DSCN: double-target selection guided by CRISPR screening and network

Enze Liu[1,2,3], Xue Wu[2], Lei Wang[2], Yang Huo[2,3], Huanmei Wu[4], Lang Li[2], and Lijun Cheng[2*]

[1]Division of Hematology and Oncology, School of Medicine, Indiana University, Indianapolis, Indiana, USA

[2]Department of Biomedical Informatics, College of Medicine, The Ohio State University, Columbus, Ohio, USA

[3]School of Informatics and Computing, Indiana University, Indianapolis, Indiana, USA

[4]College of Public Health, Temple University, Philadelphia, Pennsylvania, USA

*Corresponding author:
Lijun Cheng, Ph.D.
Department of Biomedical Informatics, College of Medicine
The Ohio State University
340H Lincoln Tower, 1800 Cannon Drive
Columbus Ohio 43210
Telephone: (614) 366-4980
Email: Lijun.Cheng@osumc.edu

We declare no potential conflict of interest.

Abstract

Cancer is a complex disease with usually multiple disease mechanisms. Target combination is a better strategy than a single target in developing cancer therapies. However, target combinations are generally more difficult to be predicted. Current CRISPR-cas9 technology enables genome-wide screening for potential targets, but only a handful of genes have been screend as target combinations. Thus, an effective computational approach for selecting candidate target combinations is highly desirable. Selected target combinations also need to be translational between cell lines and cancer patients.

We have therefore developed **DSCN (double-target selection guided by CRISPR screening and network)**, a method that matches expression levels in patients and gene essentialities in cell lines through spectral-clustered protein-protein interaction (PPI) network. In DSCN, a sub-sampling approach is developed to model first-target knockdown and its impact on the PPI network, and it also facilitates the selection of a second target. Our analysis first demonstrated high correlation of the DSCN sub-sampling-based gene knockdown model and its predicted differential gene expressions using observed gene expression in 22 pancreatic cell lines before and after MAP2K1 and MAP2K2 inhibition ($R^2 = 0.75$). In our DSCN algorithm, various scoring schemes were evaluated. The 'diffusion-path' method showed the most significant statistical power of differentialting known synthetic lethal (SL) versus non-SL gene pairs ($P = 0.001$) in pancreatic cancer. The superior performance of DSCN over existing network-based algorithms, such as OptiCon[1] and VIPER[2], in the selection of target combinations is attributable to its ability to calculate combinations for any

gene pairs, whereas other approaches focus on the combinations among optimized regulators in the network. DSCN's computational speed is also at least ten times faster than that of other methods. Finally, in applying DSCN to predict target combinations and drug combinations for individual samples (DSCNi), we showed high correlation of DSCNi predicted target combinations with synergistic drug combinations (P= 1e-5) in pancreatic cell lines. In summary, DSCN is a highly effective computational method for the selection of target combinations.

Author Summary

Cancer therapies require targets to function. Compared to single target, target combination is a better strategy for developing cancer therapies. However, predicting target combination is much complicated than predicting single target. Current CRISPR technology enables whole genome screening of potential targets. But most of the experiments have been conducted on single target (gene) level. To facilitate the prediction of target combinations, we developed DSCN **(double-target selection guided by CRISPR screening and network)** that utilize single target-level CRISPR screening data and expression profiles for predicting target combinations by connecting cell-line omics-data with tissue omics-data. DSCN showed great accuracy on different cancer types and superior performance compared to existing network-based prediction tools. We also introduced DSCNi derived from DSCN and designed specific for predicting target combinations for single-paitent. We showed synergistic target combinations predicted by DSCNi accurately reflected synergies on drug combination levels. Thus, DSCN and DSCNi have the potential be further applied in personalized medicine field.

67

68 # Introduction

69 The complexity of cancer is widely recognized, with heterogeneous disease mechanisms
70 underlying primary, metastatic, and drug-resistant tumors [3, 4]. Therefore, translational
71 cancer research now focuses on the identification of combinational rather than single targets
72 and the selection of drug combinations instead of single drugs [5, 6]. Synthetic lethality (SL),
73 a key concept in the simultaneous targeting of two genes that contribute to tumor vulnerability
74 [7], requires the loss of both genes in a pair to be lethal to a cancer cell. A CRISPR-based
75 double knockout (CDKO) system has recently been developed to effectively screen gene pairs
76 or target combinations [8, 9]. In this paper, we will use the terms gene pair and target
77 combination interchangeably because they represent equivalent concepts. Screening using
78 the CDKO system, however, is limited by the number of genes to be screened. For instance,
79 if we screen target combinations among 100 genes, and each gene has four gRNAs, there
80 will be $(4 \times 100)^2/2 = 80{,}000$ combinations, a scale that is feasible in a CDKO system.
81 However, across the genome, if we screen target combinations among 10,000 genes and
82 select only one gRNA per gene, the resulting $10{,}000^2/2 = 50{,}000{,}000$ combinations will be
83 practically infeasible. Therefore, a computational approach is very much needed to rank and
84 select top candidate gene pairs for CDKO analysis.

85

86 The many computational methods developed to identify potential candidate SL gene pairs fall
87 into two major categories: machine learning and statistical inference. The machine-learning

approach has a much longer history, with the generation of large-scale double knockout data in yeast in 2004 [10]. Several methods, including multiple network decision tree [11], protein interaction network [12], and multi-network and multi-classifier [13] approaches, have demonstrated the significant predictive performance of SL gene pairs using features derived from network topology, gene ontology, and gene function sets. Recently, researchers applied a systems-biology framework called 'mashup' that allows for the acquisition and synthesis of data from diverse sources [14]. This method performed even better than the other network analyses in predicting SL gene pairs, further demonstrating the ability of such computational algorithms as random walk with restart to integrate and characterize the biological network topology successfully. Group-sparse collective matrix factorization (gCMF), another unique machine-learning method and recent major contribution to SL prediction [15], performed matrix factorizations among input data, such as gene expression, mutations, copy number variations (CNV), and CDKO, and identified a shared sub-matrix in which SL gene pairs can be predicted. Its performance was comparable to that of the mashup approach in several CDKO datasets derived from human cancer cell lines.

Statistical inference, on the other hand, relies on strong biological assumptions regarding the mechanisms of synthetic lethality. These methods infer SL gene pairs utilizing multi-omics data, such as CNV, mutations, gene expression, and single gene essentiality generated from CRISPR screening. In particular, CDKO data are NOT employed to train SL prediction in the statistical inference. DAISY is a notable early SL inference method [16], in which one primary

assumption is that if the cancer cell is viable, the SL pair comprises one gene that is both active and essential if the other is inactive. MiSL (mining synthetic lethals), another statistical-inference approach [17], assumes that if one gene in an SL gene pair is inactive, the other must be active, and its activity is demonstrated through concordant changes in CNV and gene expression. Several other methods, such as ISLE (identification of clinically relevant synthetic lethality) [18], DiscoverSL [19], and ASTER (analysis of synthetic lethality by comparison with tissue-specific disease-free genomic and transcriptomic data) [20] were developed similarly, each making different biological assumptions regarding SL gene pairs.

Considering network topology, recent systems-biology-based statistical inference methods differed significantly from the other established SL statistical-inference methods. Here, we highlight two notable approaches, OptiCon (optimal control nodes) [1] and VIPER (virtual inference of protein activity by enriched regulon analysis) [2]. Both approaches primarily utilize gene-expression data to construct a biological network and rank and select target combinations that demonstrate optimal control of the network. OptiCon relies on a protein-protein interaction (PPI) network and models both signaling transduction and gene regulation during the selection of target combinations, and VIPER focuses on a gene-regulatory network model derived from mutual information among genes. Both approaches assume that the more a combination of two gene targets controls the network, the more likely those targets will be an SL pair.

Both statistical inference and machine learning have their advantages and disadvantages. A

machine-learning approach optimizes prediction based on training data, but the quantity of training data limits the validity of its prediction of SL gene pairs. Extrapolation of the SL gene pairs from machine-learning prediction to other pathways will be challenging because most CDKO data generated from human cancer cells are sparse and biased toward several specific pathways. However, the prediction of genome-wide SL gene pairs using statistical-learning approaches relies strongly on biological assumptions and is technically unbiased. This can be particularly useful in the case of very limited CDKO data.

