking the real-world scenario. It helps to establish the model's ability to be used in practical applications. The preliminary results are encouraging when taking into account the number of SL label for each cell-lines.

Our model was able to make accurate predictions when using data from different cell-lines for training and testing (Fig. 3). Detailed information can be found in the supplementary table S5. Our model had the best performance when using A549 as training data. The large amount of single-label (SL) data available for A549, 1345 positives and 2820 negatives, may have contributed to its performance. This dataset is almost five times bigger than the amount available for the other two cell lines (A375 and HT29). Besides, our approach also managed to maintain a satisfactory level of performance when using the other cell-lines as training data. In these tests, EXP2SL performed poorly. These results demonstrate that our model is transferable among cell lines, and has the potential to predict SL pairs for unknown cell-lines.

3.4. Evaluation of SLWise and other SL Prediction Models on Cell Context-specific SL pairs

While machine learning models can predict many synthetic lethal (SL) gene pairs, evaluating real-world utility remains challenging due to the extensive experimental validation required. To ensure practical applicability, we benchmark our model's accuracy at prioritizing validated SL pairs within the top predictions. Focusing evaluation on highly-ranked candidates provide critical insight into predictive power on the most relevant pairs.

It is clear that our model (SLWise) outperformed the EXP2SL, NSF4SL, and MGE4SL models in accuracy using true positive labels (Table 2). Among the top 100 predictions, SLWise successfully identified 22 true SL pairs in A375 and two true SL pairs in HT29 (Table 2, Table S6), while EXP2SL failed to identify any true SL pair. Notably, 20 out of 22 true SL pairs in A375 and one out of two in HT29 exhibited cell context-specificity, meaning that the SL gene pairs had lethal effects in one cell type but not in another due to their different cellular context. Expanding the evaluation to the top 200 predictions, EXP2SL exhibited even poorer performance, failing to identify additional SL pairs. In contrast, SLWise identified 30 true SL pairs (27 of which are cell context-specific) in A375, six true SL pairs (five of which are cell context-specific) in HT29, and there is no true SL pair in A549. Neither NSF4SL nor MGE4SL identified any SL pairs within the top 200 results.

Table 2
Top100 and 200 predicted SL pairs performance evaluation.

		Taining SL data	Testing SL data	SL pairs	Cell context-specific SL pair	General SL pair
Top100	SLWise	A549	A375	**22**	**20**	2
		A549	HT29	**2**	**1**	1
		A375	A549	0	0	0
	EXP2SL	A549	A375	0	0	0
		A549	HT29	0	0	0
		A375	A549	**5**	**4**	1
Top200	SLWise	A549	A375	**30**	**27**	3
		A549	HT29	**6**	**5**	1
		A375	A549	0	0	0
	EXP2SL	A549	A375	0	0	0
		A549	HT29	0	0	0
		A375	A549	**5**	**4**	1
	NSF4SL	SynLethDB 2.0		0	0	0
	MGE4SL	SynLethDB		0	0	0

Moreover, when extending the analysis to the top 1000 predicted SL pairs, SLWise consistently demonstrated better performance. It successfully identified 30 true SL pairs (27 of which were cell context-specific) in the A375 cell line, seven true SL pairs (six of which are cell context-specific) in the HT29 cell line, and five true SL pairs (all of which are cell context-specific) in the A549 cell line (Table S7). In contrast, EXP2SL only identified seven true SL pairs (five of which are cell context-specific) in the A549 cell line and failed to identify any SL pairs in other cells. MGE4SL identified only four true SL pairs, and NSF4SL failed to identify any true pair. Furthermore, we conducted a performance comparison by employing the same approach as the EXP2SL method, focusing on using the same cell line for training and testing. In this evaluation, SLWise still demonstrated better performance over EXP2SL (Table S8). These findings demonstrate the evident superiority of our SLWise model's accuracy in predicting cell context-specific SL interactions compared to the baseline methods.

For a more intuitive visualization, we have listed several SL gene pairs labeled as positive among the top 100 prediction results (Table 3). It is worth noting that many of these predicted SL pairs have been validated using low-throughput experiments. For example, the paralog pairs BCL2L1-MCL1 [39,43] and PARP1-PARP2 [44] have been verified for their SL interactions, and the combination of BCL2L1 or BCL2L2

Table 3
SL pairs with the positive label in the top 100 predicted results.

