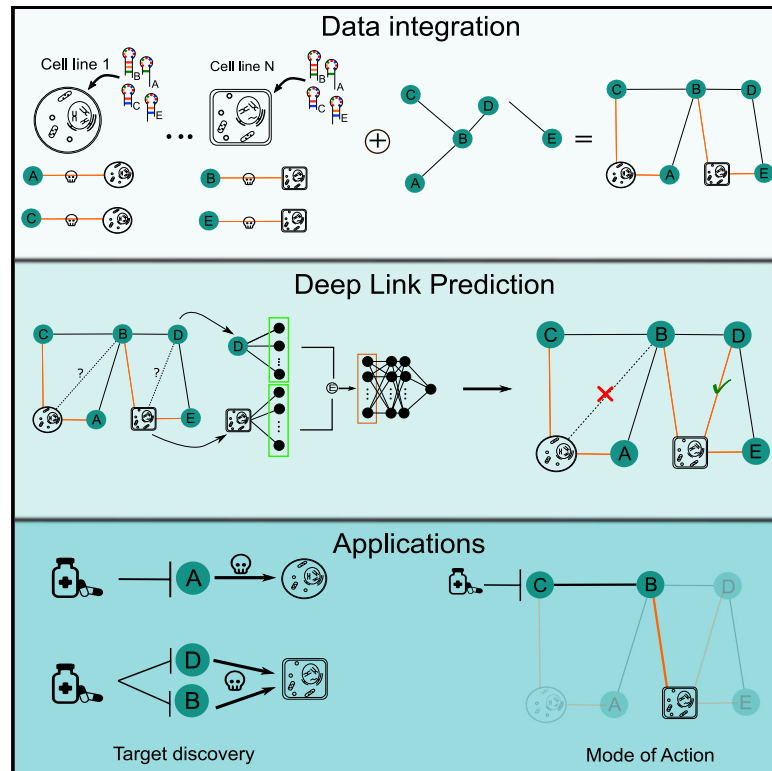**Article**

Extracting functional insights from loss-of-function screens using deep link prediction

Graphical abstract

Authors
Pieter-Paul Strybol, Maarten Larmuseau, Louise de Schaetzen van Brienen, Tim Van den Bulcke, Kathleen Marchal

Correspondence
kathleen.marchal@ugent.be

In brief
Strybol et al. model genetic dependencies in cell lines as a link prediction problem on a graph. Leveraging this approach, the authors present deep link prediction (DLP), a method to predict dependencies in cell lines. DLP improves the identification of drug targets and can unveil their mode of action.

Highlights

- Essential dependencies in cell lines can be modeled as a heterogeneous graph

- Deep link prediction (DLP) outperforms current methods in identifying dependencies

- Highly prioritized genes by DLP are enriched in known drug targets

- DLP can be used to infer the mode of action of a drug

CellPress

Cell Reports Methods

Article

Extracting functional insights from loss-of-function screens using deep link prediction

Pieter-Paul Strybol,[1,3] Maarten Larmuseau,[1,3] Louise de Schaetzen van Brienen,[1] Tim Van den Bulcke,[2] and Kathleen Marchal[1,4,*]

[1]Department of Plant Biotechnology and Bioinformatics, Department of Information Technology, IDLab, imec, iGent Toren, 9000 Gent, Belgium
[2]Galapagos NV, Generaal De Wittelaan L11 A3, 2800 Mechelen, Belgium
[3]These authors contributed equally
[4]Lead contact
*Correspondence: kathleen.marchal@ugent.be
https://doi.org/10.1016/j.crmeth.2022.100171

MOTIVATION Loss-of-function screens are routinely used to identify genes that, when knocked out, result in a phenotype of interest. However, screening experiments are inherently noisy, making it hard to distinguish true from spurious hits. Previous approaches have tried to address this issue by using very stringent thresholds for the identification of true hits; however, this limits the discovery of new targets. Moreover, such a gene-centric approach does not provide insights into the molecular mechanisms underlying a drug response. Our study aims to provide an approach that avoids these limitations.

SUMMARY

We present deep link prediction (DLP), a method for the interpretation of loss-of-function screens. Our approach uses representation-based link prediction to reprioritize phenotypic readouts by integrating screening experiments with gene-gene interaction networks. We validate on 2 different loss-of-function technologies, RNAi and CRISPR, using datasets obtained from DepMap. Extensive benchmarking shows that DLP-DeepWalk outperforms other methods in recovering cell-specific dependencies, achieving an average precision well above 90% across 7 different cancer types and on both RNAi and CRISPR data. We show that the genes ranked highest by DLP-DeepWalk are appreciably more enriched in drug targets compared to the ranking based on original screening scores. Interestingly, this enrichment is more pronounced on RNAi data compared to CRISPR data, consistent with the greater inherent noise of RNAi screens. Finally, we demonstrate how DLP-DeepWalk can infer the molecular mechanism through which putative targets trigger cell line mortality.

INTRODUCTION

Loss-of-function (LOF) screens have become a powerful tool for studying key cellular functions and their relation to disease. Comparing the phenotypic changes between perturbed and normal cells allows the identification of genes essential for maintaining or inducing a certain phenotypic change. This strategy is of particular interest for the discovery of new disease-specific vulnerabilities and potential therapeutic targets (Bortone et al., 2004; Campbell et al., 2016; van Es and Arts, 2005). However, LOF screens (Campbell et al., 2016; Olst et al., 2017) also come with technical challenges resulting in false-negatives (McDonald et al., 2017), false-positives (McFarland et al., 2018; Tsherniak et al., 2017), and variations in the measured phenotypic change

(Lord et al., 2020). Previous studies have sidestepped these issues by imposing a stringent threshold on the observed signals—for example, by focusing only on genes that show strong differential phenotypic effects in a minority of the cell lines (Tsherniak et al., 2017). The downside of this approach is that it results in a limited list of hits, restricting the discovery of new disease genes, potential drug targets, or targeted pathways. However, meta-analysis of individually collected screening experiments offers the opportunity to resolve the distinction more accurately between weakly measured true effects and spurious signals, consequently recovering some of the missed hits.

We therefore propose a meta-analysis framework that leverages available LOF screening data with prior gene-gene interactions to reprioritize measured phenotypic effects. The framework

is based on network-based data interpretation techniques (Cloots and Marchal, 2011; Dimitrakopoulos and Beerenwinkel, 2017; Reyna et al., 2020) inspired by the growing body of work on network representation learning (NRL) (Stanfield et al., 2017; Turki and Wei, 2017; Yue et al., 2020). NRL aims to represent vertices in a graph as low-dimensional, dense vectors that capture the topology of the vertices. Most NRL methods attempt to preserve distances between vertices, such that vertices that are close in the network are also close in the representation space. Using standard machine learning techniques, it is then possible to perform basic tasks on these representations, such as link prediction (LP) (i.e., predicting the probabilities of the edges between the vertices in the graph). NRL-based LP methods are either end-to-end predictors, immediately returning both the vertex representations and the edge probabilities, or methods that learn vertex representations from the input graph (see Table S7) (Perozzi et al., 2014; Tang et al., 2015; Cao et al., 2015). In the latter case, to obtain edge probabilities, the vertex representations of the 2 genes are first combined into an edge representation, for instance, by using a simple binary operator such as the absolute difference of the vertex representations. Subsequently, a classification model such as logistic regression is trained on the obtained edge representations that allows assigning to all pairs of vertices in the graph a probability of interaction.

In our study, we show how prioritizing genes that affect a cell line phenotypically based on LOF screening data can be formulated as an LP problem. We introduce a model coined deep link prediction (DLP) capable of capturing complex network topologies by means of non-linear representations. Using publicly available data from the DepMap Consortium (Dempster et al., 2019; Tsherniak et al., 2017), we show that our method outperforms existing LP methods in prioritizing cancer gene dependencies on both RNAi and CRISPR screens. In addition, we find that a significant positive correlation exists between correctly predicting dependencies in cancer cell lines and retrieving drug targets. These results demonstrate how LP models enable a valuable reprioritization of LOF screening results for target discovery. Finally, we illustrate how predictions made by the LP methods can be used to infer the pathways affected by a knockdown of a gene.

RESULTS

Integrating LOF screens with known functional interactions

The noisiness of LOF screening data hampers distinguishing true from spurious hits. To alleviate this issue, we propose a meta-analysis aggregating screening experiments across different cell lines and combining them with *a priori* known gene-gene interactions. Accordingly, we cast cell line-gene interactions inferred from available LOF screening experiments and a prior gene-gene interaction network in a single heterogeneous graph. Then, we use NRL-based LP modeling to predict the probabilities with which 2 entities in the graph interact. The connectivity in this heterogeneous graph inherently contains information that can improve the prioritization of genes with a phenotypic effect on a cell line. First, not all true dependencies necessarily

display a strong phenotypic effect but are likely part of a pathway in which other genes do display stronger phenotypic effects (McFarland et al., 2018; Tsherniak et al., 2017). Hence, not only should genes that display a strong effect in a cell line be prioritized but also genes displaying relatively weak effects but are close to other genes in the interaction network that do exhibit strong phenotypic effects in the same cell line. Second, by leveraging the size of the screening data, we can assume that weakly measured phenotypic effects can be upgraded in a cell line for genes that display a strong dependency in another cell line that shares many dependencies with the first one. NRL-based LP methods naturally capture such connectivity in the heterogeneous graph to predict cell line-gene probabilities.

Key to our approach is a heterogeneous graph that summarizes information on gene-gene interactions and dependencies obtained from LOF screening data. Cancer dependency of a cell line is here defined as the gene that when knocked down results in the mortality of the cell line (Dempster et al., 2019; Tsherniak et al., 2017). The resulting heterogeneous graph is defined as $G = (V, E)$, where $V = \{v_1, v_2, \ldots, v_n\}$ is the set of vertices and $E = \{e_{i,j}\}$ is the set of edges. Note that there can only be 1 edge between any 2 vertices. The vertex v_i is either a gene or a cell line and the edge between v_i and v_j represents either a gene-gene interaction or a dependency relation between a cell line and a gene.

To illustrate our approach, we used publicly available RNAi screening data from the DepMap Consortium, in which up to 17,309 different genes were knocked down and the effect on the population size of the cell lines was measured. From the DepMap data, we used available cell line-gene dependencies of 7 cancer types (lung, breast, brain, skin, bladder, prostate, and bile duct) and combined them with STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) gene-gene interaction data to reconstruct, for each cancer type, a cancer-specific heterogeneous graph. As the heterogeneous graph is to be used for training, it contains only highly reliably gene-gene and cell line-gene interactions (see STAR Methods). For each cancer type, a different number of cell lines were included in the LOF screening, such that the heterogeneous graphs of the different cancer types have a different number of cell line vertices and dependency interactions (see Table S1). To assess the effect of a larger LOF screening dataset, we also built a heterogeneous graph by combining the cell lines from all 7 cancer types, hereafter referred to as the pan-cancer setting.