In this paper, we describe a new statistical inference method we have developed called DSCN (double target selection guided by CRISPR screening and network). It more resembles OptiCon and VIPER than other methods, such as DAISY or MiSL. Similar to OptiCon and VIPER, DSCN is built upon a biological network and transcriptome data, and the top combination targets are ranked and selected based on their control of or impact on the network.

DSCN differs from OptiCon and VIPER in its use of single-gene-library-based CRISPR-cas9 screening data, its focus on the overlapped networks between the cancer cell line and the corresponding primary tumor data, and most important, its consideration of the sequential selection of two targets, which involves the perturbation of transcriptome data for selection of the second target after selection of the first. This third model strategy, we believe, will make the selection of combination targets by DSCN more closely resemble the true biology. DSCN is also built upon our early research in the selection of single targets, SCNrank (spectral clustering for network-based ranking) [21], which ranks and selects consensus single-gene

153 targets between cancer cell lines and tumor samples.

154

155 # Materials and Methods

156 **Table 1.** Datasets used in this study

157

158 **Tables 1** details our data sources, including the types of cancer screened, data platforms and
159 types, and sample numbers. We retrieved gene-expression and -mutation data for normal
160 tissue and tumor samples for pancreatic and breast cancers from the Gene Expression
161 Omnibus (GEO) [22, 23] and The Cancer Genome Atlas (TCGA) [24] and gene-expression
162 and -essentiality data from the Cancer Cell Line Encyclopedia (CCLE) [13] and Project Achilles
163 [25-27], downloaded PPI data from STRING [28], extracted drug-target data from DrugBank
164 [12], and downloaded synthetic lethal gene-pair data from the SynlethDB database [29] and
165 drug-sensitivity data from the DrugComb database [30].

166

167 **Steps of DSCN algorithm**

168 DSCN algorithm consists of six steps (**Figure 1**):

169

170 **Figure 1.** Overview of double-target selection guided by CRISPR screening and network
171 (DSCN)

172

173 Step 1: Network construction

174 In this step, we construct two integrated function networks, a tissue network G_t and a cell-line
175 network G_c. G_t consists of a skeleton from the STRING PPI network and edge weights from

gene pair-wise Pearson correlations in tumor samples, and node weights are the fold changes in gene expression between tumors and normal tissue. A high fold change indicates higher gene expression in the tumor than in normal tissue. Assume that there are a total of n genes (nodes) in G_t. The affinity matrix S_t denotes the edge weights, and diagonal matrix D_t denotes the node weights in Equation (1):

$$G_t = S_t + D_t, \quad S_t = \begin{pmatrix} 0 & \cdots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{n1} & \cdots & 0 \end{pmatrix}, \quad D_t = \begin{pmatrix} w_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & w_n \end{pmatrix}, \tag{1}$$

where $w_{\mathrm{ab}}, a \neq b \in (1,n)$ in S_t indicates the edge weight (correlation) between genes a and b in the tissue network; and w_i in D_t is the tumor versus normal fold change in the expression of gene $i, i = 1,\ldots,n$.

Similarly, G_c consists of an identical skeleton from the same STRING PPI network and edge weights from pair-wise gene correlations in cell-line samples. Unlike G_t, the node weight of G_c is from CRISPR-Cas9 screening data, which is indicated as the gene essentiality value. The gene essentiality value can be generally interpreted as the fold change in cell count before and after gene knockout. Genes demonstrating smaller fold change are more essential. In this study, all the essentiality values are log2 transformed. Similarly, G_c is decomposed into affinity matrix S_c for edge weight and diagonal matrix D_c for node weight in the cell-line network $G_c = S_c + D_c$.

Step 2: Construction of Laplacian matrices for the tissue and cell-line networks

A Laplacian matrix measures all properties of a network, including node weight, edge weight, and connectivity. In this second step, we construct Laplacian matrices for the tissue network G_t and cell-line network G_c as:

$$L = D - S, \quad (2)$$

in which *D* is the diagonal matrix and *S,* the affinity matrix, defined in Equation (1), and L_t is the Laplacian matrix for the tissue network and L_c, that for the cell-line network.

Step 3: Spectral clustering for tissue network

We perform spectral clustering only on the Laplacian matrix of the tissue network L_t as:

I. Normalize the Laplacian matrix L_t to L'_t:

$$L'_t = \begin{pmatrix} w_1 & \cdots & -\frac{abs(w_{1n}w_1)}{\sum_{k=1}^{n} abs(w_{1k})} \\ \vdots & \ddots & \vdots \\ -\frac{abs(w_{n1}w_1)}{\sum_{k=1}^{n} abs(w_{nk})} & \cdots & w_n \end{pmatrix} \quad (3)$$

In the normalized Laplacian matrix L'_t, all diagonal elements are positive, and all other elements are negative. The row sum of non-diagonal elements is equal to its corresponding diagonal.

II. Perform eigen decomposition for matrix L' to obtain the spectrum E = $\{\lambda_1, \lambda_2 ... \lambda_n\}$, where $0 = \lambda_1 \leq \lambda_2 \leq ... \leq \lambda_n$, and their corresponding eigenvector.

III. Choose the k smallest non-negative eigenvalues $\{\lambda_i,...,\lambda_{i+k}\}$ and their corresponding eigenvectors, and combine these k eigenvectors into an $n \times k$ matrix, H.

IV. In this H eigenvector matrix, each row represents a gene node, and k columns represent the coordinate values of a gene node. The row vectors in H are used to calculate the Euclidean distance between a pair of gene nodes. We then perform K-means clustering for n nodes. To select the number of clusters, K', to produce a good fit, we calculate

Hartigan's number, which measures the quality of clustering results. We select the optimal K' and constrain it further to less than 10 for practical consideration. This spectral clustering leads to K' exclusive clusters (i.e., subnetworks). From the tissue network G_t, subnetworks $g_{t1},...g_{tK'}$ are classified.

Step 4: Mapping the tissue/cell-line network and calculating the impact score of Target 1

The cell-line network G_c is then mapped to the spectral clusters, $g_{t1},...g_{tK'}$, generated from tissue network G_t in Step 3. Because tissue network G_t and cell-line network G_c share the identical network structure, i.e., nodes and connections, G_t subnetworks, $\{g_{t1},...g_{tK'}\}$ are mapped to G_c subnetworks $\{g_{c1},...g_{cK'}\}$ using their common node names and connections.

The target impact score will be calculated based on the cell-line subnetworks $\{g_{c1},...g_{cK'}\}$. We focus on all Food and Drug Administration (FDA)-approved drug targets (see **Table 1**) to calculate our target score. The impact score of a target 1 $(T1)$ is calculated as the sum of the impact score itself and its impact on the rest of the genes in the network. Its general form is defined in Equation (4):

$$IS(T_1) = S(T_1) + \sum_{i\in\{1,..,n\}} S[N_i \,|\, Pa(N_i)], \qquad (4)$$

in which, $\{N_i, i = 1,...n\}$ are the gene nodes in the network other than $T1$, and $Pa(N_i)$ is a set of parent nodes of N_i. In particular, the impact score on N_i depends on its parent nodes, $Pa(N_i)$. **Figure 2** illustrates the three different methods of calculating the impact score–the most-probable, random-walk, and diffusion paths.

Most-probable path: The immediate children of $T1$ are the gene nodes directly connected to $T1$, e.g., N4 is the direct child $T1$ in **Figure 2b**. In this method, we will count only the immediate children of $T1$ in calculating the impact score. Without loss of generality, let $ch(T1)$ be the set of immediate children of $T1$. The most probable path of $T1$ is the one that has the smallest impact score among $ch(T1)$. Based on the general impact score as calculated in Equation (4), the most-probable-path impact score is defined in Equation (5):

$$IS\ (T1) = S(T1) + min_{N_i \in ch(T1)}\, S[N_i \,|\, T1]$$

$$= w_{T1} + min_{N_i \in ch(T1)}(w_{N_i} \times w_{T1,N_i}), \qquad (5)$$

where w_{T1} and w_{N_i} indicate their node weights, and w_{T1,N_i} indicates their edge weight.

