Diffusion mapping of drug targets on disease signaling network elements reveals drug combination strategies

Jielin Xu[†], Kelly Regan-Fendt[†], Siyuan Deng

Department of Biomedical Informatics, The Ohio State University, Columbus, OH 43210, U.S.A.
Email: Jielin.Xu@osumc.edu, Kelly.Regan@osumc.edu, Siyuan.Deng@osumc.edu

William E. Carson III

Comprehensive Cancer Center, The Ohio State University, Columbus, OH 43210, U.S.A.
Email: William.Carson@osumc.edu

Philip R.O. Payne

Institute for Informatics, Washington University in St. Louis, St. Louis, MO 63110, U.S.A.
Email: prpayne@wustl.edu

Fuhai Li

Department of Biomedical Informatics, Translational Data Analytics Institute, The Ohio State University, Columbus, OH 43210, U.S.A. Email: Fuhai.Li@osumc.edu

The emergence of drug resistance to traditional chemotherapy and newer targeted therapies in cancer patients is a major clinical challenge. Reactivation of the same or compensatory signaling pathways is a common class of drug resistance mechanisms. Employing drug combinations that inhibit multiple modules of reactivated signaling pathways is a promising strategy to overcome and prevent the onset of drug resistance. However, with thousands of available FDA-approved and investigational compounds, it is infeasible to experimentally screen millions of possible drug combinations with limited resources. Therefore, computational approaches are needed to constrain the search space and prioritize synergistic drug combinations for preclinical studies. In this study, we propose a novel approach for predicting drug combinations through investigating potential effects of drug targets on disease signaling network. We first construct a disease signaling network by integrating gene expression data with disease-associated driver genes. Individual drugs that can partially perturb the disease signaling network are then selected based on a drug-disease network "impact matrix", which is calculated using network diffusion distance from drug targets to signaling network elements. The selected drugs are subsequently clustered into communities (sub-groups), which are proposed to share similar mechanisms of action. Finally, drug combinations are ranked according to maximal impact on signaling sub-networks from distinct mechanism-based communities. Our method is advantageous compared to other approaches in that it does not require large amounts drug dose response data, drug-induced "omics" profiles or clinical efficacy data, which are not often readily available. We validate our approach using a BRAF-mutant melanoma signaling network and combinatorial in vitro drug screening data, and report drug combinations with diverse mechanisms of action and opportunities for drug repositioning.

Keywords: Drug repositioning; drug combination; signaling network; network diffusion

1. Introduction

Despite the discovery of many disease-causing molecular aberrations, the vast majority are not successfully targeted by approved drugs. Furthermore, widespread drug resistance to targeted therapies is still a major challenge in cancer treatment [1]. Thus, the design of multi-target agents and rationale drug combinations seeks to address some of these issues and accomplish specific objectives: increased overall efficacy, improved initiation for first-line therapies, reduced drug resistance, reduced required doses and reduced drug toxicities. However, the high costs and low success rates of high-throughput drug screening are exponentially prohibitive to screen drug combinations across different cellular contexts and doses [2]. Therefore, computational methods have the potential to focus research efforts on optimal drug combination in the preclinical testing setting, and eventually to aid in clinical decision-making [3, 4].

Malignant melanoma represents an important use case for precision medicine research and systematizing the design of rationale combination therapies, including recent developments in targeted and immune-based therapies. Melanoma tumors are primarily driven by two oncogenes, *BRAF* (~50%) and *NRAS* (~25%), that converge on the MAPK signaling pathway to promote growth, survival and evade apoptosis. Currently approved targeted therapies for patients with *BRAF*-mutant melanoma include first-line treatment with BRAF inhibitors vemurafenib and dabrafenib, which have improved survival by 6-12 months [5]. However, most patients eventually become resistant to BRAF inhibitor therapies, and heterogenous resistance mechanisms have been observed [6]. For instance, while mutations in *BRAF* and *NRAS* genes are observed to be mutually exclusive across primary and metastatic tumors, acquired *NRAS* mutations have been described as a mechanism of BRAF-inhibitor resistance [7]. Reactivation of the MAPK signaling pathway can also occur via MEK over-activation, and the first combination of targeted therapy including BRAF and MEK inhibition was recently approved for patients with *BRAF*-mutant melanoma. While this combination extends patient survival an additional 5-10 months, additional non-MAPK pathway resistance mechanisms arise, and new, more durable drug combination regimens are needed [8].

In our previous work, we described a novel computational method, SynGeNet [9], which i) integrates transcriptomics and protein-protein interaction data into a comprehensive disease signaling network to map signal flow from an initial set of disease-related "root" genes; and ii) determines drug combinations that maximally reverse disease-associated gene expression signals and targets topologically important nodes in the overall network. We showed that SynGeNet outperformed two other transcriptomics-based drug combination methods in predicting drug combinations validated *in vitro*, and that it could recapitulate genotype-specific results across diverse melanoma cell lines. Importantly, we observed that the network-mining step, which utilized an average of centrality metrics for drug target pathways, was the most crucial aspect of our method in validating drug combination efficacy. Due to the importance of modeling drug-target network connections, we sought to evaluate additional methods that exploit drug-target and disease signaling network structure. Additionally, we sought to overcome the limitation of requiring drug-

induced gene expression profiles as part of the original SynGeNet method. Compared with other unsupervised approaches based on correlating gene expression profiles alone, network-based approaches can more explicitly indicate possible mechanism of action in terms of inhibited signaling targets, and consequently specify a measure for predicting efficacy.