As a second LOF dataset, we ran the same analysis on the CRISPR knockout data provided by DepMap, using the same cancer types as for RNAi screening data, but with a different number of cell lines screened per cancer type (see Table S1). The CRISPR knockout data contain LOF information on 17,645 genes, 15,421 of which are also knocked down in the RNAi dataset.

A deep end-to-end LP algorithm

The resulting heterogeneous graph was used to train and benchmark several state-of-the-art NRL-based LP methods. We also implemented a deep learning based model specifically geared toward the biological setting we envisage, DLP (see STAR Methods). Our DLP model is different from existing state-of-the-art methods in its architecture and in the way vertex
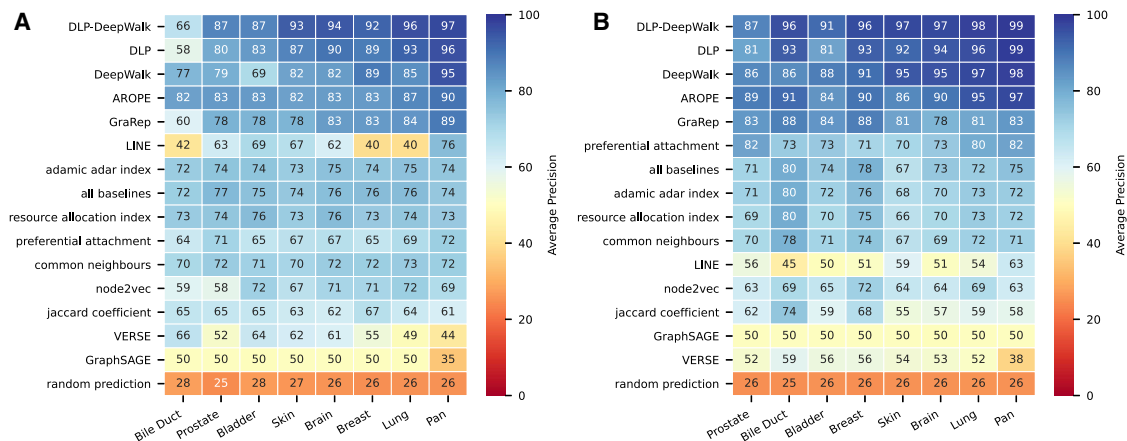
Figure 1. Performance benchmark of several state-of-the-art LP methods on retrieving cancer dependencies from different cancer types
(A and B) Average precision (AP) of LP methods in predicting cell line-gene interactions, based on (A) RNAi- or (B) CRISPR-derived screening scores. Note that the cancer types are listed in ascending order of the number of available cell lines per cancer type. The final column is the AP trained on all cancer types combined.

representations are learned. It uses a deep learning architecture, consisting of a projection layer containing the representation for each vertex, followed by several non-linear hidden layers. Both the representations and the weights in the non-linear layers are trained directly on the LP problem, by predicting whether an interaction is present between pairs of vertices from a training set (see STAR Methods), making it an end-to-end predictor. We also investigated whether the performance of DLP could be improved if the weights in the projection layer were initialized using the representations learned by another method.

Predicting unseen cancer dependencies using LP methods

After training, LP methods can predict new edges between vertices, revealing potential previously unidentified cell line-gene dependencies. To validate the predictive performance, we randomly omitted edges from the heterogeneous graph when training each model—in other words, both cell line-gene and gene-gene—and tested to what extent the LP methods could recover them. Each method is trained using the remaining edges in the graph as positive samples and randomly sampled edges that are not in the graph as negative samples. For negative cell line-gene interactions, we specifically sampled extremely weak dependencies as measured by the original corrected screening score for either RNAi or CRISPR (McFarland et al., 2018; Meyers et al., 2017) (see STAR Methods).

In addition to state-of-the-art NRL-based LP methods, we benchmarked against several baselines based on simple heuristics. These reprioritize dependencies by computing a similarity measure between the neighborhoods of 2 vertices in the heterogeneous graph to infer a likelihood of interaction. In contrast to NRL-based LP methods, these baseline methods do not use vertex representations. As references, we included the Adamic-Adar Index (AAI), the Resource Allocation Index (RAI), Common Neighbors (CN), preferential attachment (PA), and the Jaccard coefficient (JC) (see STAR Methods). In addition, Mara et al. (2020) constructed a simple all baselines NRL-based LP method by combining AAI, RAI, CN, PA, and JC in a single 5-dimensional

edge representation that can be used in combination with logistic regression to compute interaction probabilities.

The performance of the different methods was validated on an unseen test set, in which we considered both the performance on recovering gene-gene interactions and cell line-gene interactions. For each method, we repeated the experiment 3 times, each repeat using a different subsampling of the graph, resulting in 3 different training and independent test sets. This was done for the following 7 cancer types in both the RNAi and CRISPR datasets: bile duct, prostate, bladder, skin, brain, breast, and lung. See Table S1 for an overview of the number of cell lines used in each setting.

The performance of each method was evaluated in terms of the average precision (AP). Note that the dependency problem is highly imbalanced, as typically a cell line has only ~25 strong dependencies for ~10,000 weak dependencies (Table S2). For this reason, we opted to focus on AP as it is shown to be more informative in cases of imbalanced problems (Saito and Rehmsmeier, 2015). To ensure that our benchmarking was independent of the dataset, we applied the same procedure to both the RNAi and CRISPR LOF screens (see STAR Methods). Below, we first elaborate on the results of the benchmarking obtained with the RNAi data and then proceed to validate these results on CRISPR.

For the recovery of gene-gene interactions, most methods achieved a very high AP (Figure S1A), showing that it is feasible to reliably recover known gene-gene interactions. A major reason for this superior performance is the substantial number of edges in the training set. Conversely, predicting unseen cell line-gene interactions of the RNAi dataset is harder, as both the number of cell lines and known dependencies per cell line are limited. Nonetheless, all of the methods achieve an AP that is well above random on predicting cell line-gene interactions. The discrepancy between gene-gene and cell line-gene performance is most pronounced for the baseline methods, suggesting that predicting cell line-gene interactions requires more complex LP methods that also consider higher-order interactions. In addition, Figure 1A emphasizes that if more cell lines have been profiled for a cancer type, then more training data are available, which results
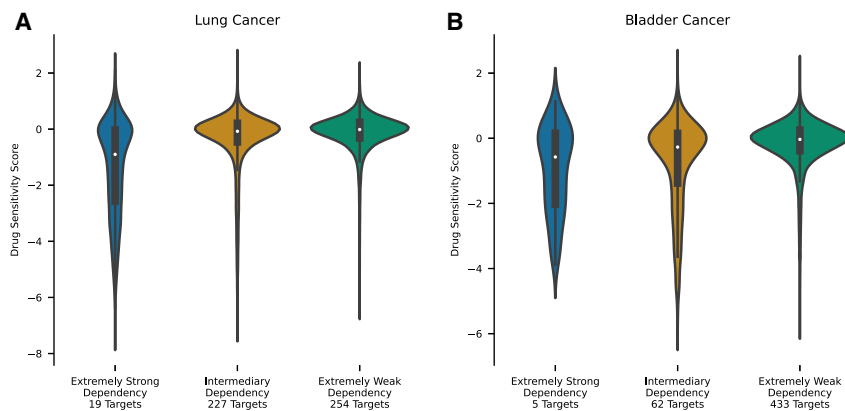
A Lung Cancer

B Bladder Cancer

Figure 2. Discrepancy between RNAi dependency and drug sensitivity scores

(A and B) Distribution of drug sensitivity scores for each RNAi dependency type, specific to (A) lung and (B) bladder cancer. The x axis shows for all known drug targets the cell line-gene interactions binned in 3 categories according to the RNAi dependency score: extremely weak, intermediary, and extremely strong (see STAR Methods). The y axis shows drug sensitivities in the same cell lines in which the dependencies occur. Lower drug sensitivities correspond to a stronger effect. For each category, the number of targets is indicated.

in a better performance for most LP methods. This in stark contrast to the baseline methods, whose performance is seemingly independent of the number of cell lines. These results show that information can flow between different cell lines to improve the performance of NRL-based models. In the case of bile duct cancer, there is only 1 cell line, such that test dependencies are irrevocably lost and their vertex representation is learned solely from the gene-gene interaction network. For other cancer types, the model can use the information from all cell lines from that cancer type to predict a specific cell line-gene interaction. The pan-cancer setting confirms that having more cell lines screened results in an overall higher AP. As shown in Figure S2A, cancer types represented by a lower number of cell lines benefit the most from such a pan-cancer approach. However, Figure S2B also shows that generalizing interactions across cancer types improves the identification of recurrently occurring cell line-gene interactions at the expense of losing some cancer-type-specific interactions. These specific interactions contribute less to the performance than recurrent interactions, but they may be more relevant to the biology of a specific cancer type.

LP methods can only meaningfully predict interactions between vertices in the graph if both of the vertices are seen during training, as otherwise no vertex representation is learned. However, it is possible to learn a vertex representation of a gene solely from gene-gene interactions and then predict the cell line-gene interactions for that gene. Hence, LP methods could predict new dependencies in a cell line, even for genes that were never screened in any of the assessed cell lines, provided these genes were present in the gene-gene interaction network. To assess the performance of NRL-based LP method in correctly predicting unseen cell line-gene interactions, we randomly removed all cell line-gene interactions for 20% of the genes that are at least once a strong dependency in lung cancer cell lines (see STAR Methods). These genes then mimic unscreened genes for which no cell line-gene interaction is available in the training data. Then, we calculated for each unscreened gene the AP of correctly predicting their association with each of the 133 cell lines. Figure S2C shows how DLP outperforms methods from Figure 1A in such a challenging setting.

Furthermore, we assessed whether the performance of DLP could be improved by using representations from a different method as initialization for the projection layer of DLP. We

assumed that simultaneously learning representations and classification is challenging, as the embedding layer alone contains $V_{vertices} \times 128$ weights that need to be trained. To improve the learning, we initialized the projection layer of DLP using the representations of DeepWalk, hereafter referred to as DLP-DeepWalk. We chose DLP-DeepWalk as it is the best-performing LP method, after DLP, when considering all cancer types (Figure 1). Clearly, providing already-information-rich representations as a starting point for DLP facilitates the learning of the model, resulting in a better performance. We also observed that this method outperformed all of the other methods, even the combination of baselines defined by Mara et al. (2020) on predicting both cell line-gene and gene-gene interactions (Figures 1 and S1, respectively). Moreover, even in the setting in which genes are completely omitted from all cell line-gene interactions during training, DLP-DeepWalk seems to achieve superior performance (Figure S2C). Finally, we also tested the robustness of these findings with respect to the network scaffold used. Therefore, we repeated the benchmarking procedure on a different gene-gene interaction network (Reactome FI 2020) (Jassal et al., 2020). The gene-gene and cell line-gene performances, respectively, are presented in Figures S1C and S1D for the RNAi screening data. As the results between these 2 interaction networks are similar, with a slightly better performance on STRING, we use the latter throughout the remainder of this work.