Cell line	Gene A	Gene B	Cell context-specific
A375	BCL2L2	UBC	Yes
	BCL2L2	WEE1	Yes
	MAPK3	UBC	Yes
	PARP1	UBC	Yes
	BCL2L1	BCL2L2	Yes
	BCL2L1	WEE1	Yes
	BCL2L2	MAPK3	Yes
	BCL2L2	PARP1	Yes
	MAPK3	WEE1	
	PARP1	WEE1	Yes
	MAPK3	PARP2	Yes
	PARP1	PARP2	Yes
	BCL2L2	MCL1	Yes
	MCL1	WEE1	
	AKT3	UBC	Yes
	AKT3	BCL2L2	Yes
	AKT3	WEE1	Yes
	AKT3	PARP2	Yes
	BCL2L1	MAPK3	Yes
	BCL2L1	PARP1	Yes
	MAPK3	PARP1	Yes
	BCL2L1	MCL1	Yes
HT29	MAPK3	WEE1	
	MAPK1	MAPK3	Yes

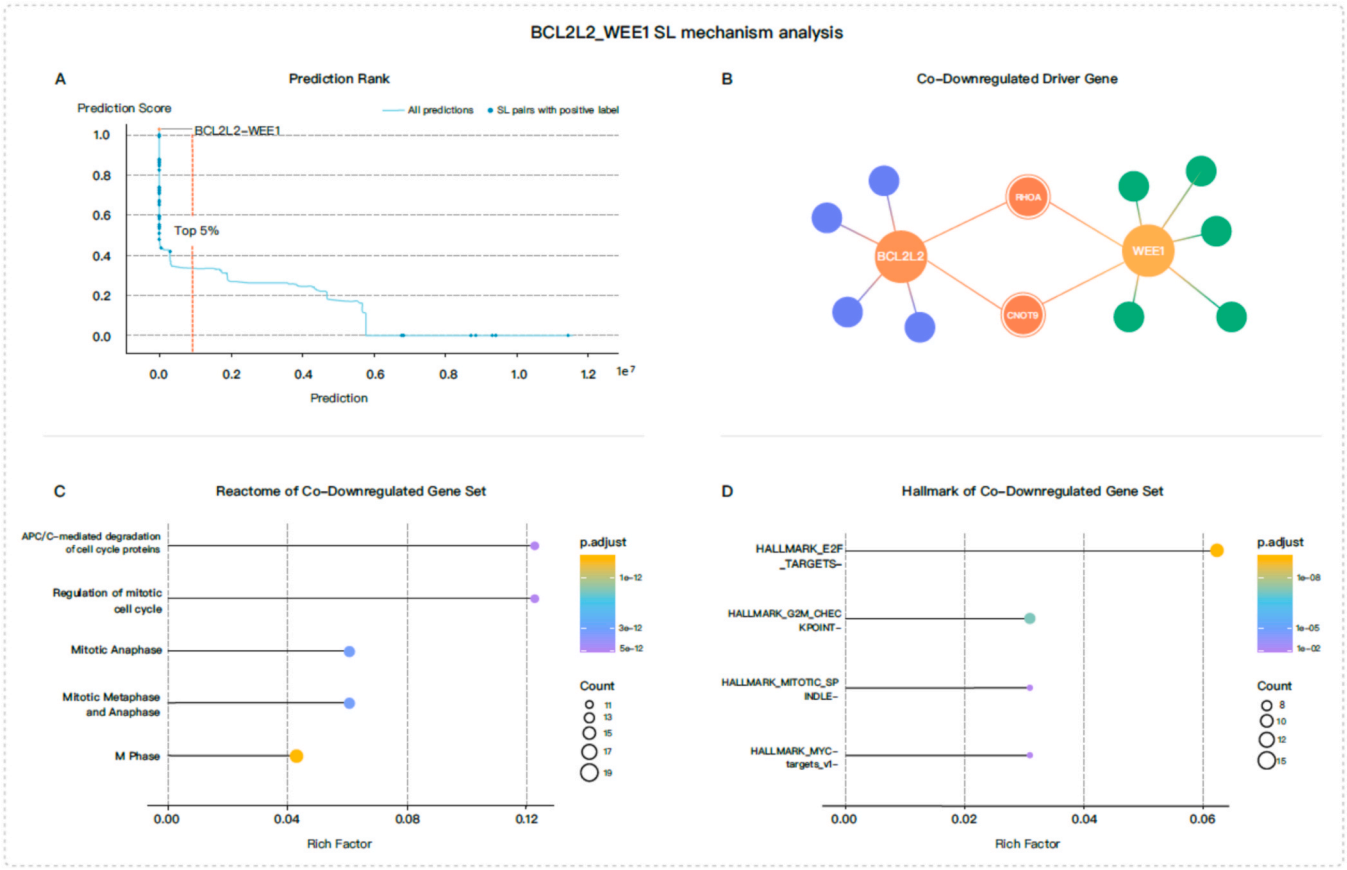

Fig. 4. Representative predicted gene pair (BCL2L2-WEE1) and its SL mechanism analysis in A375 cell line. (A). The prediction score and rank of all candidate SL gene pairs. The blue dots are SL pairs with positive labels. Predictions falling on the left side of the dotted orange line represent the top 5% ranking. (B). Among the downregulated genes of BCL2L2 and WEE1, RHOA and CNOT9 are the common significant driver gene with a CERES score below − 0.5 in A375. (C). The Reactome pathway analysis demonstrates the presence of cellular damage. (D). The Hallmark analysis demonstrates the presence of cellular damage.

knockdown and PARP inhibitor has demonstrated a significant reduction in the viability of certain tumor cells [45]. The combined inhibition of WEE1 and PARP1 has been shown to induce SL interactions, particularly in combination with radiation [46,47]. The co-inhibition of AKT3 and WEE1 has been proven to decrease the development of melanoma [45].

We utilized the logic of synthetic lethal associated gene detection and cell damage evaluation, to identify abnormalities in essential gene combinations involved in cell survival and death from the knockout of predicted SL genes, and aimed to establish associations between candidate SL gene pairs and cellular damage, providing insights into the underlying mechanisms of SL interactions in a cellular context-specific manner.

In the A375 cell line, we predicted an SL interaction between BCL2L2 and WEE1(Fig. 4A). We discovered that the knockdown of BCL2L2 and WEE1 in the A375 cell line resulted in abnormalities in two significant driver genes, CNOT9 and RHOA (Fig. 4B). It is noteworthy that RHOA plays a crucial role in the growth, progression, and metastasis of various cancer types and has been considered a therapeutic target [48]. Moreover, the cell damage is enriched in mitotic anaphase and M