In this study, we propose a novel approach to prioritize drug combinations that can potentially impact a *BRAF*-mutant melanoma signaling network that is constructed from the integration of gene expression and protein-protein interaction data. We employed a random walk with restart (RWR) model to traverse a *BRAF*-mutant melanoma disease signaling network to derive a drug-disease "impact matrix". We then selected drugs that can maximally perturb the disease signaling network for subsequent drug combination modeling. Additionally, we hypothesized that drugs with different mechanisms of action or targets on distinct network modules may have a higher potential for synergy according to the independent mechanism theory for drug combinations. Therefore, we divided prioritized drugs into communities (sub-groups) based on drug target similarity matrices and subsequently ranked drug combinations representing different drug communities. To our knowledge, this is the first study to apply this paradigm to evaluate drug combination hypotheses. Furthermore, we apply the RWR method to determine drug mechanisms by delineating shortest local paths within the network.

2. Methods

An overview figure of the RWR approach to predict and validate drug combinations is shown in **Supplemental Figure S1**. All supplemental materials can be found at the following URL: https://www.kaggle.com/osubmi/diffusion-mapping-for-drug-combinations. All code is made available upon request.

2.1. *Melanoma disease signaling network construction.* In our previous work, we defined a melanoma disease signaling network integrating gene expression data from a publicly available dataset of melanoma patient tumors harboring driver $BRAF^{V600E/K}$ mutations (GSE15605) with protein-protein interaction data from the BioGRID database [10]. Briefly, the network was constructed using the belief propagation approach to map signal flow from a set of frequently mutated "root" melanoma disease genes (n=30) from the DisGeNET database [11, 12]. We hypothesized that an estimated driver disease network could be constructed by minimizing the cost function that weighs the trade-off of including highly activated genes via gene expression fold-changes against including experimentally validated protein-protein interactions with decreasing confidence. This resulted in a disease signaling network of 131 genes to be integrated with drug target information. The mathematical function is reproduced here: Given the BioGRID background network, G = (V, E), where V and E represent the vertices and edges in the BioGRID network, the sub-network, G' = (V', E'), is constructed to minimize the objective function:

$$\min_{E' \subseteq E, V' \subseteq V} \sum_{e \in E'} c_e - \lambda \sum_{i \in V'} b_i \tag{1}$$

where Ce (cost of edge) is set to 0.2, b_i is the patient tumor gene expression fold change representing "activated" state of signaling components (i.e. positive fold change), and λ (set as 0.02) regulates the size of the sub-network. Detailed information regarding disease signaling network construction and empirical rationale for parameter selection can be found in our previous work [13].

2.2. Drug target-disease signaling impact matrix construction using Random Walk with Restart Model (RWR). To estimate the potential impact of inhibiting a drug target on the disease signaling network, the random walk with restart model (RWR) was employed. RWR describes a stochastic process of network signaling flow as follows: at each iteration step, the network signal beginning at an individual gene travels randomly, with equal probability, to a neighboring gene or remains in its current location. In this application, the RWR model is initiated for a gene representing a known direct drug target. The updated location of the network signal can be viewed as a probability expectation, which is defined mathematically in Eq. (2).

$$\vec{r_i} = (1 - c)W\vec{r_i} + c\vec{e_i} \tag{2}$$

Here, $\vec{r_i}$ is a probability vector with elements r_{ij} that denotes the probability of signal flow at gene i travels to gene j, and the sum of all r_{ij} with respect to j should equal to 1. In our method, we set r_{ij} as the impact of gene i to gene j. Here c denotes the restart probability, and W is the normalized adjacency matrix, which is constructed on edge connections (protein-protein interactions) with respect to the BioGRID network. $\vec{e_i}$ represents the starting vector, and has all elements equal to 0 except the i-th element, which is set equal to 1. By solving (2) iteratively, we can extract the target-disease impact matrix $M \in \mathbb{R}^{|V_D| \times |V_T|}$, with M_{ij} denoting the impact of j-th drug target on the i-th gene on the disease signaling network.