Figure 2. Network configurations for three methods to calculate impact score

Random walk path: The random-walk score is calculated in two steps. Step 1 is a random walk in the network, in which the random walk has a transition probability of traveling from one node to another. In **Figure 2c**, starting from $T1$, each node N_i is randomly visited. Here we used normalized edge weight for transition probability as defined in Equation (6):

$$P_{j,i} = \frac{w_{j,i}}{\sum_{x \in e} w_{j,x}}\ , \qquad (6)$$

where $P_{j,i}$ is the transition probability from N_j to N_i, $w_{j,i}$ is the edge weight between them, and $\sum_{x \in e} w_{j,x}$ is the sum of all edge weights of N_j. In this Markov process, a node can be visited multiple times. We set the total number of random-walk steps as $2n$, where n is the total number of nodes in the network.

Then, in Step 2, we defined the parent node as the node that visited N_i first, i.e., $Pa(N_i)$. Hence, the impact score of $T1$ becomes:

$$IS(T1) = S(T1) + \sum_{i\in\{1,..,n\}} S[N_i \,|\, Pa(N_i)]$$
$$= S(T1) + \sum_{i\in\{1,..,n\}} w_i \times w_{i,Pa(N_i)}. \quad (7)$$

Diffusion path: Starting from $T1$, each node is visited in a hierarchical order. Therefore, the parent nodes of a node, N_i, can be from the upper tier, i.e., $UpperTier\ (N_i)$, or the same tier, i.e., $SameTier\ (N_i)$. For instance, in **Figure 2d**, there are four tiers in the hierarchical structure starting from $T1$. The impact of $T1$ transmits from Tier 1 to Tier 4 in the network. Therefore, the impact score is defined in Equation (8):

$$IS(T1) = S(T1) + \sum_{i\in\{1,..,n\}} S[N_i \,|\, Pa(N_i)]$$
$$= S(T1) + \sum_{i\in\{1,..,n\}} \{\sum_{j\in UpperTier} W_{ij}W_i + \sum_{w\in SameTier} W_{iw}W_i\} \quad (8)$$

Step 5: Subsampling and Target 2 (*T2*) score and selection

Once *T1* is selected, we remove cancer cell lines with higher expression of the *T1* than its sample mean and only keep cell lines with its expression lower than mean.This subsampling method characterizes the knockdown of the *T1* . Similarly, we also remove cancer cell lines with higher *T1* essentiality scores than the sample in our subsampling. After the resampling, we construct the cell-line network G_c as Equation (2) using the subsampled cell-line subsamples. We follow the same **Step 3** in mapping G_c to $\{g_{t1},...g_{tK'}\}$ and calculate the *T2* impact score following the same algorithms defined in **Step 4**. The *T2* impact score is then denoted as $IS\ (T2|T1)$, because the subsampling and network depend on $T1$.

Step 6: Calculation of impact score for target combinations

Because $T1$ and $T2$ and their impact scores are computed sequentially, the combinational impact score will consider both sequential orders in Equation (9), in which $T1 \neq T2$:

$$IS\,(T1, T2) = IS(T1) + IS(T2 \,|\, T1) \tag{9}$$

Tissue cell-line subnetwork similarity measure

We measure the similarity of each subnetwork pair <g_{t_i},g_{c_i}>, $i \in (1,\ldots,K')$ between tissue and cell-line using the following scheme:

I. Normalization of node weight (diagonal)

To make two subnetworks, g_{t_i} and g_{c_i}, comparable, we normalize the cell-line diagonal matrix D_{c_i} according to the tissue diagonal matrix D_{t_i} using the following formula:

$$D'_{c_i} = \begin{pmatrix} \frac{w_{c,i,1}\sum_{j=1}^{J} w_{t,i,j}}{\sum_{j=1}^{J} w_{c,i,j}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{w_{c,i,j}\sum_{j=1}^{J} w_{t,i,j}}{\sum_{j=1}^{J} w_{c,i,j}} \end{pmatrix}, \tag{10}$$

in which $w_{c,i,j}$ denotes the node weight $j \in (1,J)$ in the cell-line subnetwork, and $w_{t,i,j}$, that in the tissue subnetwork. J is the total number of nodes in g_{c_i} and g_{t_i}.

II. Normalization of edge weight

The Laplacian matrices for each subnetwork pair, <g_{t_i},g_{c_i}> , $i \in (1,K')$, are defined similarly as Equation (3): $L_{t_i} = D_{t_i} - S_{t_i}$ and $L_{c_i} = D'_{c_i} - S_{c_i}$. After node-weight normalization, trace (L_{c_i}) = trace (L_{t_i}). Then, their edge weights (non-diagonal elements) are normalized accordingly using the formula:

$$L'' = \begin{pmatrix} w_1 & \cdots & \frac{w_{1J}abs(w_1)}{\sum_{j=1}^{J} abs(w_{1j})} \\ \vdots & \ddots & \vdots \\ \frac{w_{J1}abs(w_1)}{\sum_{j=1}^{J} abs(w_{Jj})} & \cdots & w_J \end{pmatrix} . \quad (11)$$

Until this step, all edges (non-diagonal elements) in both Laplacian matrices, L''_{t_i} and L''_{c_i}, acquired node features during normalization. We keep the original directions (positive or negative) of node weights and edge weights for the following distance calculation.

III. Distance calculation

For two corresponding subnetworks g_{t_i} and g_{c_i} in tissue and cell-line, we calculate the distance using their normalized Laplacian matrices L''_{t_i} and L''_{c_i}:

$$Distance(g_{t_i}, g_{c_i}) = \sum_{j=1}^{J} \sum_{l=1}^{J} (L''_{t_i}(j,l) - L''_{c_i}(j,l))^2, \ l \neq j, \quad (12)$$

where $L''(i,j)$ $i \neq j$ indicates the edge weight between nodes *I* and *j* in a given Laplacian matrix, and $(L''_{t_i}(i,j) - L''_{c_i}(i,j))^2$ indicate the Euclidean distance between the same edges in two Laplacian matrices.

Construction of a DSCN algorithm for an individual cancer cell-line sample (DSCNi)

We apply DSCNi algorithm for scoring target combinations in a single cancer cellline for a single patient. Very similar to DSCN, in building up G_c, DSCNi relies on a set of expression profiles for a cancer cell line to calculate the edge weights (i.e., correlations) between gene nodes. However, unlike DSCN, DSCNi uses a cell-line-specific essentiality score for node weights. Its impact score calculation for *T1*, $IS(T1)$, follows exactly from **Steps 1, 2, 3, and 4.** In modeling the knockdown of *T1* in the subsampling in **Step 5**, we maintain the same *T1* subsampling as DSCN, i.e., we remove samples with higher expression of T1 than its sample

mean. However, we will keep the same essentiality score for this individual cancer cell-line sample to calculate the Target 2 impact score. We calculate the final combination target impact score similarly as in DSCN, such that it has a comparable meaning to that calculated from DSCN.

Analysis of association between drug- and target-combination synergy

The Bliss score [32] measures the synergistic effect of a drug combination, i.e., the effect of the drug combination on cell viability rather than the additive effects of its two component drugs. A two-drug combination is considered synergistic if its Bliss score exceeds 0.12 [33]. On the other hand, the target combination is predicted to be synergistic if the impact score of the drugs in combination is larger than the additive score of the constituent drugs, as in Equation (13), in which the impact scores of $IS(T1,T2)$, $IS(T1)$ and $IS(T2)$ are calculated by (9) and (8):

$$IS(T1,T2) > IS(T1) + IS(T2) \quad (13)$$

In this section, we will define an association analysis between drug-combination scores and target-combination synergy scores. Consider a cancer cell line screened by a set of drug combinations, and these drug combinations can be categorized as either synergistic or non-synergistic based on their Bliss scores. Then, for each drug combination, we identify all its two-target combinations, calculate their synergy scores, and classify the drug combinations as either synergistic or not as in Equation (13). In a 2 by 2 contingency table, the rows are drug synergy (Y/N), and columns are target synergy (Y/N). For each drug combination, all

counts of target-combination synergy and non-synergy are added to the corresponding row with respect to drug-combination synergy or non-synergy. The association between drug- and target-combination synergy is tested using a Chi-square test.