phase, as observed in the Reactome analysis (Fig. 4C), and in E2F_targets and G2M checkpoint, observed in the Hallmark analysis (Fig. 4D). These findings are consistent with previous reports indicating that the inhibition of WEE1 in tumor cells increases the dependency on BCL2L2 [49], providing a plausible explanation for the observed cellular damage. In contrast, in the HT29 cell line, the knockout of BCL2L2 and WEE1 did not result in abnormalities in any essential driver gene, and there was no significant enrichment observed in the Hallmark and Reactome pathways. Additionally, although a significant downregulation of the essential driver gene was detected in the A549 cell line, no significant enrichment was observed in the Hallmark and Reactome pathways.

Similarly, in the HT29 cell line, we observed an SL interaction between UBC and UBE2L6 (Fig. 5A). UBC and UBE2L6 encode Ubiquitin and E2 ubiquitin-conjugating enzymes, respectively, both of which are critical post-translational modifiers involved in maintaining genome stability. Upon analyzing the effects of UBC and UBE2L6 knockout in the HT29 cell line, we found that the only essential gene STIL abnormality (Fig. 5B). Notably, inhibition of STIL has been shown to suppress tumor progression, indicating its importance in cancer development [50]. Further analysis using Reactome and Hallmark enrichment revealed that the downregulation of UBC and UBE2L6 led to mitotic damage (Figs. 5C and 5D). This suggests that the disruption of UBC and UBE2L6 may induce cellular damage in the HT29 cell line through the inhibition of STIL, leading to impaired mitotic processes. In contrast, in the A375 cell line, no Hallmark or Reactome enrichments in this disfunction were observed.

4. Discussion

In our approach, we select some proven SL mechanisms relevant to cell context and incorporated them into our models. Specifically, paralogs, which arise from duplicated sequences of a shared ancestor and often perform similar functions, exhibit functional redundancy. Their loss is more common in tumors [51], making them potential precision targets for cancer treatment and an essential dataset for SL discovery [52,53]. Mutually exclusive mutation patterns suggest incompatible driver mutations in tumorigenesis, indicating a potential source of SL interactions [54]. In addition, the combination of