2.3. Single drug scoring model. The single drug scoring model is constructed in the following sequential steps. First, with target-disease impact matrix M_{ij} with $M \in \mathbb{R}^{|V_D| \times |V_T|}$, the drug-disease matrix is defined by summing all related target-disease impacts, i.e., we define drug-disease matrix $\widetilde{M} \in \mathbb{R}^{|V_D| \times |DRUG|}$, as:

$$\widetilde{M}_{ij} = \sum_{s \in \{V_T[DRUG_j]\}} M_{is} \tag{3}$$

In other words, Eq. (3) gives us the overall impact of a set of drugs across disease genes within the network. Second, considering the relative "influence" (i.e. number of connections) of disease genes with respect to the topological structure of disease network, we weight drug-disease impact by the degree of disease genes. Thus, the larger the disease gene degree, the higher magnitude the drug impact is amplified. Mathematically speaking, we define a weighted drug-disease impact matrix $\widehat{M} \in \mathbb{R}^{|V_D| \times |DRUG|}$ as follows:

$$\widehat{M}_{ij} = \widetilde{M}_{ij} \cdot \overrightarrow{DEG}[i] \tag{4}$$

with $\overrightarrow{DEG}[i] = deg(V_D^i, N_D)$, where $1 \le i \le |V_D|$. Finally, a score for an individual drug is defined as its average impact on all genes in disease network.

$$DS[i] = \frac{\sum_{s=1}^{|V_D|} \widehat{M}_{si}}{|V_D|} \qquad (5)$$

with $\overrightarrow{DS} \in \mathbb{R}^{|DRUG| \times 1}$. The computed drug scores are ranked in decreasing order.

2.4. *Drug combination scoring model.* The first assumption embedded in the drug combination model is independent mechanism theory, which states that drugs with different mechanisms of action or targets on different disease signaling modules have a higher potential for synergy. Based on this assumption, we first divided drugs into communities based on target similarity, and then estimate the drug combination synergy based on their impact disease signaling elements. The second assumption for the drug combination score model is that isolated disease genes do not contribute to the disease signaling, and that a disease gene can only influence the disease network if it is connected to other nodes inside the disease network along "important" paths. Relative importance for network paths is described in subsection 2.4.2.

2.4.1. Drug community clustering. We clustered drugs that passed criteria described in the single drug score model into different functional groups via the affinity propagation (AP) clustering algorithm [14]. Drug-target interaction information was extracted from the DrugBank database, and the resulting Jaccard index coefficient was used as the basis for affinity propagation clustering. Drug-drug similarity was evaluated by Jaccard index:

$$S_{ij} = \frac{\left| V_T[SEL\_DRUG_i] \cap V_T[SEL\_DRUG_j] \right|}{\left| V_T[SEL\_DRUG_i] \cup V_T[SEL\_DRUG_j] \right|} \qquad (6)$$

with S_{ij} denotes the similarity between selected drug i and selected drug j, and $S = \{S_{ij}\} \in \mathbb{R}^{|SEL\_DRUG| \times |SEL\_DRUG|}$.

2.4.2. Drug combination prediction. We select drug pairs from different communities as candidate combinations. We then determine the disease genes that are highly impacted by different drug combinations. The disease genes are scored by truncating impact above a certain threshold (T) as described below:

$$T = alpha \, * \max_{i,j} M \qquad (7)$$

with $1 \leq i \leq |V_D|$ and $1 \leq j \leq |V_T|$. An *alpha* value is determined as follows: We define a parameter p that represents the percentage of disease genes exclusively impacted by drugs involved in a given drug combination. The optimal alpha should provide the highest 'p' as defined below. In other words, the optimal alpha allows both drugs within a combination to provide as much unique information as possible. In this way, we prioritize drug pairs that influence non-redundant disease sub-networks.

$$p = \frac{|D_1| + |D_2|}{|D_1| + |D_2| + |D_c|} \qquad (8)$$

With D_1, D_2, D_c represent the number of exclusive disease genes impacted by drug 1, drug 2, and both drugs, respectively. Based on the optimal alpha selected, for a given drug combination $(DRUG_i, DRUG_j)$, we can extract highly impacted disease genes for both

drugs, i.e., IDG_i and IDG_j, with $IDG_i = \bigcup_{k=1}^{|IDG_i|} V_D^k$ and $IDG_j = \bigcup_{l=1}^{|IDG_j|} V_D^l$. The drug combination score $CS(DRUG_i, DRUG_j)$ is then defined as follows:

$$CS(DRUG_i, DRUG_j) = \begin{cases} 0, & |IDG_i| = 1, |IDG_j| = 1 \\ PS_{1D}(DRUG_i), & |IDG_i| > 1, |IDG_j| = 1 \\ PS_{1D}(DRUG_j), & |IDG_i| = 1, |IDG_j| > 1 \\ PS_{2D}(DRUG_i, DRUG_j), & |IDG_i| > 1, |IDG_j| > 1 \end{cases} \tag{9}$$

with $PS_{1D}(DRUG_k)$ and $PS_{2D}(DRUG_k, DRUG_l)$ defined below respectively. Here, $|IDG_k| = 1$ refers to $DRUG_k$ impacts only one isolated gene in the entire disease network, and therefore, it does not contribute to the drug combination score. Using the RWR model, we then evaluate $DRUG_k$ impact to all Dijkstra's shortest paths that connects two arbitrary disease genes V_D^m and V_D^n in set IDG_k for $DRUG_k$, and then rank them in decreasing order as I_{DRUG_k}, then we have:

$$PS_{1D}(DRUG_k) = \sum_{m=1}^{|PATH_{NT}|[k]} I_{DRUG_k}[m] \tag{10}$$

with

$$|PATH_{NT}|[k] = \begin{cases} |I_{DRUG_k}|, & |I_{DRUG_k}| < 3 \\ 3, & |I_{DRUG_k}| \geq 3 \end{cases} \tag{11}$$