Results

Validation of the subsampling scheme for determining the impact of target-gene knockdown in the DSCN algorithm

In the DSCN algorithm, we designed our subsampling method (**Step 5)** to model the impact of Target 1 knockdown in the cancer cell line. To demonstrate the validity of this sampling scheme, we identified a GEO dataset, GSE45757, that provided transcriptome profiles across 22 pancreatic cell lines before and after MAP2K1 and MAP2K2 inhibition. Our analysis focused on 1,301 neighbor genes of MAP2K1 and MAP2K2 in the PPI network. Using the subsampling approach, we calculated the log-fold changes in these 1,301 genes between groups with either high or low expression of MAP2K1 and MAP2K2 group, which represent the predicted impact of Target 1 knockdown in the subsampling scheme. On the other hand, the observed log-fold changes in these 1,301 gene expressions were calculated during MAP2K1 and MAP2K2 inhibition. **Figure 3** shows a strong correlation, $R^2 = 0.75$, between the predicted and observed fold changes among these 1,301 neighbor genes of MAP2K1 and MAP2K2. Findings of this analysis strongly support subsampling as a valid model for determining the impact of target-gene knockdown.

Figure 3. Correlation between the predicted and observed log-fold changes in gene expression among MAP2K1 and MAP2K2 neighbor genes in the protein-protein interaction (PPI) network

Comparison of impact scores of target combinations using known synthetic lethal gene pairs in pancreatic cancers

We proposed three different scoring schemes to model the impact of target-gene knockdown on the network–those of the most-probable, random-walk, and diffusion paths. In addition, the impact score can be calculated based on either the global or local PPI network. The local PPI network is the product from spectral clustering of the whole genome PPI network (global network). To compare the performance of these impact scores, we used the 23 reported synthetic lethal pancreatic gene pairs in SynlethDB [29] as benchmarks. We compared impact scores between them and the other 164 gene pairs, which were derived from 21 unique genes among the 23 SL gene pairs. We constructed a tissue-function network using 153 tumor and 58 normal expression profiles of the pancreas from the GEO database (**Table 1**) and a cell-line function network using CRISPR screening data of 26 pancreatic cell lines from Project Achilles and 92 pancreatic tumor cell-line expression profiles from the GEO database (**Table 1**). All expression profiles are generated by Affymetrix U1332.0 microarray.

Smaller impact scores indicated the stronger impact of the gene knockdown on the network. Calculation of the impact scores using the local network generated from spectral clustering revealed significant difference in diffusion-path-based impact scores (IS) between synthetic and non-synthetic lethal gene pairs (*P*-values) as well as lower impact scores of synthetic than

non-synthetic lethal gene pairs. We observed the same trends with the other two impact scoring schemes, the most-probable and random-walk paths, i.e., lower IS score in the synthetic than non-synthetic lethal gene pairs that were not statistically significant.

Calculation of the impact scores using the global network and diffusion-path scoring scheme also yielded lower diffusion impact scores in the synthetic than non-synthetic gene pairs, though the differences were not statistically significant. The scores of the most-probable and random-walk paths, on the other hand, showed the reverse direction between synthetic and non-synthetic gene pairs. We therefore believe that using the diffusion-path and local networks, evaluation of the target-combination impact score is an ideal approach in selecting synthetic lethal gene pairs.

Figure 4. Comparison of target-combination impact scores using synthetic versus non-synthetic lethal gene pairs in pancreatic cancer.

Compare the selection of target combinations among DSCN, OptiCon, and VIPER

We compared the performance of DSCN with that of two existing algorithms for the selection of target combinations–OptiCon and VIPER. Both of these use transcriptome profiles to select combination targets, and their top target combinations are master regulators of synergy that have optimal control of their corresponding networks. OptiCon requires tumor transcriptome profiles and corresponding mutation data as input to infer master regulators and predict synergies among them, whereas VIPER uses transcriptome profiles from both tumor and normal samples to select regulons and infers synergies among the regulons. Because the

pancreas microarray expression profile used in the previous section has no corresponding mutation information, we utilized pancreatic expression profiles in TCGA to construct a tissue function network. We used 179 pancreatic tumor expression profiles along with their mutation data and 41 adjacent normal expression profiles (Table 1). We also used expression profiles of 92 pancreatic tumor cell lines from GEO and CRISPR-screening data of 26 pancreatic cell lines from Project Achilles (Table 1). Together, these data served for benchmark comparison of the performance of the three algorithms.

In total, DSCN predicted 37,275 synergistic target combinations, OptiCon, 2,778 , and VIPER, 191. After mapping them onto all 12,821 synthetic lethal gene pairs within SynlethDB, neither OptiCon nor VIPER showed any overlap. However, DSCN demonstrated overlap of 936 target combinations. Among the 936 overlapped SL pairs, 79 were annotated as SL pairs specific to pancreatic ductal adenocarcinoma (PDAC). Of these 79, their predicted IS scores showed a 0.34 Spearman correlation with their SynlethDB score ($P < 0.01$), and the predicted IS scores were significantly lower than that of 6,162 random combinations on t-test ($P = 0.05$). These 6,162 random combinations were derived from genes in the 79 SL pairs, but 79 were removed.

These benchmark comparison analyses were performed on Indiana University's supercomputer, 'Carbonate'. DSCN completed its search of target combinations on the single central processing unit core in 12 hours, a significantly faster speed than those using OptiCon (320 hours) and VIPER (141 hours). OptiCon mainly performs two computational tasks, calculating subnetworks and null distributions, each of which uses about 160 hours. Most of

the computational time of VIPER, on the other hand, involves the generation of the transcriptome mutual information network using ARACNe [34], a classical tool of reconstructing regulatory network by calculating pair-wise mutual information between genes.

Top-ranked target combinations and their associations with overall survival in patients with pancreatic cancer

We used expression profiles of tissues and cell lines from the GEO database (**Table 1**) to construct function networks and predict impact scores. Our dataset consisted of expression profiles of 153 tumors and 58 normal pancreas samples from GEO, CRISPR-screening data of 26 pancreatic cell lines from Project Achilles, and 92 pancreatic tumor cell-line expression profiles from the GEO database. This yielded 14,066 overlapped genes.

In this analysis, we focused on 1,437 drug targets of all FDA approved drugs in DrugBank and calculated their possible target combinations. Most interestingly, all genes in the top 230 target combinations are within the same subnetwork–the PDAC tissue subnetwork (**Supplementary Figure 1 A**) and cell-line subnetwork (**Supplementary Figure 1 B**). **Supplementary File 1** includes the full list of genes in this subnetwork.

Figure 5. Kaplan-Meier curves for the nine top-ranked target combinations.

Table 2 displays the nine top-ranked target combinations and their annotations. Their Kaplan-

Meier curves (**Figure 5**) are generated using TCGA PDAC clinical annotations from the Gene Expression Profiling Interactive Analysis (GEPIA) database [35]. Patient samples are categorized into two groups based on a target combination in which both genes are expressed either above (i.e., high-2) or below their means (i.e., low-2). Using log-rank test and Cox proportional hazard model to analyze the association between the expression of a target combination (high-2 versus low-2) and overall survival of patients with PDAC, we observed significant survival difference ($P < 0.05$, Table 2) of three of the nine top-ranked target combination comparisons, (EGLN1, TRFC), (FRK, TRFC), and (XDH, TRFC), their overall survival was worse for patients with high expression of these two genes than those with low expression.

Interestingly, seven of the top nine target combinations include transferrin receptor (TFRC), which encodes a surface receptor responsible for cellular iron intake. High expression of TFRC in PDAC and its strong association with PDAC growth and survival have been reported [36]. Recent studies suggest several key pathways of ferroptosis induction, including mitogen-activated protein kinases (MAPK) and reactive oxygen species (Ros) pathways [37]. Hence, targeting upstream genes (e.g., MAP2K2, EGLN2) along with downstream genes (e.g., TFRC, FTL) might lead to a synergistic effect.