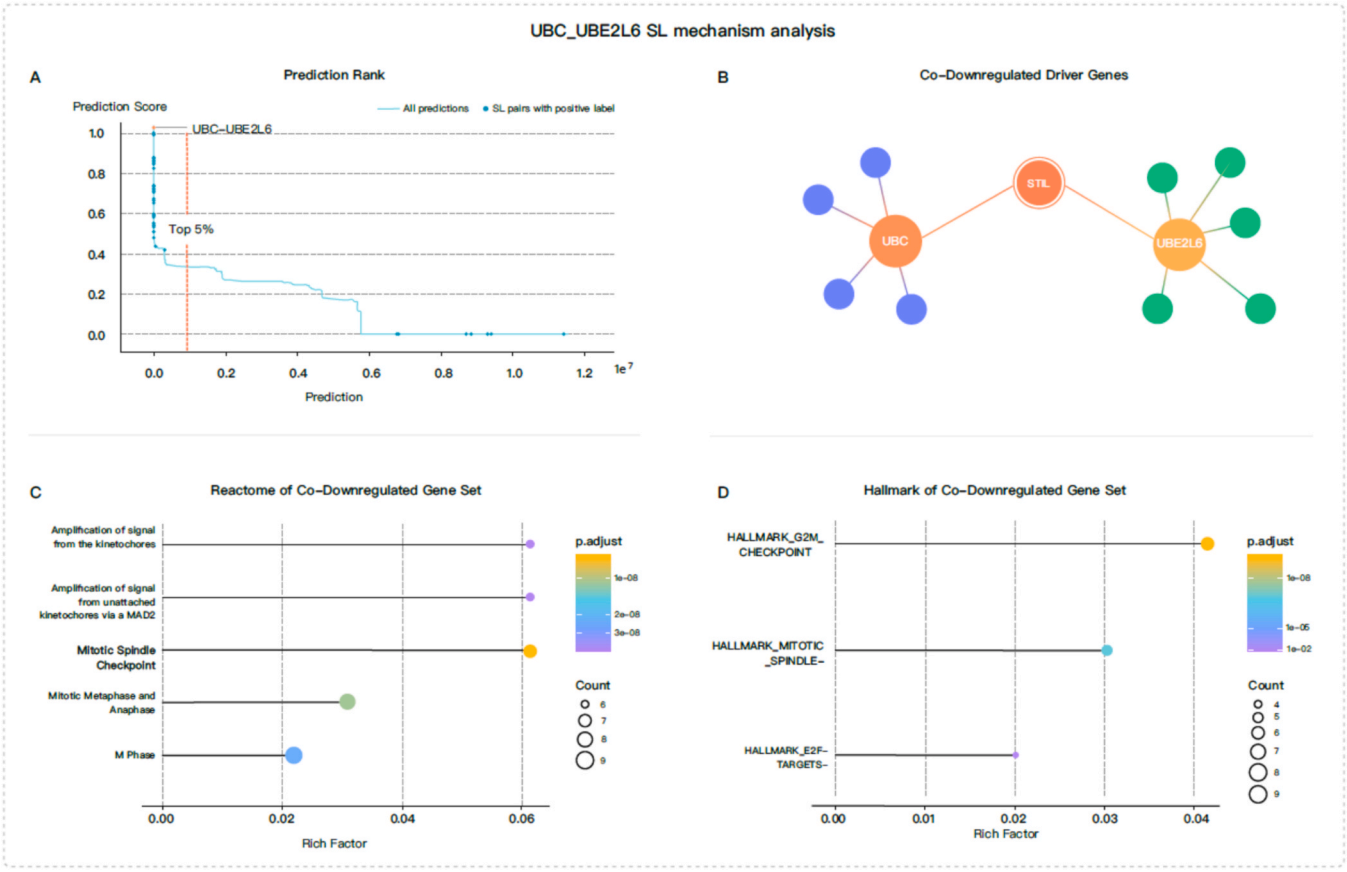

Fig. 5. Representative predicted gene pair (UBC-UBE2L6) and its SL mechanism analysis in HT29 cell line. (A). The prediction score and rank of all candidate SL gene pairs. The blue dots are SL pairs with positive labels. Predictions falling on the left side of the dotted orange line represent the top 5% ranking. (B). Among the downregulated genes of UBC and UBE2L6, STIL is the only common significant driver gene with a CERES score below − 0.5 in A375. (C). The Reactome pathway analysis demonstrates the presence of cellular damage. (D). The Hallmark analysis demonstrates the presence of cellular damage.

high-throughput CRISPR-Cas9 screens with the background gene's low expression (which might be caused by mutation, copy number variation, or epigenetic modification) is also a major logic for SL pair discovery [55–57]. Furthermore, to understand specific SL interactions in the cellular context, it's important to consider the dynamic relationship of genes. The L1000 data is a valuable resource for tracking the cellular responses and helps characterize specific gene relationships based on expression changes caused by gene perturbations, providing a comprehensive understanding of the cellular context for SL interactions.