Here, $|I_{DRUG_k}|$ denotes the number of all shortest paths that connects all non-isolated disease genes that are highly impacted by the kth drug, and $|PATH_{NT}|[k]$ denotes the truncated numbers of all shortest paths, which are used for drug combination score evaluation. For construction of $PS_{2D}(DRUG_k, DRUG_l)$, considering a certain path score threshold T_{PS} which is chosen as the magnitude separator, i.e., local path scores above T_{PS} stay within the highest magnitude, and local path scores below T_{PS} are with lower magnitudes, then we have:

$$PS_{2D}(DRUG_k, DRUG_l) = \begin{cases} \sum_{k,l} PS_{1D}(DRUG_m), & |PATH_T|[k] + |PATH_T|[l] = 0 \\ PS_{1D}(DRUG_k), & |PATH_T|[k] > 0, |PATH_T|[l] = 0 \\ PS_{1D}(DRUG_l), & |PATH_T|[k] = 0, |PATH_T|[l] > 0 \\ \sum_{k,l} PS_{1D}(DRUG_m), & |PATH_T|[k] * |PATH_T|[l] > 0 \end{cases} \tag{12}$$
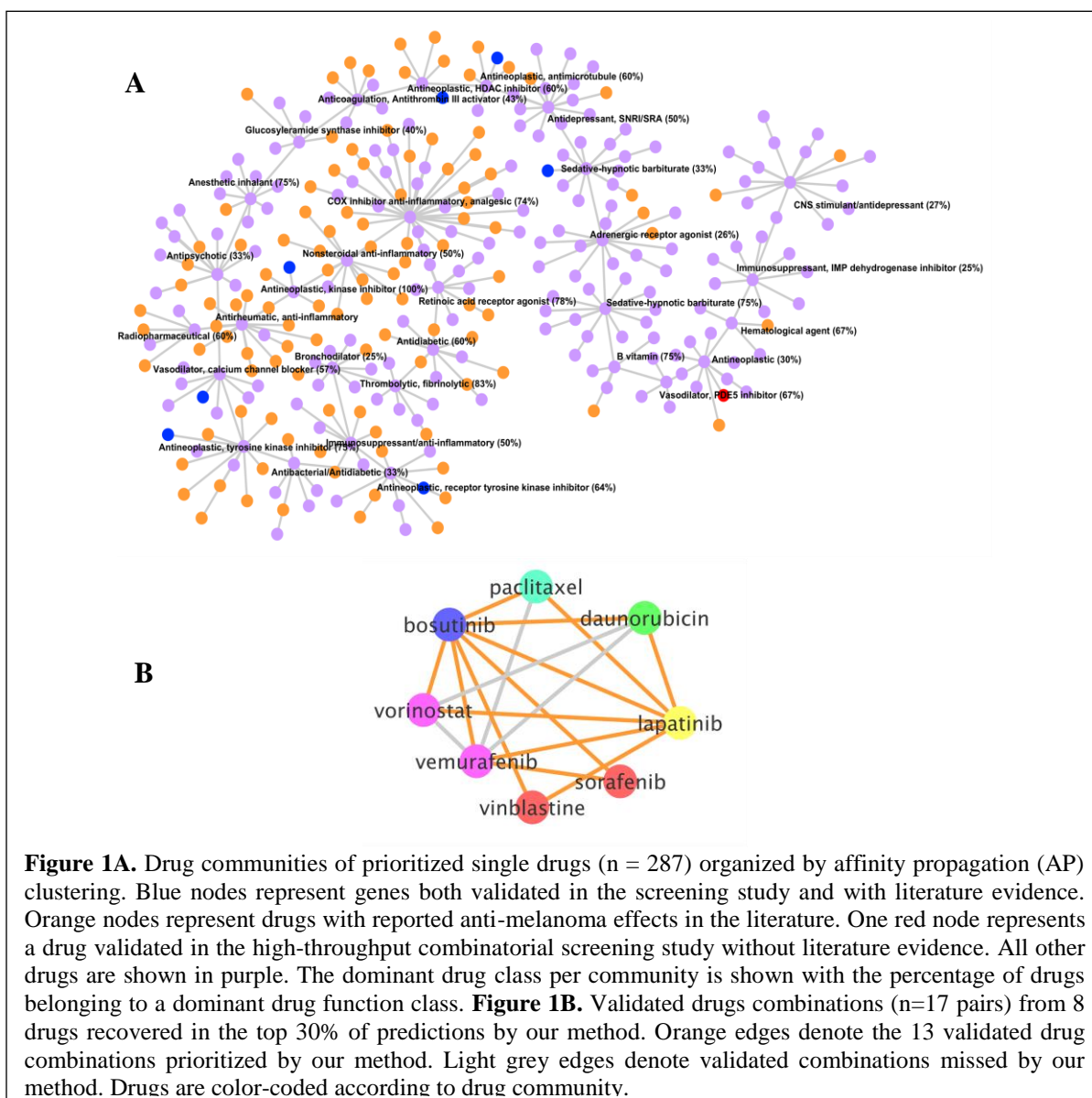
and