Table 2. Analysis of overall survival among the nine top-ranked target combinations in pancreatic ductal adenocarcinoma (PDAC)

482 **Performance of DSCNi in predicting drug synergy in cancer cell lines**

483 DSCNi predicts target combinations for individual patients using gene-expression and -
484 essentiality profiles. In this study, we assessed whether DSCNi predicted any association
485 between target- and drug-combination synergy at each individual cell-line level. DrugComb
486 [30] is a comprehensive database that incorporates information regarding the synergy of drug
487 combinations from numerous well-known projects, such as the National Cancer Institute
488 (NCI)-60 [38] for Human Tumor Cell Lines Screen. Because DrugComb includes only one
489 PDAC cell line with five associated combinational drug treatments, we decided to use the cell-
490 line data of triple-negative breast cancer (TNBC). We used 115 TNBC expression profiles from
491 TCGA to generate edge weights in the tissue-function network, 12 TNBC cell lines from the
492 Cancer Cell Line Encyclopedia (CCLE) database [39] to generate edge weights for the cell-
493 line function network, and CRISPR screening data of the TNBC cell line “HS578T” from Project
494 Achilles to generate node weights in the cell-line function network. Among all TNBC cell-lines,
495 HS578T has the largest number (N = 5,226) of drug-combination screening data in the
496 DrugComb database, and our focus on drugs with known targets in DrugBank led to screening
497 data for1,031 drug combinations in the HS578T cell line.*)* In turn, these drug combinations
498 correspond with 14,066 target combinations in our network model (**Supplementary File 2**).
499
500 To measure the association between predicted synthetic lethal pairs and synergistic drug
501 combinations, we constructed a 2 by 2 contingency table (Table 3), in which rows correspond
502 with drug-combination synergy (Y/N), and columns, with target-combination synergy (Y/N).
503 Among synergistic drug combinations, synergy is predicted in 2,594 of their corresponding

target combinations with DSCNi, but not in the other 7,097. Neither is synergy predicted in any of the other non-synergistic drug combinations in iDSCN. The *P*-value of the chi-squared test is 0.00001, and the odds ratio is 1,599. This is strong evidence of the greater likelihood that synergistic drug combinations have synergistic target combinations.

Table 3. Contingency table between drug- and and target-combination synergy

Discussion

Our new DSCN method, double target selection guided by CRISPR screening and network, uses both cancer tissue and cell-line models to discover and rank target combinations, and it has several unique features and advantages in comparison with existing methods of selecting combination targets.

For the first time, DSCN uses a subsampling approach that characterizes the knockdown of the first target and models its impact on all the other genes. To demonstrate the validity of this assumption, we studied a set of transcriptome profiles across 22 pancreatic cell lines before and after MAP2K1 and MAP2K2 inhibition. Among 1,301 neighbor genes of MAP2K1 and MAP2K2 in the PPI network, our analysis revealed high correlation of observed log-fold changes in these genes before and after MAP2K1 and MAP2K2 inhibition with log-fold changes calculated from the sub-sampling approach, $R^2 = 0.75$.

DSCN also differs from all other methods by focusing on the overlapped functional network

between cancer tissues and cell lines and further matching the differential gene expression in the tissue to gene essentialities in the cell line. This framework for the selection of target combinations is highly translational and practical. We investigated a number of scoring schemes for calculating impact score, including the most-probable paths, random-walk paths, and diffusion paths, and we studied whether the global network and spectrum clustering-based local network lead to different calculations of impact score. Using tumor samples of pancreatic cancer and cell-line samples and known synthetic lethal data in SynlethDB, we showed statistically significant lower impact scores of target combinations in synthetic lethal gene pairs than other target pairs utilizing a diffusion-path approach on the local network. This analysis clearly demonstrates the validity of our proposed algorithm for calculating the impact scores of target combinations that reflect synthetic lethalilty.

Furthermore, DSCN is broadly defined for every target and target combination, unlike existing network-based target selection algorithms, such as OptiCon[1] or VIPER [2], that are limited by their initial step in the selection of single targets (i.e., master regulators). This advantage of DSCN is demonstrated in the analysis of overlap among the the top-ranked target pairs between DSCN, Opticon, and VIPER and synthetic lethal target pairs reported in the analysis of pancreatic cancer data in SynlethDB. DSCN identified 79 overlapped synthetic lethal target combinations, whereas OptiCon and VIPER showed zero overlap. In addition, three of these top nine predicted synergistic target combinations in pancreatic cancer show statistically significant association with overall survival in patients with pancreatic cancer, and all three contain the TRFC gene, which encodes a surface receptor for cellular iron intake. Hence, the

548 targeting of upstream genes (e.g., MAP2K2, EGLN2) along with downstream genes (e.g., FTL)
549 might lead to a synergistic effect.
550
551 Finally, we investigate two relevant but different concepts, drug- and target-combination
552 synergy, hypothesizing the greater likelihood of synergistic than non-synergistic drug
553 combinations to target more synergistic target combinations. Using DSCNi, a model derived
554 from DSCN for the prediction of target combinations for individual patients, we showed the
555 truth of our hypothesis using triple-negative breast-cancer tissue and cell-line data. Based on
556 1,031 drug combination screening data in HS578T, a TNBC cell line, and its corresponding
557 14,067 DSCNi-predicted target combination synergy scores, we showed the 1,599-fold higher
558 odds of synergistic than non-synergistic drug combinations to predict synergistic target
559 combinations (P = 0.00001).

560 ## Author's contributions

561 Enze Liu executed the study. Lei Wang, Xue Wu, Yang Huo and Huanmei Wu provided
562 technical support and valuable comments to the study. Lang Li and Lijun Cheng designed
563 the study.

564 ## Code availability

565 The python programming was used to implement and train all models. The training and
566 validation datasets used to create each model are available as part of the experimental dataset
567 released as described in materials. The code required to construct the training and validation

dataset data and also to analyze the experimental data is provided for download (https://github.com/tzcoolman/DSCN).

References

1. Hu, Y., et al., *Optimal control nodes in disease-perturbed networks as targets for combination therapy.* Nature communications, 2019. **10**(1): p. 1-14.
2. Alvarez, M.J., et al., *Functional characterization of somatic mutations in cancer using network-based inference of protein activity.* Nature genetics, 2016. **48**(8): p. 838.
3. Parhi, P., C. Mohanty, and S.K. Sahoo, *Nanotechnology-based combinational drug delivery: an emerging approach for cancer therapy.* Drug discovery today, 2012. **17**(17-18): p. 1044-1052.
4. Al-Lazikani, B., U. Banerji, and P.J.N.b. Workman, *Combinatorial drug therapy for cancer in the post-genomic era.* 2012. **30**(7): p. 679-692.
5. Hammer, S.M., et al., *Treatment for adult HIV infection: 2006 recommendations of the International AIDS Society–USA panel.* 2006. **296**(7): p. 827-843.
6. Stephenson, D., et al., *Charting a path toward combination therapy for Alzheimer's disease.* 2015. **15**(1): p. 107-113.
7. Ryan, C.J., I. Bajrami, and C.J.J.T.i.c. Lord, *Synthetic lethality and cancer–penetrance as the major barrier.* 2018. **4**(10): p. 671-683.
8. Shen, J.P., et al., *Combinatorial CRISPR–Cas9 screens for de novo mapping of genetic interactions.* 2017. **14**(6): p. 573-576.
9. Han, K., et al., *Synergistic drug combinations for cancer identified in a CRISPR screen for pairwise genetic interactions.* 2017. **35**(5): p. 463.
10. Tong, A.H.Y., et al., *Global mapping of the yeast genetic interaction network.* 2004. **303**(5659): p. 808-813.
11. Wong, S.L., et al., *Combining biological networks to predict genetic interactions.* 2004. **101**(44): p. 15682-15687.
12. Benstead-Hume, G., et al., *Predicting synthetic lethal interactions using conserved patterns in protein interaction networks.* 2019. **15**(4): p. e1006888.
13. Pandey, G., et al., *An integrative multi-network and multi-classifier approach to predict genetic interactions.* 2010. **6**(9): p. e1000928.
14. Cho, H., B. Berger, and J.J.C.s. Peng, *Compact integration of multi-network topology for functional analysis of genes.* 2016. **3**(6): p. 540-548. e5.
15. Liany, H., A. Jeyasekharan, and V.J.B. Rajan, *Predicting synthetic lethal interactions using heterogeneous data sources.* 2020. **36**(7): p. 2209-2216.
16. Jerby-Arnon, L., et al., *Predicting cancer-specific vulnerability via data-driven detection of synthetic lethality.* 2014. **158**(5): p. 1199-1209.
17. Sinha, S., et al., *Systematic discovery of mutation-specific synthetic lethals by mining pan-*

cancer human primary tumor data. 2017. **8**(1): p. 1-13.