By analyzing specific examples like BCL2L2-WEE1 in A375 and UBC-UBE2L6 in HT29, we highlight the potential effects and damage caused by these SL pairs in specific cell lines. Notably, although EXP2SL also incorporates the L1000 data, it relies solely on a limited set of genes from L1000 data as its input features, resulting in incomplete learning of gene-gene relationships. This limitation could hinder the capture of the fundamental gene interaction logic behind cell line characteristics and SL pairs, making it suboptimal performance for transferability among cell lines.

Obviously, SLWise model provide valuable insights into predicting and analyzing SL interactions. However, they have some limitations that could be addressed through further data expansion and model improvements. Our approach lacks transcription and translation regulation information, resulting in a relatively coarse representation of the underlying mechanisms and cellular specificity. To enhance the precision and explanatory power of the model, future versions could incorporate gene regulation networks and protein-protein interaction networks to provide a more detailed and comprehensive analysis. Furthermore, incorporating factors like the extracellular microenvironment and cell-cell interactions will be crucial when applying the model from tumor cells to the tumor tissue and microenvironment. These factors can influence SL interactions and should be considered to obtain a more realistic representation of SL interactions in tumors. Additionally, the current version of the model does not integrate mechanisms such as genetic epistasis and non-coding regulatory elements, which could enhance the accuracy of predictions once relevant data becomes available. Cells are highly complex, unique and specialized. The mechanisms underlying SL interactions are diverse and cell context-specific. It is important to acknowledge that the limited mechanistic logic employed in this study cannot capture all possible SL interactions. The focus here is on capturing the damage caused by differential gene expression levels and the responsible genes will result in SL interactions.

5. Conclusion

We presented a deep learning SL prediction method, SLWise, which combines graph-based representations, attention mechanisms, and multiple omics data to enhance its predictive power. The ablation study demonstrates that the GraphSAGE module effectively captures the representation of omics data. The transformer cross-attention mechanism is designed to assemble multi-source features, making it better at capturing the cell specific correlation of data and features. By integrating different biological data sources, our model can capture the complex relationships and interactions within the data, and thus outperform SOTA models in predicting cell-specific SL pairs for different cell-lines. The development of our approach is expected to be beneficial to the advancement of cancer precision medicine by supporting the discovery of cell-type specific drug targets and biomarkers in the future.

CRediT authorship contribution statement

M.P. and Y.Z. conceived the project. M.P., K.C. and X.L. designed the method and conducted the experiments. W.Z, L.W. G.P. and Y.X. curated the data. M.P., K.C. and X.L. pre-processed the data and analysed the results. S. J. assisted with data analysis. Q.T. assisted with data visualizations. M.P., X.L. and K.C. wrote the manuscript. All authors provided feedback on the manuscript.

Declaration of Competing Interest

The authors declare that they have no known competing interests.

Data Availability

The data underlying this article are available in the article and in its online supplementary material. The code and training data is available in https://github.com/promethiume/SLwise. The datasets analyzed for this study can be found in the L1000 datasets (https://clue.io/data/CMap2020#LINCS2020), the DepMap datasets (https://depmap.org/portal/download/all/) and the Integrative Onco genomics datasets (https://www.intogen.org/download).

Acknowledgements

We thank Dr. B. Fu and Dr. K. Tian from the Innovation Center for helpful discussions. We also thank Y. Wang, S. Guo, X. Xiang, and all members of the StoneWise AI. Inc. for administrative support.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2023.10.011.