$$|PATH_T|[k] = \begin{cases} |I_{DRUG_k} > T_{PS}|, & |I_{DRUG_k} > T_{PS}| < 3 \\ 3, & |I_{DRUG_k} > T_{PS}| \geq 3 \end{cases} \tag{13}$$

Here, $\left|I_{DRUG_k} > T_{PS}\right|$ denotes the number of all shortest paths above T_{PS} for the kth drug, and $|PATH_T|[k]$ denotes the truncated numbers of shortest paths with highest magnitude which are used for drug combination score evaluation.

3. Results

3.1. *Evaluation of single drug predictions.* To evaluate the proposed approach, we first obtained all the FDA approved drugs with target information available from the DrugBank [15], which resulted in 1,433 drugs selected for this study. Based on the single drug score model, we selected the top 20% ranked single drugs (287 drugs) predicted by our method targeting a *BRAF*-mutant melanoma disease signaling network constructed via integration of gene expression and protein-protein interaction data (described in **Methods**). We first sought to distinguish single drug predictions on a mechanistic basis by applying affinity

Figure 1A. Drug communities of prioritized single drugs (n = 287) organized by affinity propagation (AP) clustering. Blue nodes represent genes both validated in the screening study and with literature evidence. Orange nodes represent drugs with reported anti-melanoma effects in the literature. One red node represents a drug validated in the high-throughput combinatorial screening study without literature evidence. All other drugs are shown in purple. The dominant drug class per community is shown with the percentage of drugs belonging to a dominant drug function class. **Figure 1B.** Validated drugs combinations (n=17 pairs) from 8 drugs recovered in the top 30% of predictions by our method. Orange edges denote the 13 validated drug combinations prioritized by our method. Light grey edges denote validated combinations missed by our method. Drugs are color-coded according to drug community.

propagation clustering on the Jaccard Index of the drug-drug similarity matrix to define drug community clusters. This resulted in 30 non-overlapping drug communities, where each drug was assigned to a single community (**Figure 1A**). The average size of a drug community was 9.6 drugs, with an average of 54.8% drugs belonging to a predominant drug function class within a community (**Figure 1A**). While anti-neoplastics was the most frequent drug function category observed (6/30 communities), other common drug functions included anti-inflammatory, antidiabetic, vasodilation and anti-depressant drugs, indicating a high potential for drug repositioning. To evaluate these selected drugs, we first manually searched the literature for associations between these drugs and melanoma. We found literature associations for 112/287 drugs with melanoma having 10 or more citations in PubMed (**Supplemental Table S1**). Additionally, in order to evaluate our predictions in the *BRAF*-mutant melanoma context, we used the reported GI50 (50% growth inhibition) values from a previous drug screen testing 40 agents in combinations *BRAF*-mutant melanoma cell lines [16]. Drugs were first screened individually across melanoma cell lines (**validation dataset 1**). Of these 40 tested drugs, 13 overlapped with our initial set of FDA-approved therapies, and of these 13 drugs, we note that 8 were ranked in the top 20% of

3.2. *Evaluation of drug combination predictions.* We hypothesized that drugs paired from different drug communities would have non-redundant functions to inhibit the overall disease signaling network, and thus could represent efficacious drug combinations. From the aforementioned combinatorial drug screening study, a total of 650 drug combinations (with different doses) were considered validated by the previous criteria defined by the authors (**validation dataset 2**): both ≥ 50% growth inhibition in *BRAF*-mutant melanoma cell lines and ≥ 15% growth inhibition observed specifically in *BRAF*-mutant cell lines vs. other genetic backgrounds. Among the total 650 validated drug combinations, there were 28 drug combinations overlapping between the original subset of FDA-approved drugs. Considering the top 20% of predicted drugs, our method recovered 17 out of 28 validated drug combinations (**Figure 1B**). Interestingly and consistent with our hypothesis, 16 out of 17 of the validated drug combinations included both drugs from distinct communities in our clustering analysis. Only one validated drug combination contained two drugs from the same community and was discarded by our method for this reason. Furthermore, the drug combination score model predicted 13 out of 17 drug combinations that ranked in the top 30% of all qualified drug combinations (**orange** edges in **Figure 1B**), and the first 10 drug combinations ranked by GI50 score were all prioritized by our approach.