Reprioritizing dependencies using LP improves drug target retrieval

As we have shown above, LP methods can exploit the topology of the heterogeneous graph to correctly predict dependencies in cell lines. In fact, LP methods assign a probability to each cell line-gene interaction, and as such, perform a reprioritization of potential dependencies for each cell line. This prioritization can differ from the ranking based on the original LOF scores.

Such re-ranking could be useful to improve the prioritization of drug targets. This is illustrated in Figure 2, which shows the dependency score that a gene displays in the RNAi screening of a cell line versus the sensitivity that the cell line displays toward a drug targeting this gene, called the drug sensitivity score. Drug sensitivity scores displayed in Figure 2 were also obtained from DepMap (see STAR Methods). A more negative sensitivity score indicates the increased sensitivity of a cell line to a drug.
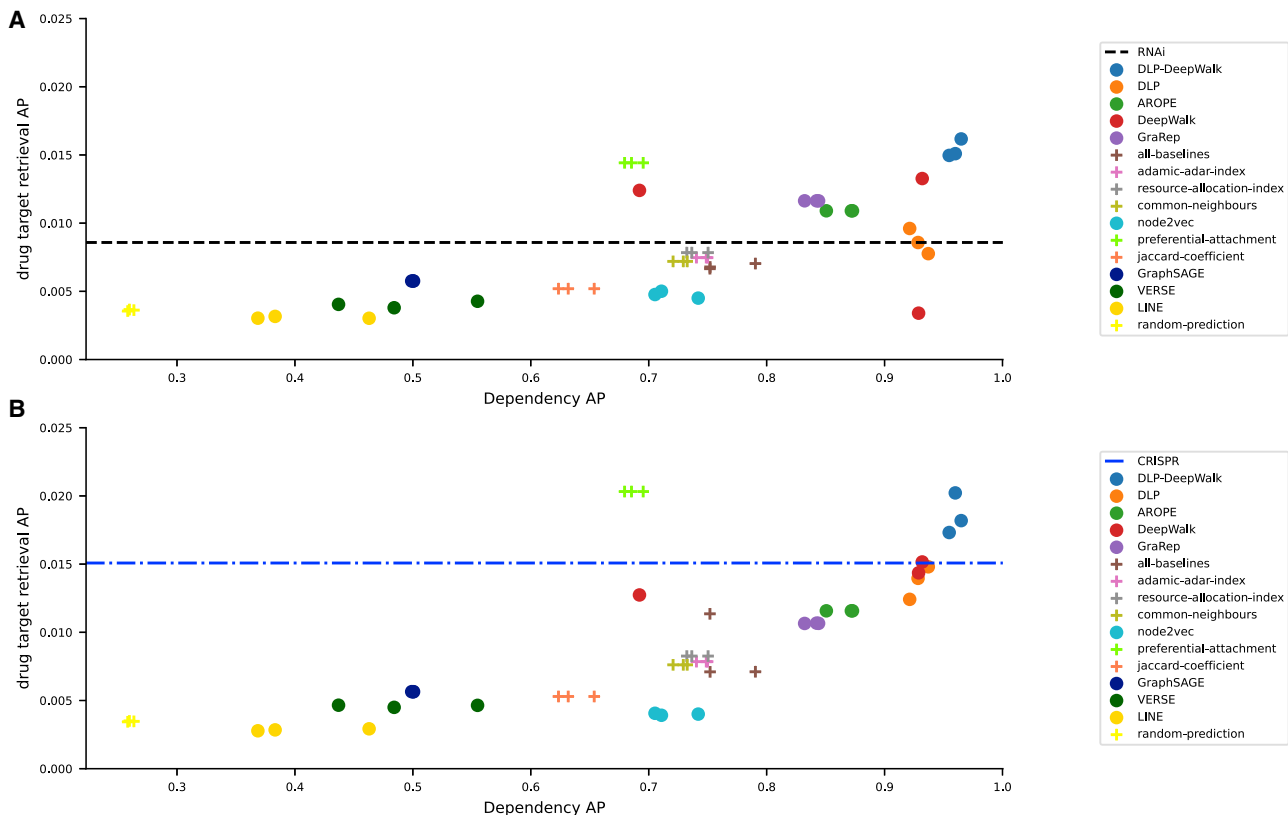
Figure 3. Relation between drug target retrieval and dependency prediction performance for all LP methods

(A and B) AP of each method for cell line-gene dependency predictions and drug target retrieval using (A) RNAi or (B) CRISPR screening data. x axis: AP on correctly predicting a gene dependency on a cell line. y axis: AP on correctly labeling a gene as being a drug target. The horizontal dashed line represents the performance of the ranking based on original RNAi (black) or CRISPR (blue) screening scores in correctly retrieving a drug target. Each method is run 3 times using a different train and test set, and each repeat is shown as a separate dot.

If the ranking based on RNAi dependency scores would allow correctly prioritizing drug targets, then we would expect that targets of drugs with high sensitivity scores in a cell line also display a strong dependency in that cell line. Figure 2 shows that although there is a relationship between the effect a drug has on the cell line (sensitivity score) and the dependency score the target of that drug displays in that cell line (Kruskal-Wallis $p < 0.05$ for both lung and bladder cancers), many targets are missed in the RNAi screening data. Table S3 shows that the same holds true for other cancer types, both in RNAi and CRISPR screens. A considerable number of drug targets with a high sensitivity score tend to display weak dependencies in the same cell lines. As most LOF screening methods put a stringent threshold on the dependency score to avoid prioritizing FP dependencies, many drug targets will be missed.

We hypothesized that sensitive targets should be, despite not displaying the strongest dependencies themselves, functionally related to strong dependencies. In principle, this gene-gene relation can be captured by LP methods using the heterogeneous graph, allowing them to prioritize targets.

To validate whether these ranked lists improved the prioritization of known targets, we generated a benchmark set of true drug targets using the DepMap drug sensitivity scores (see

STAR Methods). True cell line-specific targets were defined as those genes that are known targets of a drug for which a particular cell line is sensitive. Thus, we focused on drugs with a single target to avoid introducing a bias to the target retrieval performance. In fact, when a drug targets multiple genes and if a cell line shows sensitivity to that drug, it is unclear which gene elicited the drug response and, hence, which gene should display high dependency. We calculated the degree to which the LP-based prioritization resulted in an improved drug target retrieval as compared to the ranking obtained from DepMap using the benchmark drug targets as positive labels (see STAR Methods). This improvement is presented as the AP in drug target retrieval obtained after reprioritizing dependencies using the cell line-gene probabilities of each of the LP methods. We assessed this drug target retrieval AP on the RNAi and CRISPR screening data for the different cancer types (Table S4). Figure 3 shows the representative results obtained on lung cancer cell lines for, respectively, a model trained and assessed on the RNAi data (Figure 3A) and on the CRISPR data (Figure 3B). The obtained drug target retrieval performances should be compared to the performance obtained by prioritizing targets according to the original RNAi or CRISPR dependency scores, indicated by the black and blue dashed lines in Figures 3A and 3B,

respectively. Overall, these results show that the added value of reprioritizing targets with LP is more pronounced for RNAi screening than it is for CRISPR data. Interestingly, DLP-Deep-Walk and PA outperform the original prioritization on both the RNAi and CRISPR data. In addition, in each panel, the drug target retrieval AP is compared to the AP on retrieving cell line-gene interactions for lung cancer, hereafter referred to as the dependency AP. The drug target retrieval AP is 2 orders of magnitude lower than the dependency AP for both RNAi and CRISPR, which is to be expected, given that no information related to targets was seen during training. However, the drug target retrieval AP correlates with the dependency AP (Spearman rho = 0.65, p < 0.05). This observed correlation holds true for both RNAi and CRISPR and for all other cancer types as well (see Table S4). This indicates that the methods that perform better in predicting true dependencies by exploiting relationships in the heterogeneous graph also perform better in prioritizing true drug targets. Figure 3 confirms the hypothesis that drug targets must be functionally related to strong dependencies, allowing LP methods to infer missing targets from dependency information. Notably, DLP-DeepWalk exhibits superior performance on correctly predicting true dependencies as well as true drug targets. Although DLP-DeepWalk is among the best-performing methods at prioritizing targets, Figure 3A shows large variance in its performance between different runs of the model. Interestingly, the performance of DLP-DeepWalk is unaffected by this, showing that it learns new representations. The learning itself, however, is facilitated by the sensible initial weights provided by DLP-DeepWalk (Perozzi et al., 2014).

Although the baseline methods typically perform worse than the NRL-based methods, there is 1 notable exception, namely PA. The reason PA has a consistently good drug target retrieval AP despite its relatively low-dependency AP is because it only considers the degree of the vertices in the heterogeneous graph to prioritize dependencies. As such, its predictions are more biased toward the highly connected genes in the network than those of DLP-DeepWalk (see Figure S3A). These are by definition the most studied genes, which often correspond to known drug targets. In fact, drug targets typically have a higher degree in the network, which is also confirmed by the Mann-Whitney U test between the degree of all of the targets used in the benchmark and the remaining genes in the interaction network (1-sided p value = 1.56×10^{-81}). Therefore, PA can be expected to perform well on prioritizing known drug targets that act as hubs in the network. In contrast, DLP-DeepWalk allows the identification of a higher diversity of targets than a degree-based approach, as it does not solely rely on degree information for its predictions (Figure S3B). Therefore, it improves the potential identification of less-studied genes as potential new drug targets.

LP-based reprioritization enriches the number of putative drug targets among the top 100 highest ranked dependencies

Figure 3 shows that the re-ranking of dependencies using LP methods generally results in a better overall drug target prioritization. However, in an applied setting, only the genes showing the strongest dependencies are used for follow-up validation experiments. Hence, to achieve an optimal success rate in follow-up

analysis, the selected top genes should be enriched in genes of interest (here, drug targets). To mimic such a situation, we selected for each LP method its top 100 predicted dependencies per cell line and subsequently assessed whether the top 100 ranked genes in fact contained significantly more drug targets than expected by chance. The top 100 were chosen as an arbitrary threshold to mimic a practical situation in which a small gene list is constructed out of the top predictions for further, more detailed analysis into their potential as drug targets (see STAR Methods). Given that the added value of the LP methods was most pronounced for the RNAi dataset, we used this dataset here.