18. Lee, J.S., et al., *Harnessing synthetic lethality to predict the response to cancer treatment.* 2018. **9**(1): p. 1-12.
19. Tamada, K., et al., *Redirecting gene-modified T cells toward various cancer types using tagged antibodies.* Clin Cancer Res, 2012. **18**(23): p. 6436-45.
20. Liany, H., A. Jeyasekharan, and V.J.b. Rajan, *ASTER: A Method to Predict Clinically Actionable Synthetic Lethal Interactions.* 2020.
21. Liu, E., et al., *SCNrank: spectral clustering for network-based ranking to reveal potential drug targets and its application in pancreatic ductal adenocarcinoma.* 2020. **13**: p. 1-15.
22. Edgar, R., M. Domrachev, and A.E.J.N.a.r. Lash, *Gene Expression Omnibus: NCBI gene expression and hybridization array data repository.* 2002. **30**(1): p. 207-210.
23. Barrett, T., et al., *NCBI GEO: archive for functional genomics data sets—update.* 2012. **41**(D1): p. D991-D995.
24. Tomczak, K., P. Czerwińska, and M.J.C.o. Wiznerowicz, *The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge.* 2015. **19**(1A): p. A68.
25. Tsherniak, A., et al., *Defining a Cancer Dependency Map.* Cell, 2017. **170**(3): p. 564-576 e16.
26. Aguirre, A.J., et al., *Genomic Copy Number Dictates a Gene-Independent Cell Response to CRISPR/Cas9 Targeting.* Cancer Discov, 2016. **6**(8): p. 914-29.
27. Cowley, G.S., et al., *Parallel genome-scale loss of function screens in 216 cancer cell lines for the identification of context-specific genetic dependencies.* Sci Data, 2014. **1**: p. 140035.
28. Szklarczyk, D., et al., *The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible.* Nucleic Acids Res, 2017. **45**(D1): p. D362-D368.
29. Guo, J., H. Liu, and J. Zheng, *SynLethDB: synthetic lethality database toward discovery of selective and sensitive anticancer drug targets.* Nucleic acids research, 2016. **44**(D1): p. D1011-D1017.
30. Zagidullin, B., et al., *DrugComb: an integrative cancer drug combination data portal.* 2019. **47**(W1): p. W43-W51.
31. Wishart, D.S., et al., *DrugBank 5.0: a major update to the DrugBank database for 2018.* Nucleic Acids Res, 2018. **46**(D1): p. D1074-D1082.
32. Borisy, A.A., et al., *Systematic discovery of multicomponent therapeutics.* Proceedings of the National Academy of Sciences, 2003. **100**(13): p. 7977-7982.
33. O'Neil, J., et al., *An unbiased oncology compound screen to identify novel combination strategies.* Molecular cancer therapeutics, 2016. **15**(6): p. 1155-1162.
34. Margolin, A.A., et al. *ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context*. in *BMC bioinformatics*. 2006. Springer.
35. Tang, Z., et al., *GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses.* 2017. **45**(W1): p. W98-W102.
36. Jeong, S.M., S. Hwang, and R.H. Seong, *Transferrin receptor regulates pancreatic cancer growth by modulating mitochondrial respiration and ROS generation.* Biochemical and biophysical research communications, 2016. **471**(3): p. 373-379.
37. Xie, Y., et al., *Ferroptosis: process and function.* Cell Death & Differentiation, 2016. **23**(3): p. 369-379.
38. Shoemaker, R.H.J.N.R.C., *The NCI60 human tumour cell line anticancer drug screen.* 2006.

650 **6**(10): p. 813-823.
651 39. Barretina, J., et al., *The Cancer Cell Line Encyclopedia enables predictive modelling of*
652 *anticancer drug sensitivity.* 2012. **483**(7391): p. 603-607.
653
654

TABLE LEGENDS

Table 1. Datasets used in this study

Table 2. Analysis of overall survival among the nine top-ranked target combinations in pancreatic ductal adenocarcinoma (PDAC)

Table 3. Contingency table between drug- and and target-combination synergy

FIGURES LEGENDS

Figure 1. Overview of double-target selection guided by CRISPR screening and network (DSCN).

Figure 2. Network configurations for three methods to calculate impact score.

Figure 3. Correlation between the predicted and observed log-fold changes in gene expression among MAP2K1 and MAP2K2 neighbor genes in the protein-protein interaction (PPI) network.

Figure 4. Comparison of target-combination impact scores using synthetic versus non-synthetic lethal gene pairs in pancreatic cancer.

Figure 5. Kaplan-Meier curves for the nine top-ranked target combinations (a)-(i).

Table 1. Datasets used in this study.

Part 1. Multi-omics data			
Cancer type	**Data platform**	**Data type**	**Data (n, sample size)**
Pancreatic cancer cell lines	Affymetrix U133 2.0	Gene expression	GSE36133 (43), GSE46385 (7), GSE21654 (22), GSE17891 (20) Total sample size = 92
	CRISPR screening	Gene essentiality	Project Achilles (v3.3.8) Total sample size = 26
Pancreatic tissue samples	Affymetrix U133 2.0	Gene expression (tumor)	GSE42952 (33), GSE51978 (2), GSE16515 (36), GSE15471 (39), GSE23952 (3) Total sample size = 113
	Affymetrix U133 2.0	Gene expression (normal)	GSE46385 (3), GSE16515 (16), GSE15471 (39) Total sample size = 58
	Illumina DNA-seq & RNA-seq	Mutation and gene expression (tumor)	TCGA ductal and lobular neoplasms (150), adenomas and adenocarcinomas (29)
	Illumina RNA-seq	Gene expression (normal)	Solid tissue adjacent normal (41)
Breast cancer tissue samples	RNA-seq	Gene expression (tumor)	TCGA triple negative breast cancer sample (115)
		Gene expression (normal)	TCGA triple negative breast cancer sample (163)
Breast cancer cell lines	Affymetrix U133 2.0	Gene expression	GSE36133 (12)
	CRISPR Screening	Gene essentiality	Project Achilles (v3.3.8) Total sample size = 28
Part 2. Databases			
Data type	**Database**	**Data**	
Protein-protein interaction (PPI) network	STRING[28]	PPI data in STRING database for human (v11): 11,609,230 interactions	
Drug targets	DrugBank[31]	Food and Drug Administration (FDA)-approved drugs and their associated target proteins: 1,769 gene targets	
Synthetic lethal pairs	SynlethDB[29]	19,613 synthetic lethal gene pairs in human cancer	
Drug sensitivity data	DrugComb[30]	Drug synergies among cell lines on 5,226 drug pairs (HS5	

672

673

Table 2. Analysis of overall survival among the nine top-ranked target combinations in pancreatic ductal adenocarcinoma (PDAC).