3.3. *Drug combination mechanism discovery.* We further sought to extend the RWR network diffusion model to test the independent drug mechanism hypothesis for drug combinations from a signaling pathway perspective by determining local signaling paths within the network. First, we aimed to define an optimum threshold that provides the most exclusive information between two drugs, and tested several alpha values that control the percentage of disease genes exclusively impacted by drugs paired in combination (alpha = 0.0001, 0.0003, 0.0005, 0.0007, 0.001) for all validated drug combinations as plotted in **Supplemental Figure S2**. According to our observations, we empirically selected 0.0005 as the optimal alpha.

To demonstrate the independent mechanism hypothesis in a clinically relevant context, we first considered the two FDA approved first-line targeted therapies for BRAF-mutant melanoma, including a BRAF and MEK inhibitor, respectively: vemurafenib + cobimetinib, and dabrafenib + trametinib. It is worth noting that all four single drugs are selected among the top 20% by our method. For the sake of simplicity, only the local path plot for vemurafenib + cobimetinib is presented (**Figure 2A**). As is shown in **Figure 2A,** the two shortest local paths for this drug combination originate from the *BRAF* and *MAP2K1* genes (targets of vemurafneib and cobimetinib, respectively), and they do not intersect, thus fulfilling the independent mechanism hypothesis. Among the 17 selected validated drug combinations (drug edges in Figure 1B) involving the 8 prioritized single drugs (drug nodes in Figure 1B), we note that 5 of these drug combinations (**Table 1**, right) shared common genes with impact scores within the same order of magnitude, thus permitting testing of the independent mechanism hypothesis via RWR of shortest paths.

Of these five combinations, the top four drug combinations all exhibited independent local paths derived from each drug. For the sake of simplicity, we only show the local path plot for the bosutinib + sorafenib combination in **Figure 2B** to compare to the clinically relevant example of vemurafenib + cobimetinib. As seen with vemurafenib + cobimetinib, the bosutinib + sorafenib combination also demonstrates local paths connecting to *BRAF* and *MAP2K1* (MEK1) genes through independent mechanisms. Another interesting observation for the bosutinib + sorafenib combination is that of the 20 total disease genes that are highly impacted by both drugs (union of shared and unique genes), only 6 genes (*BRAF*, *APP*, *FLT1*, *CDK2*, *SRC* and *MAP2K1*) are connected through local paths that reveal mechanisms by which both drugs impact the *BRAF*-mutant melanoma disease signaling network. Finally, we note that although both drugs share four highly impacted disease genes *BRAF*, *BSG*, *IPO13*, and *MC1R*, the only connected gene in a local path of these is *BRAF*, suggesting the major role for *BRAF* signaling in the network. Furthermore, we note that this local path for *BRAF* is connected to genes highly impacted by sorafenib, which unlike bosutinib, directly targets the BRAF protein.

Table 1. (Left): Highest truncated local path scores of prioritized single drugs (see drug nodes in Figure 2B) and BRAF/MEK inhibitors. We note that several drugs, e.g., daunorubicin and vemurafenib, have only two highly impacted disease genes based on our algorithm, and therefore returns only one shortest path. (Right): The 5 drug combinations (see drug edges in Figure 1B) assessed for independent mechanism hypothesis based on single drugs with local paths scores presented on the left.

bosutinib	vorinostat	vinblastine	daunorubicin	paclitaxel	lapatinib	Drug A	Drug B
0.5336	0.0038	0.0015	0.0032	0.0020	0.4006	bosutinib	sorafenib
0.4002	0.0031	0.0013		0.0020	0.2006	bosutinib	lapatinib
0.4000	0.0029	0.0012		0.0020	0.2005	bosutinib	vemurafenib
sorafenib	**vemurafenib**	**dabrafenib**	**trametinib**	**cobimetinib**		vemurafenib	lapatinib
0.3201	0.2010	0.2038	0.2676	0.2676		vemurafenib	sorafenib
0.3200		0.2013	0.2016				
0.3200		0.2003	0.0023				