To estimate the expected number of targets that can be retrieved, we randomly picked 100 genes from the interaction network in 2 ways: (1) either using a uniform distribution in which each gene has an equal chance of being selected and (2) using a degree-based distribution in which the probability for each gene of being selected is proportional to its degree in the gene-gene interaction network (see STAR Methods). Due to the aforementioned bias in the gene-gene interaction network, in which high-degree genes are more likely to be drug targets, degree-based random sampling should be harder to beat than uniform sampling (Guney et al., 2016).

Prioritization performance was assessed by comparing the number of drug targets retrieved in the top 100 of each method with the number of targets present among 100 randomly sampled genes in each cell line. Cell lines were subdivided into 3 classes: (1) significantly worse than random (i.e., cell lines for which a method retrieves significantly fewer targets; p < 0.05), (2) significantly better than random (i.e., cell lines on which a method retrieves significantly more targets; p < 0.05), and (3) neutral (i.e., cell lines for which a method does not perform significantly better or worse than random).

Figure 4 shows the results for both random sampling strategies, highlighting that DLP-DeepWalk is the top performer in prioritizing drug targets and outcompetes the original RNAi dependency scores in many cell lines. Consequently, methods such as DLP-DeepWalk combine gene-gene interaction information with LOF screening data in such a way that the 100 highest-scoring genes become enriched in drug targets. None of the methods performed significantly worse than random.

The fact that LP methods still manage to outperform degree-based random sampling indicates that the higher-order connectivity captured by some LP methods, especially DLP-DeepWalk, provides additional information that can improve drug target prioritization. The added value of the prioritization obtained by DLP-DeepWalk becomes even more apparent when considering, for each cell line, the percentage of true cell line-specific targets retrieved in that cell line. Figure 5 shows that in lung cancer, DLP-DeepWalk has a median target retrieval in its top 100 genes of ∼6.5% of the benchmark drug targets, which is significantly more than the 3.1% obtained by ranking the DepMap RNAi dependency scores (Wilcoxon signed-rank test, false discovery rate [FDR] corrected p-value < 0.05). Moreover, DLP-DeepWalk recovers benchmark targets in most cell lines, whereas GraRep, DLP-DeepWalk, and AROPE do not retrieve any target in ∼ 25% of the cell lines (see Figure 5).

Next, we investigated why certain targets are preferentially retrieved by DLP-DeepWalk. Here, we again used the definition
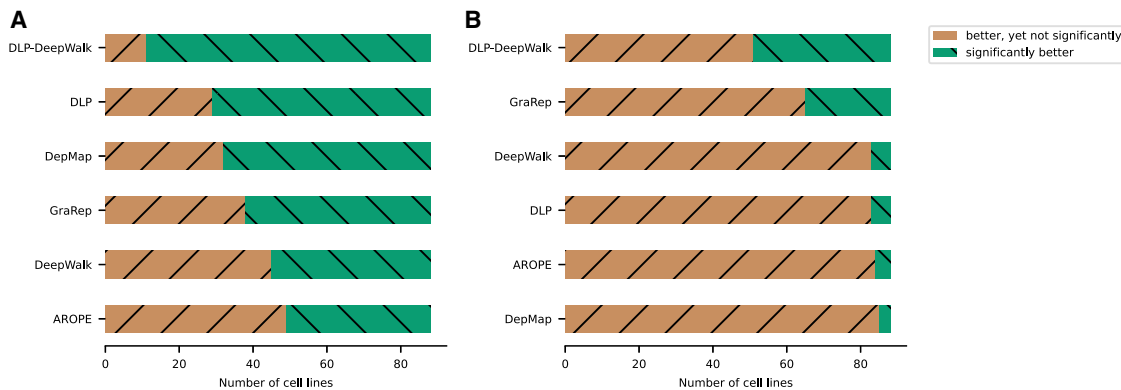
Figure 4. Performance of each method in recovering benchmark drug targets in the top 100 prioritized genes per cell line as compared to random

(A and B) This is assessed by showing the number of cell lines in which each method retrieves the targets (1) significantly better than random; (2) better, yet not significantly, than random; and (3) worse than random. The expected results were obtained by randomly sampling genes from the input graph using a scheme in which each gene has an equal chance of becoming selected—uniform (A) and based on a scheme in which a gene has a probability of being selected equal to its relative degree in the gene-gene interaction scaffold (B).

that a method correctly retrieves a gene as a target in a cell line if the gene is a benchmark target and is in the top 100 predicted cell line-gene interactions of that method for that cell line. For each such target, we could then count in how many cell lines it was correctly retrieved. This allowed us to compare, for each target, in how many cell lines it is correctly retrieved by DLP-DeepWalk compared to the original RNAi screening data. Figure 6 shows on the y axis the difference in the number of cell lines in which a target is correctly retrieved between DLP-DeepWalk and the original RNAi data. Large positive values, such as for *XPO1*, show that *XPO1* is correctly retrieved in 54 more cell lines by DLP-DeepWalk than the original data. Conversely, negative values indicate that the original screening data retrieves targets in more cell lines. Interestingly, the higher target retrieval rate of DLP-DeepWalk can mainly be attributed to 5 genes: *XPO1*, *TOP2A*, *AURKB*, *PLK1*, and *HSP90AA1*. The figure also shows why DLP-DeepWalk can retrieve these targets, as typically they have many neighbors in the heterogeneous graph that are strong RNAi dependencies.

Inferring processes triggered by interfering with a gene of interest using LP predictions

Finally, we verified whether LP methods, and more specifically DLP-DeepWalk, could also be used to infer the pathways or processes triggered by silencing a gene of interest and hence proxy the pathway on which a future drug should act. To infer these processes, we selected the 4 query genes that on average received the highest prioritization across the 133 lung cancer cell lines from the DepMap RNAi dataset and that were also known drug targets from the benchmark set: *KIF11*, *XPO1*, *VCP*, and *PLK1* (see STAR Methods). Focusing on these known targets allows us to compare the inferred processes to those described in the literature that are triggered by interfering with these targets.

For each of these 4 genes, a subnetwork was constructed to reflect the processes triggered by interfering with the gene of interest. Accordingly, we made use of the gene-gene and cell line-

gene probabilities predicted by DLP-DeepWalk trained on the lung DepMap RNAi dataset. We calculated for all of the genes in the network, other than the gene of interest, a weighted score that strikes a balance between displaying high functional similarity to the gene of interest, captured by gene-gene probabilities, and being a dependency in the same cell lines as the gene of interest, captured by the cell line-gene probabilities (see STAR Methods). Using this score, genes that are both functionally similar to the gene of interest and present as a dependency in the same cell lines as the gene of interest will be candidates to include in the subnetwork around the gene of interest. To draw the subnetwork, the 20 highest-scoring genes were mapped on the STRING interaction network, and the largest connected component induced by these genes and containing the gene of interest was selected.

Figure 7 shows as a representative example the inferred subnetwork for the benchmark target, *KIF11*. The inferred *KIF11* subnetwork consisting of 19 genes, including *KIF11* itself, contains only 8 of the 215 direct neighbors that *KIF11* has in the original STRING interaction network. Thus, our approach allows selecting only those neighbors from the interaction network that have a significant chance to also inhibit the same cell lines as *KIF11*. A gene set enrichment analysis (GSEA) (see Table S7) (Subramanian et al., 2005) indicates that the subnetwork around *KIF11* is involved in cell-cycle-related processes in line with the known role of *KIF11*, which belongs to the kinesin-like protein family involved in various kinds of spindle dynamics. The inhibition of *KIF11* by filanesib, a kinesin inhibitor, is known to prevent the formation of the mitotic spindle during the prophase causing cell-cycle arrest (Tao et al., 2005).

Remarkably, the subnetwork of *KIF11* contains 3 other benchmark targets: *AURKB*, *PLK1*, and *XPO1*. Interestingly, *AURKB* and *PLK1* in combination with *KIF11* have been found necessary to prevent excessive DNA replication and aneuploidy (Vassilev et al., 2016). Drugs that target them are known to induce cell-cycle arrest by promoting excessive DNA replication, causing damage and apoptosis (Vassilev et al., 2016).
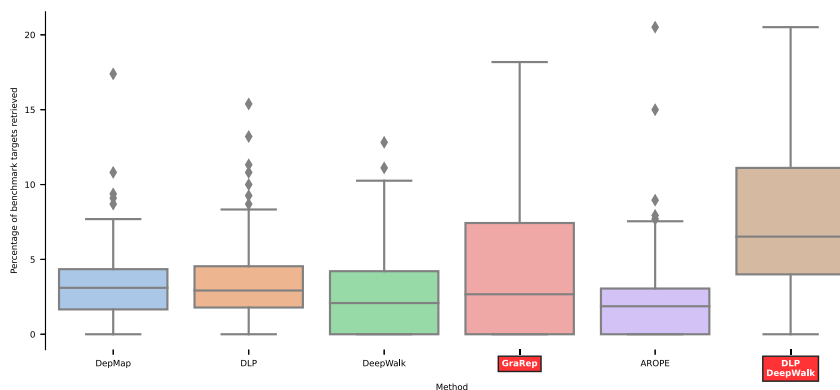
The subnetwork contains another 3 targets that were not considered in the benchmark: *POLA1*—no drugs targeting it were identified with a sensitivity in any cell line below the threshold—and *CDK1* and *PSMA4*, which were not considered as they were targets of drugs known to hit multiple targets. Interestingly, the inhibition of *CDK1*, a cyclin-dependent kinase (Malumbres et al., 2009), is known to cause a synergistic effect with kinesin inhibitors targeting *KIF11* to promote cell death by mitotic slippage (Tao et al., 2005). Synergistic effects between *KIF11* inhibitors and Aurora kinase inhibitors have also been described (Ma et al., 2014).

The resulting subnetwork nicely illustrates that known drug targets such as *KIF11*, *AURKB*, *PLK1*, *XPO1*, *POLA1*, *CDK1*, and *PSMA4* are indeed closely connected on the gene-gene interaction network, as they are involved in similar essential processes. The fact that, aside from the query gene, no target information was used to obtain the *KIF11* subnetwork clearly shows that inte-

grating LOF screens with known gene-gene interactions can reveal interesting biology. It also suggests that to discover putative targets, it may merit searching around a known target—here, *KIF11* in the induced subnetwork.