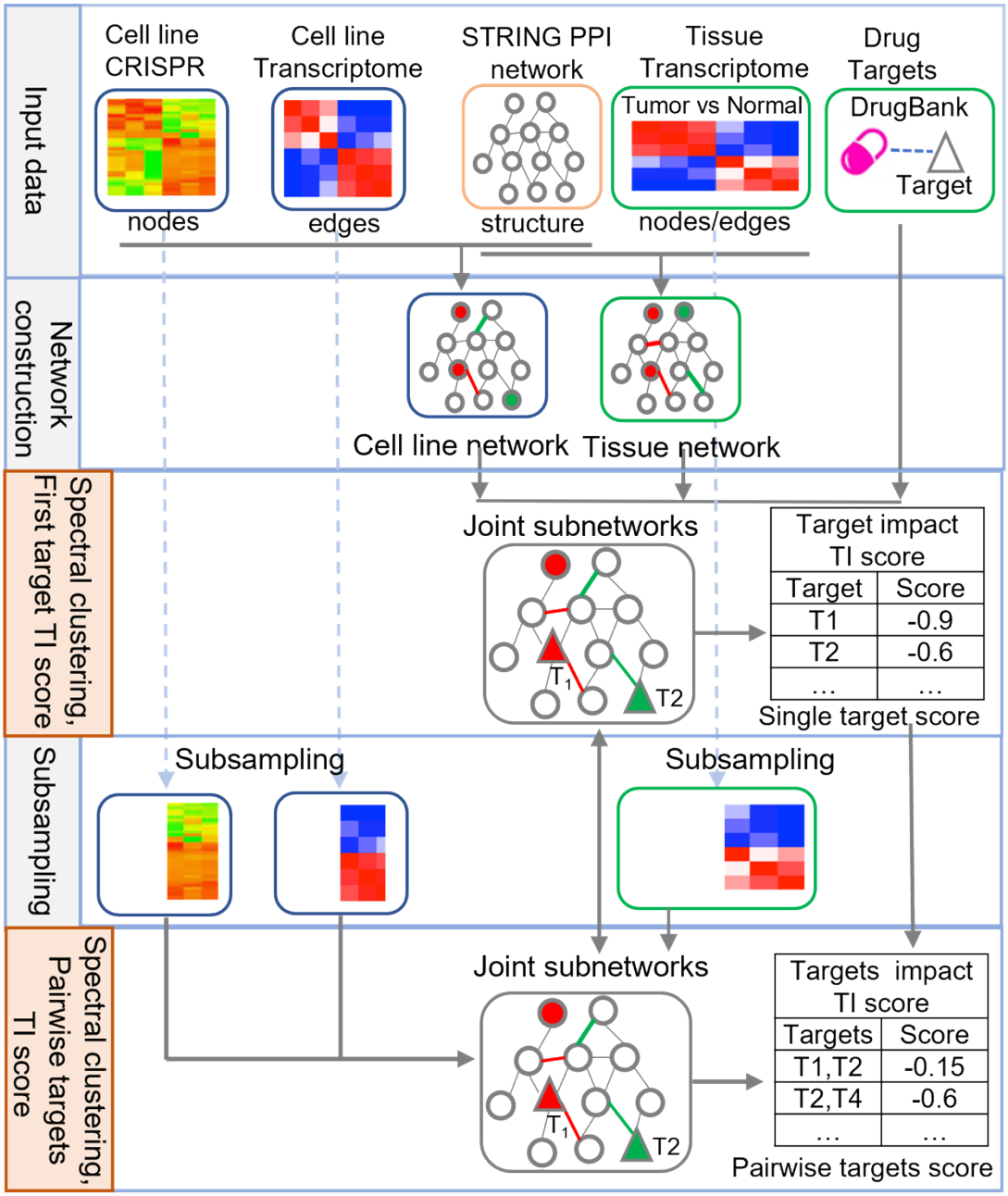
Gene 1	Gene 2	Impact score	Log rank *P*-value	Hazard Ratio (HR)	HR *P*-value	Pathways
EGLN1	TFRC	-255.12	0.02	2.00	0.02	hypoxia, ferroptosis
MAP2K2	TFRC	-255.05	0.08	1.60	0.08	MAPK, ferroptosis
HPSE	TFRC	-255.01	0.19	1.50	0.20	Metabolism, ferroptosis
PPIC	TFRC	-254.86	0.06	1.80	0.06	Immune system, ferroptosis
FRK	TFRC	-254.86	0.04	1.80	0.05	Immune system, ferroptosis
EGLN1	COX7C	-254.79	0.84	1.10	0.85	Hypoxia, metabolism
XDH	TFRC	-254.75	0.001	2.40	0.002	Metabolism, ferroptosis
MAP2K2	COX7C	-254.72	0.14	0.65	0.15	MAPK, oxidative phosphorylation
FTL	TFRC	-254.71	0.10	1.60	0.10	ferroptosis, ferroptosis

Table 3. Contingency table between drug- and and target-combination synergy.

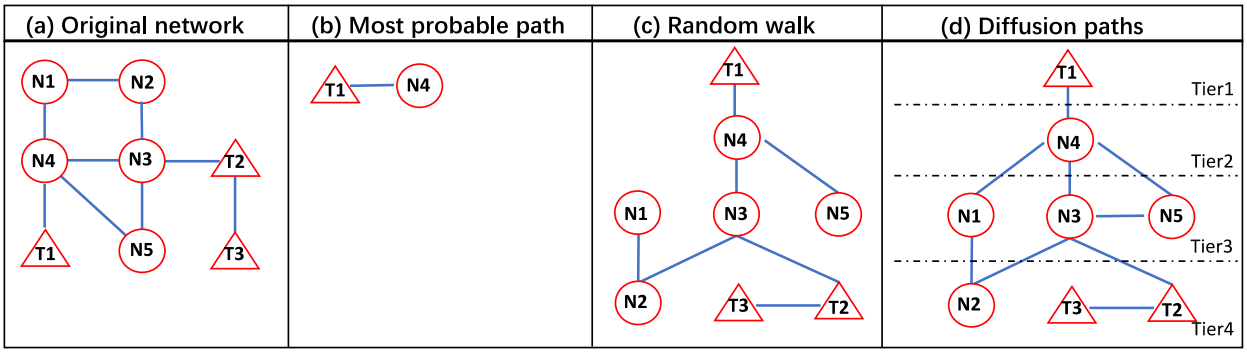
	Predicted target-combination synergy	Predicted target-combination non-synergy
Drug-combination synergy	2,594	7,097
Drug-combination non-synergy	0	4,375

677

678

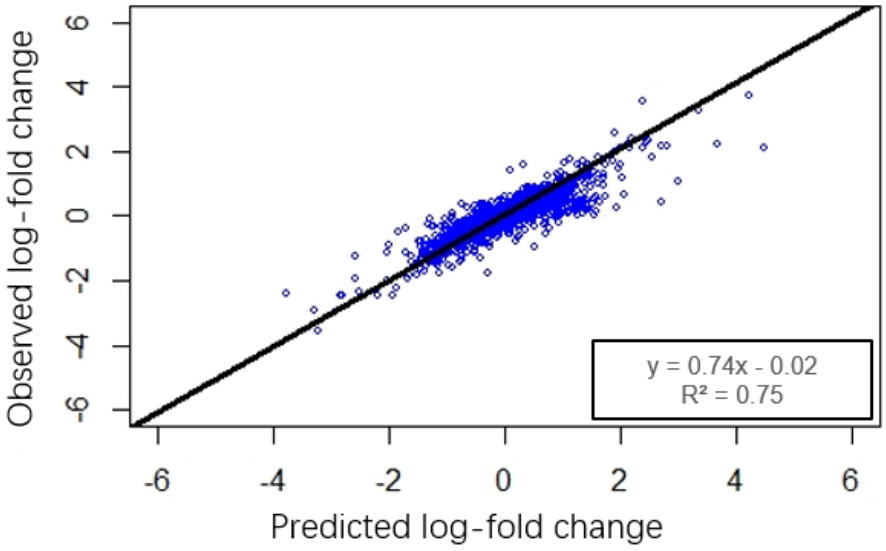


679 **Figure 1.** Overview of double-target selection guided by CRISPR screening and network
680 (DSCN).
681
682



Note: (a) Original network. (b) Most-probable path for $T1$. (c) Random-walk path from $T1$. (d) Diffusion path in four hierarchical tiers.

Figure 2. Network configurations for three methods to calculate impact score.



686

687 **Figure 3.** Correlation between the predicted and observed log-fold changes in gene
688 expression among MAP2K1 and MAP2K2 neighbor genes in the protein-protein interaction
689 (PPI) network.