3.4. *Method comparison and robustness evaluation.* First, we provide a comparison between our RWR model and our previously published SynGeNet model in **Table 2**. We observed that the RWR model recovered 76% of validated drug combinations in the top 30% of predictions, whereas SynGeNet recovered 69% of possible validated drug combinations at in the top 30% of predictions. Notably, RWR found 9 unique drug combinations not predicted by SynGeNet, while SynGeNet found 2 unique drug combinations not predicted by RWR. Additionally, to evaluate the robustness of single drug and drug combination prediction models, and consequently the confirm the validity of the highlighted signal flow path as presented in Figures 2A and 2B, we evaluated single and combination drug prediction results by re-wiring the underlying network randomly. The randomly re-wired networks were constructed by keeping the same nodes (genes) in the disease network and randomly assigning connections (edges), while maintaining the same number of connected genes in the original network. We generated 100 random disease networks, and evaluated single drug and drug combination predictions respectively (**Supplemental Figure S3 and Table 2**). We observed that the disease network implemented in our model consistently gave the best performance compared with the 100 randomly generated disease networks. For single drug predictions, our model could predict 8 out of 13 validated drugs, while the highest number of validated drugs predicted by a random network was 6 drugs. Furthermore, the highest number of validated drug combinations predicted by the random networks was 11, compared to the 17 out of 28 predictable, validated drug combinations.
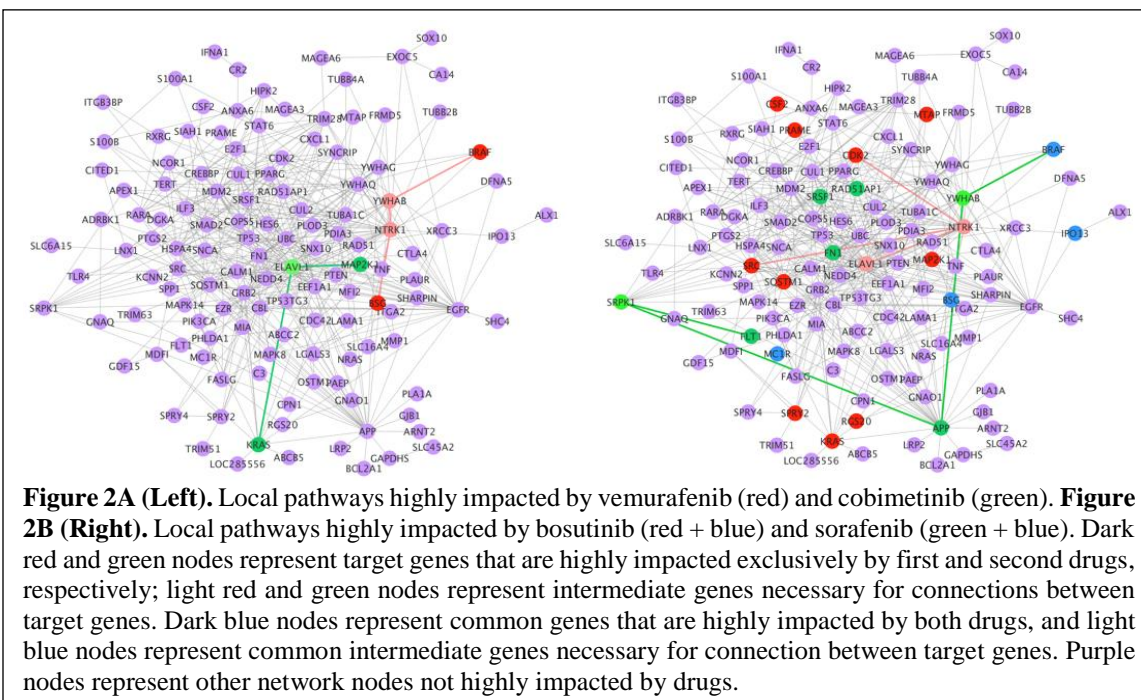

Figure 2A (Left). Local pathways highly impacted by vemurafenib (red) and cobimetinib (green). **Figure 2B (Right).** Local pathways highly impacted by bosutinib (red + blue) and sorafenib (green + blue). Dark red and green nodes represent target genes that are highly impacted exclusively by first and second drugs, respectively; light red and green nodes represent intermediate genes necessary for connections between target genes. Dark blue nodes represent common genes that are highly impacted by both drugs, and light blue nodes represent common intermediate genes necessary for connection between target genes. Purple nodes represent other network nodes not highly impacted by drugs.

Table 2. Comparison between the SynGeNet model and the RWR model proposed in this paper.

	SynGeNet Model	RWR Model
Genomic Data Input	GSE15605 BRAF melanoma gene expression	GSE15605 BRAF melanoma gene expression
Candidate Drug Input	633 FDA approved drugs with gene expression data and known targets	1,433 FDA approved drugs with known targets
Validation Set input	Melanoma cell line drug combination screening (Held *et al*, 2013)	Melanoma cell line drug combination screening (Held *et al*, 2013)
Single Drug Model	GSEA-based connectivity score for signaling network gene signature reversal	Utilize RWR model to rank single drugs according to their average impact to the entire disease network
Drug Combination Model	Drug targets that independently target highly central nodes using averaged topology score	Effective drug combinations are filtered out according to their signal flow path strength evaluation inside the entire disease network
Drug Combination Prediction Evaluation	Among top 30% drug combination predictions, 9 out 13 (69%) validated drug combinations	Among top 30% drug combination predictions, 13 out 17 (76%) validated drug combinations

4. Discussion

In summary, we present a novel approach to predict drug combinations using a RWR model to traverse a heterogeneous network integrating drug target and disease signaling elements. Drug combinations are prioritized that contain individual drugs that can partially impact the disease signaling network through distinct modules and represent different drug communities in order to fulfill the independent mechanism theory. The main advantage of this approach is that it permits a coherent framework to integrate different levels of drug, target and disease information and exploits the entire network structure, including multiple points of entry and pathways to traverse. Therefore, the issue of missing data can be resolved, where RWR can predict drug targets for drugs with no known drug-target interactions through intermediate paths. Additionally, our approach is advantageous over other methods, as it does not require large amounts of drug-induced genomics profiles or other sources of preclinical or clinical efficacy information, which are often lacking. Ultimately, we observed that the RWR model provides a precise delineation of the mechanism of action of drug combinations using propagated signal information as compared to other network-based approaches that may only highlight isolated, central genes. With increasing publicly available patient genomics datasets and ubiquitous drug-target databases, our approach can be applied to a variety of disease contexts and is highly scalable to complex drug-target-gene interaction networks.