Subnetworks inferred for each of the 3 other query genes are shown in Figure S4. The functional annotation of each of the subnetworks using GSEA can be found in Table S3. Subnetworks around *XPO1* and *PLK1* are enriched in cell-cycle pathways, related to the G1/S phase and G2/M phase, respectively. Drugs inhibiting these targets are known to trigger similar cell-cycle-related processes. For instance, exportin antagonists, which target *XPO1*, invoke DNA double-stranded breaks, associated with the G1/S phase, causing decreased DNA replication (Burke et al., 2017). Similarly, the inhibition of *PLK1* kinase with a peak expression in the G2/M phase also leads to cell-cycle arrest because *PLK1* loses its function as a cell-cycle regulator (Pezuk et al., 2013).

The subnetwork around *VCP* is enriched in protein degradation processes. This agrees with the known role of *VCP*, a member of the AAA-ATPase gene family (Beskow et al., 2009), and the known effect of inhibiting of *VCP*. ATPase inhibition of *VCP* leads to cancer cell line mortality due to increased proteotoxic stress (Bastola et al., 2019; Deshaies, 2014).

These subnetworks inferred around query genes provide insight into the functional role of the query genes in the cells and hence, also into the potential mode of action of drugs interfering with this target.

DISCUSSION

In this work, we have shown how integrating *a priori* known gene-gene interaction information with available screening data can be used to improve the inference of true dependencies from LOF screening data. Accordingly, we cast the reprioritization problem as an LP problem.

Our results demonstrate that the heterogeneous graph that integrates gene-gene with cell line-gene interactions contains information that allows capturing the original prioritization but can also predict unseen hits. Representation-based LP methods are ideally suited for this task, as they do not rely on a binary representation of edges, allowing them to naturally cope with the incompleteness of the interaction information and LOF screening data. Although cell line-gene interactions, denoting strong dependencies, represent only a small fraction of the edges in the heterogeneous graph, we have shown
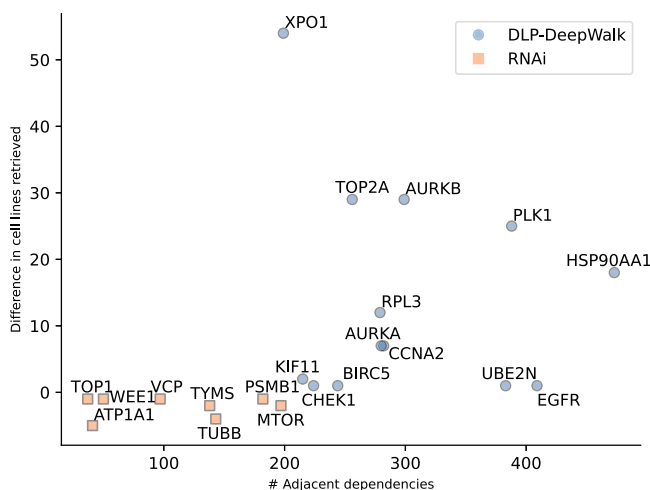

Figure 6. Genes that have more neighboring genes in the heterogeneous graph that are RNAi dependencies (x axis) are more likely to be found by DLP-DeepWalk than by the original RNAi data
The y axis represents the difference in the number of cell lines in which a gene is correctly recovered as target, between DLP-DeepWalk and DepMap. The orange squares denote drug targets that are recovered in more cell lines by ranking on the original RNAi screening score, while blue dots are recovered more by ranking based on the probabilities provided by DLP-DeepWalk.
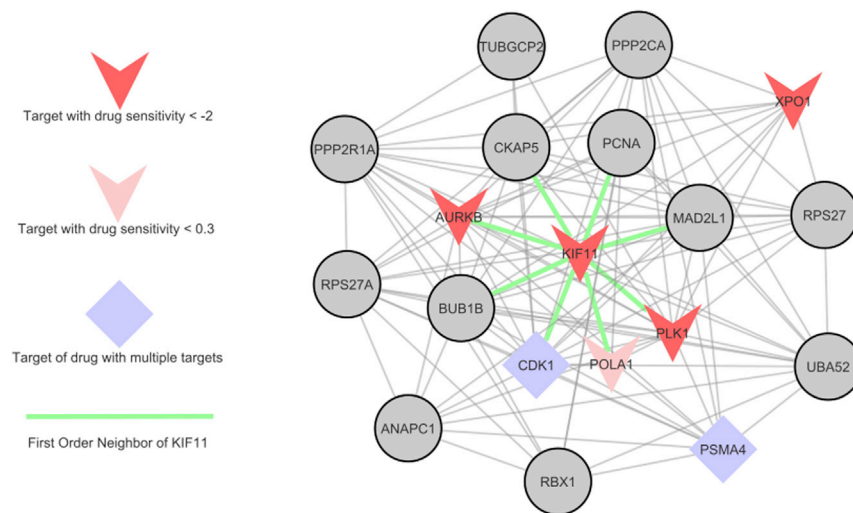
Figure 7. Subnetwork around known drug target *KIF11* proxying the molecular mechanism through which it affects cell lines
Genes connected by green edges are all first-order neighbors of *KIF11* in the original STRING interaction network.

gets are functionally related to genes exhibiting a dependency. To assess the practical use of these LP methods, we additionally verified their ability to better prioritize potential targets, restricting our attention to the top 100 genes prioritized by each method in RNAi. Several LP methods outperformed the ranking based on the original RNAi dependency scores, indicating that LP methods can aid in obtaining a better reprioritization for downstream validation experiments. Still, a large fraction of true cell line-specific drug targets was missed by both DLP-DeepWalk and RNAi in their top 100 predictions. Explicitly using target information during training (Sachdev and Gupta, 2019) could enrich the top 100 predictions in drug targets even more, but this was not the intention of our approach, which aimed at an unbiased ranking based on experimental screening information only. Such an unbiased approach could be very useful for other disease areas for which less therapeutic information is available.

Finally, we have shown how these LP methods can be used to elucidate the processes affected when interfering with a gene of interest. Using 4 known drug targets as an example, we have illustrated how predictions on the probabilities of gene-gene and cell line-gene interactions made by DLP-DeepWalk can be used to infer processes triggered by interfering with a gene of interest. We could observe that the subnetworks were also enriched in other true cell line-specific drug targets, indicating that several genes in the same essential processes or pathways affect a cell line in the same way. This was confirmed by the fact that the different drugs hitting these different targets are known to interfere with the same process. For target discovery, LP methods could thus be used to identify the affected processes when interfering with candidate targets. Such an identification could result in the discovery of previously unidentified putative targets involved in similar processes.

Limitations of the study

This work focuses only on a single disease area, namely cancer, as for cancer large screening datasets were available which allowed to benchmark various LP methods. However, the proposed approach could be even more beneficial for recommending dependencies in less-studied disease areas.

All predictions only relied on topological information of the heterogeneous graph. To improve target retrieval, a useful extension of the model, it is necessary to turn to a supervised approach that considers more features related to the properties of the drugs used and to explicitly model drugs as a separate vertex type in the heterogeneous graph. In this work, we omitted all

that most LP methods can still accurately predict these interactions.

To improve the dependency prediction, we have introduced a model named DLP-DeepWalk. This model outperforms all of the other methods in predicting gene-gene and cell line-gene interactions for both the RNAi and CRISPR screening data. DLP-DeepWalk combines the power of non-linear layers from neural networks with the embeddings that are learned by another LP method, in this case, DLP-DeepWalk. For both RNAi and CRISPR, the performance of the different LP methods indicates that dependencies do not occur at random throughout the interactome but follow distinct topological patterns that can be captured using LP. Interestingly, the benchmarking results are in general better on CRISPR data, suggesting that the incidence of CRISPR dependencies on the network follows an even better-defined pattern, which could be caused by less off-target effects (Evers et al., 2016; Shalem et al., 2014). This suggests that LP performance could give an indication of the quality of screening experiments.

In addition, our results highlight the value of screening multiple cell lines from the same disease as, for most methods, the performance increases with the number of screened cell lines. We also demonstrated that the added value of NRL-based methods still holds when using a different gene-gene interaction network. This shows that not the specific gene-gene interaction network, but the underlying biology is important. Figure 6 shows that most targets recovered are genes that are neighbors of strong dependencies. These genes do not have a strong phenotypic readout themselves, but are expected to belong to the same essential pathway as many adjacent genes that do have a strong phenotypic readout.

As an application of our method, we have compared the predictions of our model to a set of cell line-specific drug targets. Interestingly, it appears that methods that excel at predicting dependencies also perform better at recovering targets, without explicitly being trained on cell line-target information. This finding was observed in all cancer types and for each dataset in both RNAi and CRISPR, indicating that at least some drug tar-

Cell Reports Methods
Article

drugs that have multiple annotated drug targets to mitigate ambiguity in the analysis, but explicitly modeling drug-related this information could also resolve ambiguity. Finally, when predicting drug targets directly using a gene-gene interaction network, one needs to carefully correct the performance of the model for the inherent bias of these networks toward known drug targets.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- METHODS DETAILS
  - DepMap data RNAi and CRISPR
  - DLP model description
  - Generating the input graph for LP
  - Training data
  - Benchmarking LP methods with EvalNE
  - Predicting unseen genes
  - Analyzing the molecular mechanism of dependencies
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Evaluation metrics
  - AP of LP methods in predicting drug targets
  - Performance of retrieving benchmark targets

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.crmeth.2022.100171.

AUTHOR CONTRIBUTIONS

Conceptualization and supervision, K.M. and T.V.d.B.; methodology, P.-P.S., M.L., and L.d.S.v.B.; validation, formal analysis, and investigation, P.-P.S. and M.L.; writing – original draft, P.-P.S., M.L., and K.M.; writing – review & editing, P.-P.S., M.L., K.M., and L.d.S.v.B.; visualization, P.-P.S. and M.L.; funding acquisition, K.M. and T.V.d.B.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Bastola, P., Bilkis, R., De Souza, C., Minn, K., and Chien, J. (2019). Heterozygous mutations in valosin - containing protein (VCP) and resistance to VCP inhibitors. Sci. Rep. 9, 11002. https://doi.org/10.1038/s41598-019-47085-9.

Beskow, A., Grimberg, K.B., Bott, L.C., Salomons, F.A., Dantuma, N.P., and Young, P. (2009). A conserved unfoldase activity for the p97 AAA-ATPase in proteasomal degradation. J. Mol. Biol. 394, 732–746. https://doi.org/10.1016/j.jmb.2009.09.050.

Bortone, K., Michiels, F., Vandeghinste, N., Tomme, P., and van Es, H. (2004). Functional screening of viral siRNA libraries in human primary cells. DDW Drug Discov. World 5, 20–28.