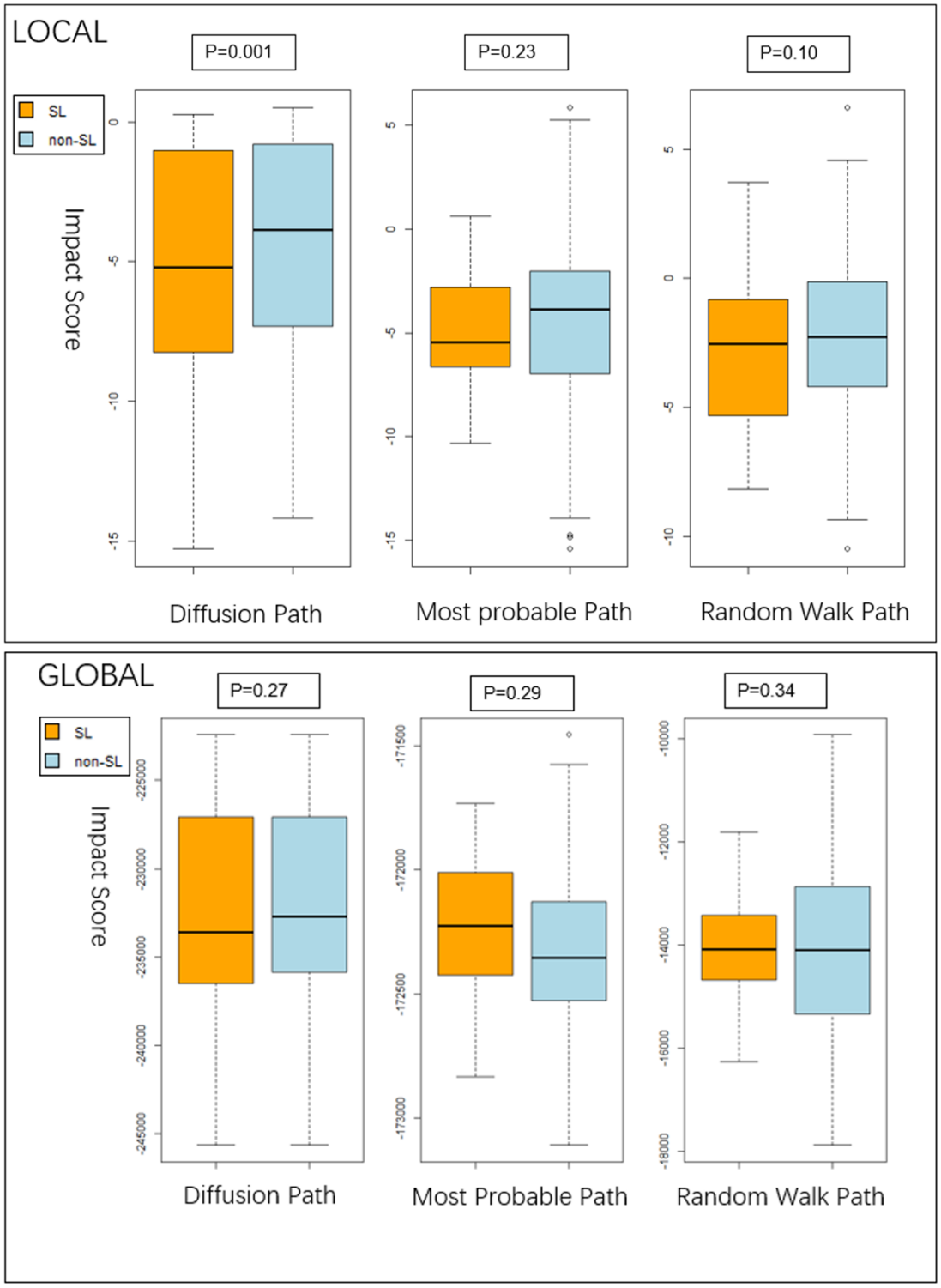


Figure 4. Comparison of target-combination impact scores using synthetic versus non-synthetic lethal gene pairs in pancreatic cancer. The three methods for calculating impact score–the most-probable, random-walk, and diffusion paths are defined in **Figure 2**. The impact scores (IS) are calculated from either the global protein-protein interaction (PPI) network (global) or the local PPI network (local).

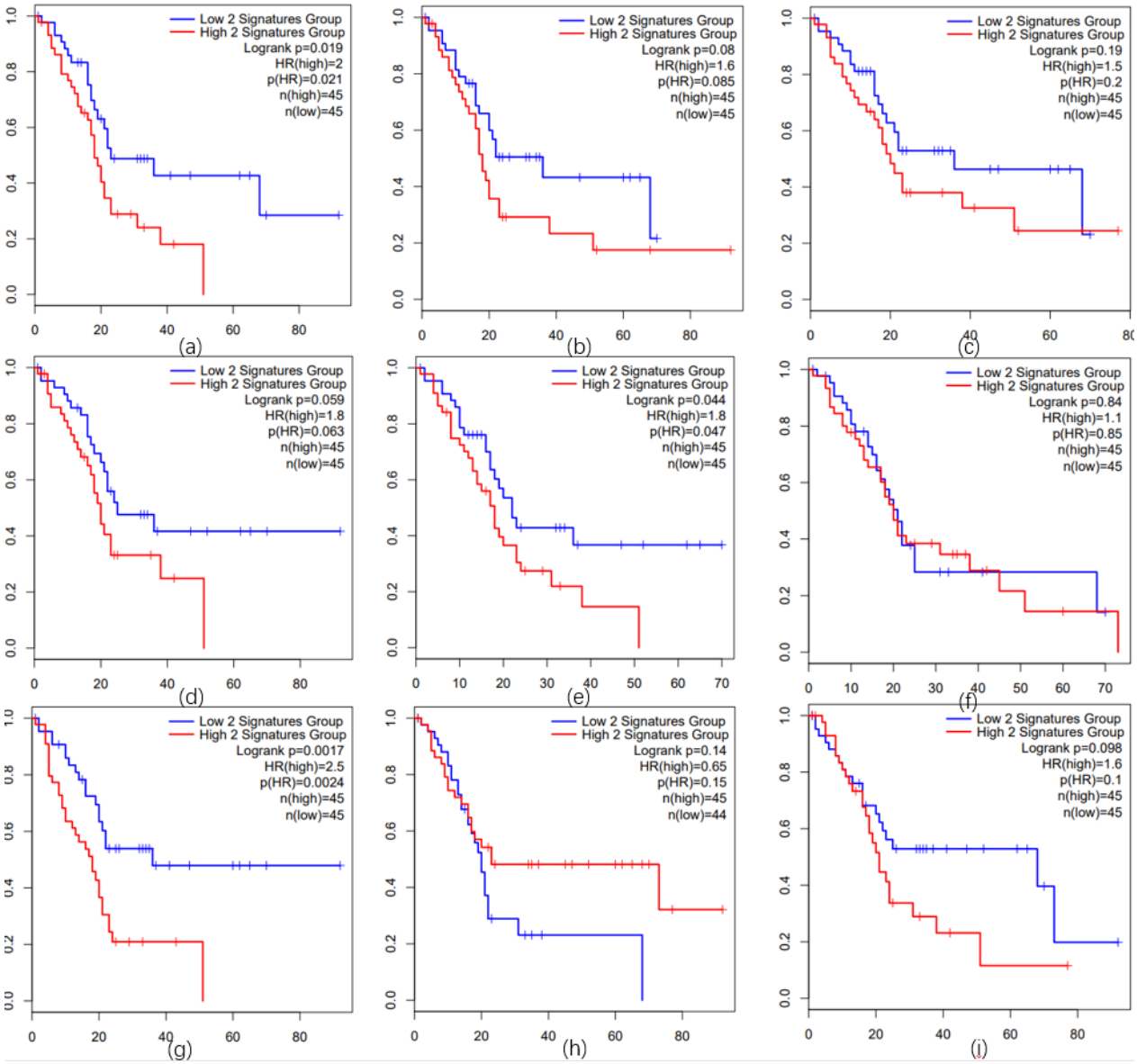


Figure 5. Kaplan-Meier curves for the nine top-ranked target combinations (a)-(i). Kaplan-Meier curves and other survival statistics for (a) < EGLN1, TRFC>, (b) < MAP2K2, TRFC>, (c) < HPSE, TRFC>, (d) < PPIC, TRFC>, (e) < FRK, TRFC>, (f) < EGLN1, COX7C>, (g) < XDH, TRFC>, (h) < MAP2K2, COX7C>, and (i) < FTL, TRFC>. Y-axis indicates survival probability while X-axis indicates months. The blue line in each plot indicates low expression of the two gene groups, and the red line, high expression.