References

[1] Warburg O. On the origin of cancer cells. Science 1956;123:309–14.
[2] Wang J, Zhang Q, Han J, Zhao Y, Zhao C, Yan B, et al. Computational methods, databases and tools for synthetic lethality prediction. Brief Bioinforma 2022;23: bbac106. https://doi.org/10.1093/bib/bbac106.
[3] Topatana W, Juengpanich S, Li S, Cao J, Hu J, Lee J, et al. Advances in synthetic lethality for cancer therapy: cellular mechanism and clinical translation. J Hematol Oncol 2020;13:1–22.
[4] Ashworth A, Lord CJ. Synthetic lethal therapies for cancer: what's next after PARP inhibitors? Nat Rev Clin Oncol 2018;15:564–76. https://doi.org/10.1038/s41571-018-0055-6.
[5] D'Andrea AD. Mechanisms of PARP inhibitor sensitivity and resistance. Dna Repair 2018;71:172–6. https://doi.org/10.1016/j.dnarep.2018.08.021.
[6] Bryant HE, Schultz N, Thomas HD, Parker KM, Flower D, Lopez E, et al. Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. Nature 2005;434:913–7. https://doi.org/10.1038/nature03443.
[7] Lord CJ, Ashworth A. BRCAness revisited. Nat Rev Cancer 2016;16:110–20. https://doi.org/10.1038/nrc.2015.21.
[8] Farmer H, McCabe N, Lord CJ, Tutt ANJ, Johnson DA, Richardson TB, et al. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. Nature 2005;434:917–21. https://doi.org/10.1038/nature03445.
[9] Diab A, Kao M, Kehrli K, Kim HY, Sidorova J, Mendez E. Multiple defects sensitize p53-deficient head and neck cancer cells to the WEE1 kinase inhibition. Mol Cancer Res 2019;17:1115–28. https://doi.org/10.1158/1541-7786.MCR-18-0860.
[10] Lecona E, Fernandez-Capetillo O. Targeting ATR in cancer. Nat Rev Cancer 2018; 18:586–95. https://doi.org/10.1038/s41568-018-0034-3.
[11] Li S, Topatana W, Juengpanich S, Cao J, Hu J, Zhang B, et al. Development of synthetic lethality in cancer: molecular and cellular classification. Signal Transduct Tar 2020;5:241. https://doi.org/10.1038/s41392-020-00358-6.
[12] Marjon K., Kalev P., Marks K. Cancer Dependencies: PRMT5 and MAT2A in MTAP/p16-Deleted Cancers. 2472–3428 2021;5:371–390. https://doi.org/10.1146/annurev-cancerbio-030419–033444.
[13] Jerby-Arnon L, Pfetzer N, Waldman YY, McGarry L, James D, Shanks E, et al. Predicting cancer-specific vulnerability via data-driven detection of synthetic lethality. Cell 2014;158:1199–209. https://doi.org/10.1016/j.cell.2014.07.027.
[14] Cai R, Chen X, Fang Y, Wu M, Hao Y. Dual-dropout graph convolutional network for predicting synthetic lethality in human cancers. Bioinformatics 2020;36: 4458–65. https://doi.org/10.1093/bioinformatics/btaa211.
[15] Wan F, Li S, Tian T, Lei Y, Zhao D, Zeng J. EXP2SL: a machine learning framework for cell-line-specific synthetic lethality prediction. Front Pharm 2020;11:112. https://doi.org/10.3389/fphar.2020.00112.
[16] De Kegel B, Quinn N, Thompson NA, Adams DJ, Ryan CJ. Comprehensive prediction of robust synthetic lethality between paralog pairs in cancer cell lines. Cell Syst 2021;12(1144–1159):e6. https://doi.org/10.1016/j.cels.2021.08.006.
[17] Long Y, Wu M, Liu Y, Zheng J, Kwoh CK, Luo J, et al. Graph contextualized attention network for predicting synthetic lethality in human cancers. Bioinformatics 2021;37:2432–40. https://doi.org/10.1093/bioinformatics/btab110.
[18] Huang J, Wu M, Lu F, Ou-Yang L, Zhu Z. Predicting synthetic lethal interactions in human cancers using graph regularized self-representative matrix factorization. Bmc Bioinforma 2019;20:657. https://doi.org/10.1186/s12859-019-3197-3.
[19] Liu Y, Wu M, Liu C, Li X-L, Zheng J. SL ^{2} MF: predicting synthetic lethality in human cancers via logistic matrix factorization. IEEE/ACM Trans Comput Biol Bioinf 2020;17:748–57. https://doi.org/10.1109/TCBB.2019.2909908.
[20] Wang S, Feng Y, Liu X, Liu Y, Wu M, Zheng J. NSF4SL: negative-sample-free contrastive learning for ranking synthetic lethal partner genes in human cancers. ii13–9 Bioinformatics 2022;38. https://doi.org/10.1093/bioinformatics/btac462.
[21] Wang S, Xu F, Li Y, Wang J, Zhang K, Liu Y, et al. KG4SL: knowledge graph neural network for synthetic lethality prediction in human cancers. Bioinformatics 2021; 37:i418–25. https://doi.org/10.1093/bioinformatics/btab271.
[22] Veličković P., Cucurull G., Casanova A., Romero A., Liò P., Bengio Y. Graph Attention Networks 2017. https://doi.org/10.48550/ARXIV.1710.10903.