Braschi, B., Denny, P., Gray, K., Jones, T., Seal, R., Tweedie, S., Yates, B., and Bruford, E. (2019). Genenames.org: the HGNC and VGNC resources in 2019. Nucleic Acids Res. 47, D786–D792. https://doi.org/10.1093/nar/gky930.

Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., Niculae, V., Prettenhofer, P., Gramfort, A., Grobler, J., et al. (2013). API design for machine learning software: experiences from the Scikit-learn project.https://arxiv.org/abs/1309.0238.

Burke, R.T., Marcus, J.M., and Orth, J.D. (2017). Inhibition of exportin-1 function results in rapid cell cycle-associated DNA damage in cancer cells. Oncotarget 8, 39460–39475. https://doi.org/10.18632/oncotarget.17063.

Campbell, J., Ryan, C.J., Brough, R., Bajrami, I., Pemberton, H.N., Chong, I.Y., Costa-Cabral, S., Frankum, J., Gulati, A., and Holme, H. (2016). Large-scale profiling of kinase dependencies in cancer cell lines. Cell Rep. 14, 2490–2501.

Cao, S., Lu, W., and Xu, Q. (2015). GraRep: learning graph representations with global structural information. In Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, CIKM'15 (Association for Computing Machinery), pp. 891–900. https://doi.org/10.1145/2806416.2806512.

Chollet, F., et al. (2015). Keras. Github.

Cloots, L., and Marchal, K. (2011). Network-based functional modeling of genomics, transcriptomics and metabolism in bacteria. Curr. Opin. Microbiol. 14, 599–607. https://doi.org/10.1016/j.mib.2011.09.003.

Corsello, S.M., Bittker, J.A., Liu, Z., Gould, J., McCarren, P., Hirschman, J.E., Johnston, S.E., Vrcic, A., Wong, B., Khan, M., et al. (2017). The Drug Repurposing Hub: a next-generation drug library and information resource. Nat. Med. 23, 405–408. https://doi.org/10.1038/nm.4306.

Corsello, S.M., Nagari, R.T., Spangler, R.D., Rossen, J., Kocak, M., Bryan, J.G., Humeidi, R., Peck, D., Wu, X., Tang, A.A., et al. (2019). Non-oncology drugs are a source of previously unappreciated anti-cancer activity. bioRxiv, 730119. https://doi.org/10.1101/730119.

Dempster, J.M., Rossen, J., Kazachkova, M., Pan, J., Kugener, G., Root, D.E., and Tsherniak, A. (2019). Extracting biological insights from the project achilles genome-scale CRISPR screens in cancer cell lines.https://doi.org/10.1101/720243.

Deshaies, R.J. (2014). Proteotoxic crisis, the ubiquitin-proteasome system, and cancer therapy. BMC Biol. 12, 94. https://doi.org/10.1186/s12915-014-0094-0.

Dimitrakopoulos, C.M., and Beerenwinkel, N. (2017). Computational approaches for the identification of cancer genes and pathways. Wiley Interdiscip. Rev. Syst. Biol. Med. 9, e1364. https://doi.org/10.1002/wsbm.1364.

Evers, B., Jastrzebski, K., Heijmans, J.P.M., Grernrum, W., Beijersbergen, R.L., and Bernards, R. (2016). CRISPR knockout screening outperforms shRNA and CRISPRi in identifying essential genes. Nat. Biotechnol. 34, 631–633. https://doi.org/10.1038/nbt.3536.

Grover, A., and Leskovec, J. (2016). node2vec: scalable feature learning for networks.https://arxiv.org/abs/1607.00653.

Guney, E., Menche, J., Vidal, M., and Barábasi, A.-L. (2016). Network-based in silico drug efficacy screening. Nat. Commun. 7, 10331. https://doi.org/10.1038/ncomms10331.

Jassal, B., Matthews, L., Viteri, G., Gong, C., Lorente, P., Fabregat, A., Sidiropoulos, K., Cook, J., Gillespie, M., Haw, R., et al. (2020). The reactome pathway knowledgebase. Nucleic Acids Res. 48, D498–D503. https://doi.org/10.1093/nar/gkz1031.

Kingma, D.P., and Ba, J. (2017). Adam: a method for stochastic optimization.https://arxiv.org/abs/1412.6980.

Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P., and Mesirov, J.P. (2011). Molecular signatures database (MSigDB) 3.0. Bioinformatics 27, 1739–1740. https://doi.org/10.1093/bioinformatics/btr260.

Lord, C.J., Quinn, N., and Ryan, C.J. (2020). Integrative analysis of large-scale loss-of-function screens identifies robust cancer-associated genetic interactions. Elife 9, e58925.

Ma, H.T., Erdal, S., Huang, S., and Poon, R.Y.C. (2014). Synergism between inhibitors of Aurora A and KIF11 overcomes KIF15-dependent drug resistance. Mol. Oncol. 8, 1404–1418. https://doi.org/10.1016/j.molonc.2014.05.007.

Malumbres, M., Harlow, E., Hunt, T., Hunter, T., Lahti, J.M., Manning, G., Morgan, D.O., Tsai, L.-H., and Wolgemuth, D.J. (2009). Cyclin-dependent kinases: a family portrait. Nat. Cell Biol. 11, 1275–1276. https://doi.org/10.1038/ncb1109-1275.

Mara, A., Lijffijt, J., and De Bie, T. (2020). Network representation learning for link prediction: are we improving upon simple heuristics?.http://arxiv.org/abs/2002.11522.

Mara, A., Lijffijt, J., and De Bie, T. (2019). EvalNE: a framework for evaluating network embeddings on link prediction.https://arxiv.org/abs/1901.09691.

McDonald, E.R., de Weck, A., Schlabach, M.R., Billy, E., Mavrakis, K.J., Hoffman, G.R., Belur, D., Castelletti, D., Frias, E., Gampa, K., et al. (2017). Project DRIVE: a compendium of cancer dependencies and synthetic lethal relationships uncovered by large-scale, deep RNAi screening. Cell 170, 577–592.e10. https://doi.org/10.1016/j.cell.2017.07.005.

McFarland, J.M., Ho, Z.V., Kugener, G., Dempster, J.M., Montgomery, P.G., Bryan, J.G., Krill-Burger, J.M., Green, T.M., Vazquez, F., Boehm, J.S., et al. (2018). Improved estimation of cancer dependencies from large-scale RNAi screens using model-based normalization and data integration. Nat. Commun. 9, 4610. https://doi.org/10.1038/s41467-018-06916-5.

Meyers, R.M., Bryan, J.G., McFarland, J.M., Weir, B.A., Sizemore, A.E., Xu, H., Dharia, N.V., Montgomery, P.G., Cowley, G.S., Pantel, S., et al. (2017). Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. Nat. Genet. 49, 1779–1784. https://doi.org/10.1038/ng.3984.

Olst, E.S., Blijlevens, M., de Menezes, R.X., van der Meulen-Muileman, I.H., Smit, E.F., and van Beusechem, V.W. (2017). A genome-wide siRNA screen for regulators of tumor suppressor p53 activity in human non-small cell lung cancer cells identifies components of the RNA splicing machinery as targets for anticancer treatment. Mol. Oncol. 11, 534–551. https://doi.org/10.1002/1878-0261.12052.

Perozzi, B., Al-Rfou, R., and Skiena, S. (2014). DeepWalk: online learning of social representations. Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. - KDD 14, 701–710. https://doi.org/10.1145/2623330.2623732.

Pezuk, J.A., Brassesco, M.S., Morales, A.G., de Oliveira, J.C., de Oliveira, H.F., Scrideli, C.A., and Tone, L.G. (2013). Inhibition of polo-like kinase 1 induces cell cycle arrest and sensitizes glioblastoma cells to ionizing radiation. Cancer Biother. Radiopharm. 28, 516–522. https://doi.org/10.1089/cbr.2012.1415.

Reyna, M.A., Haan, D., Paczkowska, M., Verbeke, L.P.C., Vazquez, M., Kahraman, A., Pulido-Tamayo, S., Barenboim, J., Wadi, L., Dhingra, P., et al. (2020). Pathway and network analysis of more than 2500 whole cancer genomes. Nat. Commun. 11. https://doi.org/10.1038/s41467-020-14367-0.

Sachdev, K., and Gupta, M.K. (2019). A comprehensive review of feature based methods for drug target interaction prediction. J. Biomed. Inform. 93, 103159. https://doi.org/10.1016/j.jbi.2019.103159.

Saito, T., and Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS ONE 10, e0118432. https://doi.org/10.1371/journal.pone.0118432.

Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelsen, T.S., Heckl, D., Ebert, B.L., Root, D.E., Doench, J.G., and Zhang, F. (2014). Genome-Scale CRISPR-Cas9 knockout screening in human cells. Science 343, 84–87. https://doi.org/10.1126/science.1247005.

Stanfield, Z., Coşkun, M., and Koyutürk, M. (2017). Drug response prediction as a link prediction problem. Sci. Rep. 7, 40321. https://doi.org/10.1038/srep40321.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U S A 102, 15545–15550. https://doi.org/10.1073/pnas.0506580102.

Tang, J., Qu, M., Wang, M., Zhang, M., Yan, J., and Mei, Q. (2015). LINE: large-scale information network embedding. Proc. 24th Int. Conf. World Wide Web 15, 1067–1077. https://doi.org/10.1145/2736277.2741093.

Tao, W., South, V.J., Zhang, Y., Davide, J.P., Farrell, L., Kohl, N.E., Sepp-Lorenzino, L., and Lobell, R.B. (2005). Induction of apoptosis by an inhibitor of the mitotic kinesin KSP requires both activation of the spindle assembly checkpoint and mitotic slippage. Cancer Cell 8, 49–59. https://doi.org/10.1016/j.ccr.2005.06.003.

Tsherniak, A., Vazquez, F., Montgomery, P.G., Weir, B.A., Kryukov, G., Cowley, G.S., Gill, S., Harrington, W.F., Pantel, S., Krill-Burger, J.M., et al. (2017). Defining a cancer dependency map. Cell 170, 564–576.e16. https://doi.org/10.1016/j.cell.2017.06.010.

Tsitsulin, A., Mottin, D., Karras, P., and Müller, E. (2018). VERSE: versatile graph embeddings from similarity measures. Proc. 2018 World Wide Web Conf. 18, 539–548. https://doi.org/10.1145/3178876.3186120.

Turki, T., and Wei, Z. (2017). A link prediction approach to cancer drug sensitivity prediction. BMC Syst. Biol. 11, 94. https://doi.org/10.1186/s12918-017-0463-8.

van Es, H., and Arts, G. (2005). Biology calls the targets: combining RNAi and disease biology. Drug Discov. Today 10, 1385–1391. https://doi.org/10.1016/s1359-6446(05)03590-7.