There are several limitations of our method to be improved and other challenges for further investigation. First, the constructed signaling network is an undirected graph. The estimation of the signaling diffusion process of signaling on the network may not be entirely accurate. For instance, future studies should implement directed signaling networks (e.g., KEGG signaling network). Also, while affinity propagation is a well-established unsupervised clustering method that has been successfully applied to constructing communities based on drug-drug similarity metrics [17], other methods may be explored including Markov clustering and other energy-model layout algorithms [18]. Second, we can include more extensive drug datasets, as expanding known drug-target interactions will also likely increase the method performance. It will be important to further validate importance of disease-specific signaling networks using larger drug combination validation datasets. Future work could evaluate the effect of incorporating different types

of biological and drug information to construct similarity matrixes (e.g. sequence, chemical structure), and to evaluate the approach in a pan-cancer setting. Another interesting aspect to explore in future work would be the impact of drug community structure and the similarity of toxicity profiles to derive drug combination models that balance efficacy and toxicity simultaneously [19]. Finally, it will be important to confirm predicted mechanisms of drug synergy in prospective *in vitro* experiments (e.g., CRISPR gene editing) to assess the impact of local paths and gene sub-networks in the overall disease signaling network.

References

1.　Holohan, C., et al., *Cancer drug resistance: an evolving paradigm.* Nat Rev Cancer, 2013. **13**(10): p. 714-26.
2.　Bunnage, M.E., *Getting pharmaceutical R&D back on target.* Nat Chem Biol, 2011. **7**(6): p. 335-9.
3.　Madani Tonekaboni, S.A., et al., *Predictive approaches for drug combination discovery in cancer.* Brief Bioinform, 2016.
4.　Huang, L., et al., *DrugComboRanker: drug combination discovery based on target network analysis.* Bioinformatics, 2014. **30**(12): p. i228-36.
5.　Chapman, P.B., et al., *Improved survival with vemurafenib in melanoma with BRAF V600E mutation.* N Engl J Med, 2011. **364**(26): p. 2507-16.
6.　Rizos, H., et al., *BRAF inhibitor resistance mechanisms in metastatic melanoma: spectrum and clinical impact.* Clin Cancer Res, 2014. **20**(7): p. 1965-77.
7.　Lo, R.S. and H. Shi, *Detecting mechanisms of acquired BRAF inhibitor resistance in melanoma.* Methods Mol Biol, 2014. **1102**: p. 163-74.
8.　Long, G.V., et al., *Combined BRAF and MEK inhibition versus BRAF inhibition alone in melanoma.* N Engl J Med, 2014. **371**(20): p. 1877-88.
9.　Regan, K., P. Payne, and F. Li, *Integrative network and transcriptomics-based approach predicts genotype-specific drug combinations for melanoma.* 2017 Joint Summits on Translational Bioinformatics, San Francisco, March 27 ~ 30, 2017, 2017.
10.　Stark, C., et al., *BioGRID: a general repository for interaction datasets.* Nucleic Acids Res, 2006. **34**(Database issue): p. D535-9.
11.　Pinero, J., et al., *DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes.* Database (Oxford), 2015. **2015**: p. bav028.
12.　Bailly-Bechet, M., et al., *Finding undetected protein associations in cell signaling by belief propagation.* Proc Natl Acad Sci U S A, 2011. **108**(2): p. 882-7.
13.　Regan, K., et al., *Drug Repurposing Hypothesis Generation Using the "RE:fine Drugs" System.* J Vis Exp, 2016(118).
14.　Frey, B.J. and D. Dueck, *Clustering by passing messages between data points.* Science, 2007. **315**(5814): p. 972-6.
15.　Wishart, D.S., et al., *DrugBank: a knowledgebase for drugs, drug actions and drug targets.* Nucleic Acids Res, 2008. **36**(Database issue): p. D901-6.
16.　Held, M.A., et al., *Genotype-selective combination therapies for melanoma identified by high-throughput drug screening.* Cancer Discov, 2013. **3**(1): p. 52-67.
17.　Sirci F., et al., *Comparing structural and transcriptional drug networks reveals signatures of drug activity and toxicity in transcriptional responses*. NPJ Syst Biol Appl, 2017. **3**(23).
18.　Udrescu L., et al., *Clustering drug-drug interaction networks with energy model layouts: community analysis and drug repurposing*. Sci Rep, 2016. **6**: p.32745.
19.　Huang H., et al., *Systematic prediction of drug combinations based on clinical side-effects*. Sci Rep, 2014. **24**(4):p.7160.