[23] Lai M., Chen G., Yang H., Yang J., Jiang Z., Wu M., et al. Predicting Synthetic Lethality in Human Cancers via Multi-Graph Ensemble Neural Network. 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Mexico: IEEE; 2021, p. 1731–4. https://doi.org/10.1109/EMBC46164.2021.9630716.
[24] Kipf T.N., Welling M. Semi-Supervised Classification with Graph Convolutional Networks 2016. https://doi.org/10.48550/ARXIV.1609.02907.
[25] Hao Z, Wu D, Fang Y, Wu M, Cai R, Li X. Prediction of synthetic lethal interactions in human cancers using multi-view graph auto-encoder. IEEE J Biomed Health Inf 2021;25:4041–51. https://doi.org/10.1109/JBHI.2021.3079302.
[26] Cheng K, Nair NU, Lee JS, Ruppin E. Synthetic lethality across normal tissues is strongly associated with cancer risk, onset, and tumor suppressor specificity. Sci Adv 2021;7:eabc2100. https://doi.org/10.1126/sciadv.abc2100.
[27] Fan K, Tang S, Gökbağ B, Cheng L, Li L. Multi-view graph convolutional network for cancer cell-specific synthetic lethality prediction. Front Genet 2023;13: 1103092. https://doi.org/10.3389/fgene.2022.1103092.
[28] Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, et al. A next generation connectivity Map: L1000 platform and the first 1,000,000 profiles. e17 Cell 2017;171:1437–52. https://doi.org/10.1016/j.cell.2017.10.049.
[29] Janssen BD, Chen Y-P, Molgora BM, Wang SE, Simoes-Barbosa A, Johnson PJ. CRISPR/Cas9-mediated gene modification and gene knock out in the human-infective parasite Trichomonas vaginalis. Sci Rep-Uk 2018;8:270. https://doi.org/10.1038/s41598-017-18442-3.
[30] McFarland JM, Ho ZV, Kugener G, Dempster JM, Montgomery PG, Bryan JG, et al. Improved estimation of cancer dependencies from large-scale RNAi screens using model-based normalization and data integration. Nat Commun 2018;9:4610. https://doi.org/10.1038/s41467-018-06916-5.
[31] Meyers RM, Bryan JG, McFarland JM, Weir BA, Sizemore AE, Xu H, et al. Computational correction of copy number effect improves specificity of CRISPR–Cas9 essentiality screens in cancer cells. Nat Genet 2017;49:1779–84. https://doi.org/10.1038/ng.3984.
[32] Behan FM, Iorio F, Picco G, Gonçalves E, Beaver CM, Migliardi G, et al. Prioritization of cancer therapeutic targets using CRISPR–Cas9 screens. Nature 2019;568:511–6. https://doi.org/10.1038/s41586-019-1103-9.
[33] Dempster JM, Rossen J, Kazachkova M, Pan J, Kugener G, Root DE, et al. Extracting biological insights from the project achilles genome-scale CRISPR screens in cancer cell lines. Cancer Biol 2019. https://doi.org/10.1101/720243.
[34] Kim Y-A, Madan S, Przytycka TM. WeSME: uncovering mutual exclusivity of cancer drivers and beyond. Bioinformatics 2017;33:814–21. https://doi.org/10.1093/bioinformatics/btw242.
[35] Zamanighomi M, Jain SS, Ito T, Pal D, Daley TP, Sellers WR. GEMINI: a variational Bayesian approach to identify genetic interactions from combinatorial CRISPR screens. Genome Biol 2019;20:137. https://doi.org/10.1186/s13059-019-1745-9.
[36] Zhao D, Badur MG, Luebeck J, Magaña JH, Birmingham A, Sasik R, et al. Combinatorial CRISPR-Cas9 metabolic screens reveal critical redox control points dependent on the KEAP1-NRF2 regulatory axis. e7 Mol Cell 2018;69:699–708. https://doi.org/10.1016/j.molcel.2018.01.017.
[37] Dede M, McLaughlin M, Kim E, Hart T. Multiplex enCas12a screens detect functional buffering among paralogs otherwise masked in monogenic Cas9 knockout screens. Genome Biol 2020;21:262. https://doi.org/10.1186/s13059-020-02173-2.
[38] Ito T, Young MJ, Li R, Jain S, Wernitznig A, Krill-Burger JM, et al. Paralog knockout profiling identifies DUSP4 and DUSP6 as a digenic dependence in MAPK pathway-driven cancers. Nat Genet 2021;53:1664–72. https://doi.org/10.1038/s41588-021-00967-z.
[39] Najm FJ, Strand C, Donovan KF, Hegde M, Sanson KR, Vaimberg EW, et al. Orthologous CRISPR–Cas9 enzymes for combinatorial genetic screens. Nat Biotechnol 2018;36:179–89. https://doi.org/10.1038/nbt.4048.
[40] Oh J., Cho K., Bruna J. Advancing GraphSAGE with A Data-Driven Node Sampling 2019. https://doi.org/10.48550/ARXIV.1904.12935.
[41] Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., et al. Attention Is All You Need 2017. https://doi.org/10.48550/ARXIV.1706.03762.