Vassilev, A., Lee, C.Y., Vassilev, B., Zhu, W., Ormanoglu, P., Martin, S.E., and DePamphilis, M.L. (2016). Identification of genes that are essential to restrict genome duplication to once per cell division. Oncotarget 7, 34956–34976. https://doi.org/10.18632/oncotarget.9008.

Wu, G., Feng, X., and Stein, L. (2010). A human functional protein interaction network and its application to cancer data analysis. Genome Biol. 11, R53. https://doi.org/10.1186/gb-2010-11-5-r53.

Yue, X. (2019). BioNev github, GitHub. https://github.com/xiangyue9607/BioNEV/tree/master/data/STRING_PPI.

Yue, X., Wang, Z., Huang, J., Parthasarathy, S., Moosavinasab, S., Huang, Y., Lin, S.M., Zhang, W., Zhang, P., and Sun, H. (2020). Graph embedding on biomedical networks: methods, applications and evaluations. Bioinformatics 36, 1241–1251. https://doi.org/10.1093/bioinformatics/btz718.

Zhang, Z., Cui, P., Wang, X., Pei, J., Yao, X., and Zhu, W. (2018). Arbitrary-order proximity preserved network embedding. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD '18 (Association for Computing Machinery), pp. 2778–2786. https://doi.org/10.1145/3219819.3219969.

Zhou, T., Lu, L., and Zhang, Y.-C. (2009). Predicting missing links via local information. Eur. Phys. J. B 71, 623–630. https://doi.org/10.1140/EPJB/E2009-00335-8.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
RNAi knock-down data	https://depmap.org/portal/download/	D2_combined_gene_dep_scores.csv
CRISPR knock-out data	https://depmap.org/portal/download/	CRISPR_gene_effect.csv
Drug sensitivity screening data	https://depmap.org/portal/download/	primary-screen-replicate-collapsed-logfold-change.csv
STRING interaction network	https://github.com/xiangyue9607/BioNEV/tree/master/data/STRING_PPI	N/A
Reactome FI 2020 interaction network	https://reactome.org/download/tools/ReatomeFIs/FIsInGene_122220_with_annotations.txt	N/A
Drug target annotation file	https://clue.io/repurposing#download-data	N/A
Software and algorithms		
Deep Link Prediction	This paper	https://github.com/pstrybol/DeepLinkPrediction_Public, https://doi.org/10.5281/zenodo.5844581
EvalNE	Mara et al. (2019)	https://github.com/Dru-Mara/EvalNE

RESOURCE AVAILABILITY

Lead contact
Further information and requests for code should be directed to and will be fulfilled by the lead contact, Kathleen Marchal (Kathleen.Marchal@UGent.be).

Materials availability
This study did not generate new unique reagents.

Data and code availability
- All RNAi and CRISPR screening data as well as drug sensitivity screening data are publicly available from the DepMap portal (https://depmap.org/portal/download/). Specifically for the RNAi data: DEMETER2 Data v6, file name "D2_combined_gene_dep_scores.csv" added on 04/20. And for the CRISPR data: DepMap Public 21Q2, file name "CRISPR_gene_effect.csv", released on 05/21. Finally, for the drug sensitivity dataset: PRISM Repurposing 19Q4, file name "primary-screen-replicate-collapsed-logfold-change.csv", released on 12/19.
- All original code has been deposited at https://github.com/pstrybol/DeepLinkPrediction_Public and is publicly available as of the date of publication. DOIs are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

METHODS DETAILS

DepMap data RNAi and CRISPR
The RNAi dataset consists of LOF screening results from 713 cell lines covering 17,309 genes in total, while the CRISPR data cover 990 cell lines and 13,345 genes in total. For each cell line-gene pair, the dataset contains a dependency score, a continuous value ranging from below −5 to almost 3 for RNAi and from below −2 to around 1 for CRISPR. RNAi measurements were corrected for off-target effects using the DEMETER2 tool (McFarland et al., 2018) while the CRISPR data were corrected using the CERES tool (Meyers et al., 2017). The more negative the dependency score, the stronger the dependency, i.e., cancer cell proliferation is halted more when these genes are knocked down. If the dependency score of a gene is close to zero or positive, the effect of knocking down the gene does not result in any cell death or change in proliferation. The latter type of dependencies is referred to as weak dependencies. In total, cell lines from seven different cancer types were considered. Cell lines from lung, breast, brain, and skin cancer were selected because these types have the largest number of cell lines profiled in DepMap. Bile duct, prostate, and bladder were also added to represent cancer types for which less samples are available. Having a wide range of sample sizes allowed us to assess

the impact of the sample size on the method performance and particularly for bile duct to test whether LP-methods could also deal with a single cell line. Finally, we also considered a pan cancer setting that combined all cell lines from these seven cancer types.

Because dependency scores in DepMap are expressed as continuous values, we subdivided them in three categories: strong, intermediary, or weak. To this end, we used two stringent thresholds: a certain gene G being a dependency of cell line C with DepMap score $S_{(C,P)}$ above T_{neg} falls in the weak dependency category, while genes with DepMap score below T_{pos} are categorized as strong dependencies, with $E_{(C, G)}$ representing an edge between cell line C and gene G, V_c and V_G representing the collection cell line and gene nodes, respectively.

$$E_{(C,G),C \in V_c, G \in V_G} = \begin{cases} 0, & if\ S_{(C,G)} > T_{neg} \\ 1, & if\ S_{(C,G)} < T_{pos} \end{cases}$$

Interactions with a dependency score in between T_{neg} and T_{pos} are referred to as interactions with an intermediary dependency score. For those, it is harder to judge whether they represent a true dependency or not. As the ground truth label for these interactions is unknown, these interactions with intermediary dependencies are not used to train any of the LP methods. For RNAi data, T_{neg} was -0.5 and T_{pos} -1.5. The thresholds for CRISPR data were chosen in such a way that the number of positive and negative samples exactly matches those of the RNAi data to obtain comparable results (the specific thresholds can be found in Table S5).

Drug sensitivity data were also obtained from the DepMap portal. This dataset contains 4,686 compounds screened for sensitivity in 578 cell lines spanning 24 different cancer types (Corsello et al., 2019). We used this dataset to construct a benchmark dataset of known drug targets per cell line. Only drugs that were sensitive for a cell line for which LOF screening was available, were considered for the benchmark. For each tested cell line, the benchmark dataset contains the targets of drugs that display a high sensitivity in that cell line. To construct a conservative benchmark, we chose quite a strict sensitivity threshold of -2 on the drug sensitivity level. For each of these retained drugs, the matching targets were retrieved from the annotation file (version 3/24/2020) available on the drug repurposing hub (Corsello et al., 2017). For benchmarking, only the drugs with a single reported target were considered (see Table S7).

DLP model description

The DLP model proposed in this paper is inspired from the field of NRL and uses a deep neural network architecture as shown in Figure S5A. The model takes as input a pair of vertices and converts them using an embedding or projection layer to their -dimensional representations. Then, these two vertex representations are combined, using one of four binary operators (Mara et al., 2020), to form a vertex-pair representation. This vertex-pair representation is further used as the input to a feedforward neural network that consists of two hidden layers (32 neurons each with ReLU activation) and an output layer. The vertex representations, i.e. the weights of the embedding layer, and the feedforward network are learned simultaneously using binary cross entropy loss and Adam adaptive learning rate (Kingma and Ba, 2017), using the default values in Keras version 2.2.4 (Chollet, 2015).

Specifically, from the training set a single vertex pair, representing a positive or negative training sample, is fed to the first layer of the DLP model known as the input layer. This layer is connected to the projection or node embedding layer in which numerical vector representations, i.e. embeddings, will be learned during training. As with other layers in a neural network, the projection layer is a $V \times D$ weight matrix, where V is the total number of vertices in the input graph and D is the embedding dimension, i.e., the number of neurons in the projection layer. In order to select the correct row of the $V \times D$ weight matrix, corresponding to one of the input vertices, a one-hot-encoded vector is constructed from each input vertex.

The flexibility of our DLP model can be further exploited by initializing the vertex embedding layer with embeddings learned by any other NRL-based LP method. As is common with neural networks, a single layer consists of a matrix of weights that are updated when the model is shown positive and negative samples. It is possible to initialize this weight matrix with a specific pre-calculated embedding matrix with dimensions $V \times D$.

The subsequent learning and prediction process remains the same as with the standard DLP model. Throughout this work, we have used the embedding from another method, DeepWalk, as initialization, using the same edge embedding operator that was selected for the original DeepWalk model (weighted-l_2). DeepWalk was chosen based on its high average performance on retrieving benchmark drug targets and dependencies as compared to other state-of-the-art LP methods.

Generating the input graph for LP

To construct the heterogeneous graph used as input to perform LP, the gene-gene interaction network from STRING was integrated with the DepMap data by adding cell line vertices and connecting those to genes from STRING that display a strong dependency. The STRING scaffold was downloaded from the GitHub repository of Yue et al. https://github.com/xiangyue9607/BioNEV/tree/master/data/STRING_PPI (Yue, 2019; Yue et al., 2020). All functional interactions contained in the largest connected component of the STRING network were used, consisting of 14,633 vertices and 350,832 interactions. The original network used Ensembl protein IDs as vertex labels. These were converted to HGNC symbols v75 using data available on the official HUGO Gene Nomenclature Committee website (Braschi et al., 2019). Ensembl protein IDs for which no suitable HGNC symbol could be retrieved were dropped from the functional interaction network. In total, there are 12,853 genes overlapping between the functional interactome and the DepMap LOF screen.

To test the robustness of each method against the use of a different gene-gene interaction scaffold we also tested the integration of DepMap data with the Reactome FI 2020 interaction network (Wu et al., 2010), consisting of 13,785 Nodes and 259,009 interactions.

A different heterogeneous graph was constructed for each cancer type separately and one was constructed combining all cancer types together (i.e., the pan cancer setting). As each cancer type has a different number of profiled cell lines, the final size of the heterogeneous graph differs per cancer type, screening technology, and used gene-gene interaction scaffold (see Table S1).

Training data

To learn the distinction between the presence and absence of an edge in the input graph, LP methods require a training set of positive and negative interactions, i.e. the problem is formulated as a binary classification. As positive samples, all edges present in the largest connected component of the heterogeneous input training graph were used. Given the heterogeneous nature of the input graph, the presence of an edge corresponds to either the presence of an interaction in the gene-gene interaction network or to an interaction in the cancer specific cell line-gene dependency network (or pan cancer network in case all cell lines were combined). Negative samples correspond to randomly sampled vertex pairs for which no interaction occurs in the original network, using a ratio of five negatives per positive sample to capture the sparseness of networks while simultaneously restricting computational burden as compared to sampling according to the real imbalance.

For the cancer dependency network, for each cell line, positive and negative edges were obtained by selecting respectively strong dependencies (DepMap score $< T_{pos}$) and weak dependencies (DepMap score $> T_{neg}$), see Table S5 for the specific thresholds used for RNAi and CRISPR. Here we used a slightly lower positive-negative ratio of one to three, to prevent the model from being too conservative. Focusing only on weak and strong dependencies ensures that the cell line-gene interactions seen during training are reliable representatives of true positive and true negative dependencies (Tsherniak et al., 2017).

Note that for the pan cancer setting we combined all training and validation sets from each of the seven cancer types into a single training and validation set. Consequently, the performance was evaluated using the exact same test sets as was done for each cancer type separately.

Benchmarking LP methods with EvalNE

The benchmark was performed using the Python package EvalNE (Mara et al., 2019). This framework allows comparing different NRL methods for various downstream tasks, including LP, based on a separate held-out test set.

To perform benchmarking, the heterogeneous input graph is divided into three separate datasets: a training, a validation, and a test set. The first two are used for training the LP methods, i.e. learning the correct vertex representation and avoiding overfitting; while the held-out test set is used to assess the generalization performance. For each combination of method and binary operator, which is used to combine the vertex representation of a particular interaction into an edge representation necessary for LP, the benchmark was evaluated on three differently sampled training, validation, and test sets, as was described by Mara et al. (2020). Since we are dealing with two different interaction types in our heterogeneous input graph, namely gene-gene interactions and cell line-gene, separate training, validation, and test sets were constructed for each interaction type and subsequently combined. The standard split of 80–20 was used to distinguish a training and test set. For validation, 20% of the training set was used. Table S7 provides a brief description of the 13 LP methods used in our benchmarking (i.e., Jaccard coefficient, preferential attachment, resource-allocation-index (Zhou et al., 2009), academic-adar-index, common neighbors, all baselines, AROPE (Zhang et al., 2018), VERSE (Tsitsulin et al., 2018), LINE (Tang et al., 2015), DeepWalk (Perozzi et al., 2014), node2vec (Grover and Leskovec, 2016, p. 2) and GraRep (Cao et al., 2015)). All methods were used with their default parameters except for the embedding size, which was set to 128 for all methods (see Table S6). All methods were tested using each of the same four binary operators used by Mara et al., (2020), namely weighted-$l1$, weighted-$l2$, hadamard and average. Only the performance for the highest scoring operator is mentioned for each method. The specific operators for each method and each run are stored in the files returned by EvalNE.

Predicting unseen genes

As LP methods do not allow predicting interactions for vertices not seen during training, each vertex needs to be seen at least once during training for the method to learn a representation for that vertex. However, as there are two types of interactions in the heterogeneous graph, a gene representation can be learned solely from gene–gene interactions and then used to predict cell line–gene interactions. This allows identifying dependencies in cell lines from genes that were not included in the LOF experiment. To mimic such a setting where not all genes were included in the LOF screening of a certain cancer type, 20% of all vertices representing genes that were positive dependencies in at least one cell line were randomly removed before training. Consequently, no dependency information was used for these genes in the heterogeneous graph during training (no cell line-gene interactions are present for the considered genes in the graph), and each model will have to predict the interactions between those genes and the cell lines purely based on the representation constructed during training of the gene-gene interactions in the heterogeneous graph. All LP methods were trained in the same way as was done in the benchmark.

Analyzing the molecular mechanism of dependencies

To infer a subnetwork reflecting the processes through which a gene is affecting a cell line, we used the cell line-gene as well as gene-gene interaction probabilities of the DLP-DeepWalk model.

Figure S5B shows how, for a given query gene, each gene in the network is scored, based on both a cell line-gene and a gene-gene component. The gene-gene component is obtained by simply taking the predicted interaction probabilities between the query gene and all other genes in the network, denoted $v_{functional}$. The cell line-gene component relies on the cell line-gene probabilities to connect the query gene to the other genes in the network. Per cell line, each gene in the network is scored by multiplying its cell line-gene probabilities with the cell line-gene probability of the query gene, resulting in a high score when both genes have a high cell line-gene probability with that cell line. Averaging this score across all cell lines we obtain a number that lies in [0, 1], similar to the gene-gene component

This cell line component ($v_{cell\ line}$) can be obtained by a simple matrix multiplication:

$$v_{cell\ line} = P\widehat{c}$$

Where $\widehat{c}$ is a $c \times 1$ vector containing the predicted probabilities between the query gene and the c cell lines, rescaled by the number of cell lines, and P is a $g \times c$ matrix containing the predicted probabilities between the c cell lines and the g genes in the network. The final score is then obtained by simply adding $v_{cell\ line}$ and $v_{functional}$. A high total score implies that a gene is close to the query gene in the functional interaction network and exhibits a dependency in the same cell lines as the query gene. Hence, it can be expected that many highly ranked genes are involved in the same molecular processes.

To keep the resulting subnetwork as specific as possible, while still including enough genes, only the largest connected component of the resulting top 20 genes was selected to make up the final subnetwork that served as a proxy of the processes triggered by interfering with the query gene. Only edges from the original STRING network are displayed as this shows the influence of the cell line-gene component. Using only the gene-gene component would result in recovering mainly direct neighbours.

Finally, to annotate the subnetwork, a hypergeometric test was performed between all the genes located in the subnetwork and genes from pathways as defined by Reactome v75, obtained from MSigDB (Liberzon et al., 2011; Subramanian et al., 2005).

QUANTIFICATION AND STATISTICAL ANALYSIS

Evaluation metrics

Although traditionally the performance of LP methods is often assessed in terms of the Area Under the Receiver Operating Characteristic (AUROC), we opted for the Average Precision (AP) metric. The AP is a weighted mean of precisions obtained at every probability threshold, where the weight is equal to the increase in recall obtained at the nextprobability threshold relative to the previous threshold (Buitinck et al., 2013). For a large number of samples, it closely approximates the area under the Precision Recall (PR) curve. The AP is more suited to our setting, as we deal with a largely imbalanced datasets in which TN interactions vastly outnumber TP interactions (Saito and Rehmsmeier, 2015). When plotting the True Positive Rate (TPR) versus the False Positive Rate (FPR) as it is done with the AUROC, the true performance might be overestimated because the presence of TNs dominate the results. Due to the imbalanced negative to positive ratio, by chance, there will be an increase in TNs and hence a decrease in FPR resulting in an underestimation of the FPR. AP does not consider the TN and thus is better at representing the performance with imbalanced data. Additionally, in practice we are interested in correctly prioritizing the rare cases of true dependencies or TP, a performance that is better captured by the AP. The formula used to calculate the AP is provided below:

$$AP = \sum_{n}(R_n - R_{n-1}) * P_n$$

with P_n and R_n represent the precision and recall at threshold n, respectively.

AP of LP methods in predicting drug targets

For the benchmark, we selected known cell line-specific drug targets. Not all molecules or drugs are equally effective even though they hit the same target. Because we are focusing on prioritizing genes (dependencies) that could be potential drug targets, we are interested in knowing whether a cell line is sensitive, if a drug is applied that successfully affects the target. For this reason, we consider a cell line sensitive, if at least one drug hitting that target shows a large effect in that cell line (sensitivity < −2). Using this definition, we then compare the number of sensitive cell lines in which a target was retrieved in the top 100 or missed by either DLP-DeepWalk or the original RNAi screening score.

To assess the performance of the LP methods in predicting targets of effective drugs, we took for each LP method and the original screening data the top ranked genes per cell line of a particular cancer type. These corresponded to respectively cell line-gene interactions that received the highest probabilities to interact according to the LP probabilities or the strongest original dependency scores according to the screening experiment. We subsequently assessed to what extent these top ranked genes corresponded to the benchmark targets described above. The number of top ranked genes was chosen to mimic a routine protocol consisting of a series of subsequent, increasingly focused LOF screens. From an initial screen, the top prioritized genes are subjected to validation screenings with more RNAi constructs to gradually limit off-target effects. We take as a representative size of the top rank list 100 predictions. This threshold genes is low enough to allow for further downstream experiments, but also high enough to enable meaningful statistical analysis. We observed that considering the performance of the top 100 genes is representative of a method's performance at higher top K, i.e., the relative ordering of the methods in the performance assessment changes very little when selecting more than 100 genes (Figure S6).

As a baseline, we also compared the performance of all LP methods and the original DepMap data in correctly inferring targets of sensitive drugs with a prioritization that could be obtained by randomly picking 100 genes from the functional interaction network and assessing to what extent these contained true drug targets. Two types of sampling strategies were employed: picking 100 genes randomly 1) from a uniform distribution or 2) from a degree-based distribution derived from the gene-gene interaction network.

The expected number of targets were calculated using a hypergeometric Probability Mass Function (PMF) in case of a uniform sampling and using permutation tests in case of a degree-based sampling strategy (10,000 permutations). The expected number of targets retrieved by each of the sampling strategies varies per cell line with the degree-based sampling often retrieving the highest number of expected targets. Hence, the degree-based sampling results in the most conservative baseline.

In case of the degree-based sampling strategy, the PMF was derived by counting the number of retrieved drug targets in 10,000 permutations. The p value to assess that a certain number of drug targets x is observed in the top 100 of a cell line by chance is subsequently derived from the following function:

$$\text{p value} = 1 - CDF(X) = 1 - P(X < x)$$

Here, X is the drug target retrieval variable, and $CDF(X)$ is the probability that X will take a value less than x, i.e. the Cumulative Distribution Function of X. Where x is the number of drug targets retrieved in the top 100, by any LP method or the original RNAi data. However, we are interested in the probability that a degree-based random model retrieves more drug targets than any LP method or DepMap, which is why we subtract the CDF from 1.

Performance of retrieving benchmark targets

To assess the performance of the LP methods in retrieving benchmark targets, the following statistics were used: 1) the average percentage of benchmark targets recovered for a certain cancer type across cell lines and 2) the degree to which LP methods retrieve significantly more benchmark targets than what can be obtained with a ranking based on the original DepMap scores.

To calculate both statistics, we ranked, for each cell line, the genes according to the probabilities that were assigned by a certain LP method. We also ranked for each cell line the genes according to their DepMap score in that cell line. For both rankings and per cell line, we considered the top 100 genes and calculated the percentage of benchmark targets in the top 100 genes of either ranking. These numbers were used as entries in a vector with N dimensions (number of cell lines in a cancer type) representing either the results of the LP method or the original RNAi screening score ranking.

To assess whether LP methods retrieve significantly more benchmark targets than what can be obtained with a ranking based on the original DepMap scores, both vectors were compared using the one-sided Wilcoxon signed-rank (WSR) test since we are dealing with paired samples.