[42] Du J, Zhang S, Wu G, Moura JMF, Kar S. Topology Adaptive Graph Convolutional. Networks 2017. https://doi.org/10.48550/ARXIV.1710.10370.
[43] DeWeirdt PC, Sangree AK, Hanna RE, Sanson KR, Hegde M, Strand C, et al. Genetic screens in isogenic mammalian cell lines without single cell cloning. Nat Commun 2020;11:752. https://doi.org/10.1038/s41467-020-14620-6.
[44] Menissier De Murcia J. Functional interaction between PARP-1 and PARP-2 in chromosome stability and embryonic development in mouse. EMBO J 2003;22: 2255–63. https://doi.org/10.1093/emboj/cdg206.
[45] Lui GYL, Shaw R, Schaub FX, Stork IN, Gurley KE, Bridgwater C, et al. BET, SRC, and BCL2 family inhibitors are synergistic drug combinations with PARP inhibitors in ovarian cancer. Ebiomedicine 2020;60:102988. https://doi.org/10.1016/j.ebiom.2020.102988.
[46] Karnak D, Engelke CG, Parsels LA, Kausar T, Wei D, Robertson JR, et al. Combined inhibition of Wee1 and PARP1/2 for radiosensitization in pancreatic cancer. Clin Cancer Res 2014;20:5085–96. https://doi.org/10.1158/1078-0432.CCR-14-1038.
[47] Parsels LA, Karnak D, Parsels JD, Zhang Q, Vélez-Padilla J, Reichert ZR, et al. PARP1 trapping and DNA replication stress enhance radiosensitization with combined WEE1 and PARP inhibitors. Mol Cancer Res 2018;16:222–32. https://doi.org/10.1158/1541-7786.MCR-17-0455.
[48] Santos JC, Profitós-Pelejà N, Sánchez-Vinces S, Roué G. RHOA therapeutic targeting in hematological cancers. Cells 2023;12:433. https://doi.org/10.3390/cells12030433.
[49] De Jong MRW, Langendonk M, Reitsma B, Herbers P, Nijland M, Huls G, et al. WEE1 inhibition enhances anti-apoptotic dependency as a result of premature mitotic entry and DNA damage. Cancers 2019;11:1743. https://doi.org/10.3390/cancers11111743.
[50] Wang J, Zhang Y, Dou Z, Jiang H, Wang Y, Gao X, et al. Knockdown of STIL suppresses the progression of gastric cancer by down-regulating the IGF-1/PI3K/AKT pathway. J Cell Mol Med 2019;23:5566–75. https://doi.org/10.1111/jcmm.14440.
[51] De Kegel B, Ryan CJ. Paralog dispensability shapes homozygous deletion patterns in tumor genomes. Cancer Biol 2022. https://doi.org/10.1101/2022.06.20.496722.
[52] Ryan CJ, Mehta I, Kebabci N, Adams DJ. Targeting synthetic lethal paralogs in cancer. Trends Cancer 2023;9:397–409. https://doi.org/10.1016/j.trecan.2023.02.002.
[53] Xin Y, Zhang Y. Paralog-based synthetic lethality: rationales and applications. Front Oncol 2023;13:1168143. https://doi.org/10.3389/fonc.2023.1168143.
[54] El Tekle G, Bernasocchi T, Unni AM, Bertoni F, Rossi D, Rubin MA, et al. Co-occurrence and mutual exclusivity: what cross-cancer mutation patterns can tell us. Trends Cancer 2021;7:823–36. https://doi.org/10.1016/j.trecan.2021.04.009.
[55] Köferle A, Schlattl A, Hörmann A, Thatikonda V, Popa A, Spreitzer F, et al. Interrogation of cancer gene dependencies reveals paralog interactions of autosome and sex chromosome-encoded genes. Cell Rep 2022;39:110636. https://doi.org/10.1016/j.celrep.2022.110636.
[56] Shields JA, Meier SR, Bandi M, Mulkearns-Hubert EE, Hajdari N, Ferdinez MD, et al. VRK1 is a synthetic–lethal target in VRK2-deficient glioblastoma. Cancer Res 2022;82:4044–57. https://doi.org/10.1158/0008-5472.CAN-21-4443.
[57] Zhang Y, Remillard D, Onubogu U, Karakyriakou B, Asiaban JN, Ramos AR, et al. Collateral lethality between HDAC1 and HDAC2 exploits cancer-specific NuRD complex vulnerabilities. Nat Struct Mol Biol 2023;30:1160–71. https://doi.org/10.1038/s41594-023-01041-4.