Reverse tracking from drug-induced transcriptomes through multilayer molecular networks reveals hidden drug targets ☆

Gwangmin Kim, Doheon Lee *

Korea Advanced Institute of Science and Technology, Daehak-ro 291, Daejeon, 34141, Republic of Korea

ARTICLE INFO

ABSTRACT

Identifying molecular targets of a drug is an essential process for drug discovery and development. The recent *in-silico* approaches are usually based on the structure information of chemicals and proteins. However, 3D structure information is hard to obtain and machine-learning methods using 2D structure suffer from data imbalance problem. Here, we present a reverse tracking method from genes to target proteins using drug-perturbed gene transcriptional profiles and multilayer molecular networks. We scored how well the protein explains gene expression changes perturbed by a drug. We validated the protein scores of our method in predicting known targets of drugs. Our method performs better than other methods using the gene transcriptional profiles and shows the ability to suggest the molecular mechanism of drugs. Furthermore, our method has the potential to predict targets for objects that do not have rigid structural information, such as coronavirus.

1. Introduction

Identifying molecular targets of drugs is an essential process for drug development regardless of discovery strategies. Most drugs are derived from phenotypic screening methods so the molecular targets are absent [1,2]. Target-based screening is another method that starts with identified targets for a given therapeutic effect [3]. However, since unintended targets are likely to exist, there have been many reports of side effects [4,5]. There are several assay-based experiments to identify targets of drugs, such as biochemical affinity purification or knocking out genes [6]. However, these approaches are largely impractical, as the experimental processes are time and labor-consuming [7]. Therefore, computational approaches to predict targets for a given drug are necessary.

Most recent computational approaches are focused on using molecular structure information. Molecular docking is one of the representative methods which simulates interactions between a drug and its possible matching proteins [8]. This method shows high accuracy and interpretability but requires significant computation resources and often unavailable information, such as 3D structures of drugs and proteins. The machine-learning-based approach is also promising in predicting drug–target interactions. This approach shows relatively less computation time, and prior knowledge of 3D structures of chemicals, drugs, and proteins is unnecessary. However, due to the lack of

non-interaction information, the drug–target prediction task only has positively-labeled data. Thus, the performance of these methods relies on how the negative samples are labeled [9,10].

An approach using transcriptome profiling was introduced due to the many consortiums such as CMap LINCS [11], CCLE [12], and GDSC [13]. This approach compares the similarity between gene expression profiles from drugs [14] so, this method can only be used on drugs with comprehensive reference gene expression profiles [15] and is not appropriate to find *de-novo* targets. Network-based approaches such as DeMAND [16] and ProTINA [17] have been proposed to overcome the above problem. They constructed a bipartite network by combining gene-regulatory network(GRN) and protein–protein interaction(PPI) network to score target proteins. However, the network has a loss of information because every protein can directly regulate the associated genes and interactions between proteins are easily ignored.

Here, we suggest a reverse tracking method that measures the propagations of a given drug-induced gene expression in multilayer molecular networks to the target proteins. Our method uses the multilayer network to separate the role of transcription factors (TFs) and the other proteins. We first scored transcription factors using gene-regulatory interactions and additional drug perturbation data. Then, we scored every protein using PPI network and the scores from the previous step. The overall schema is depicted in Fig. 1. We show
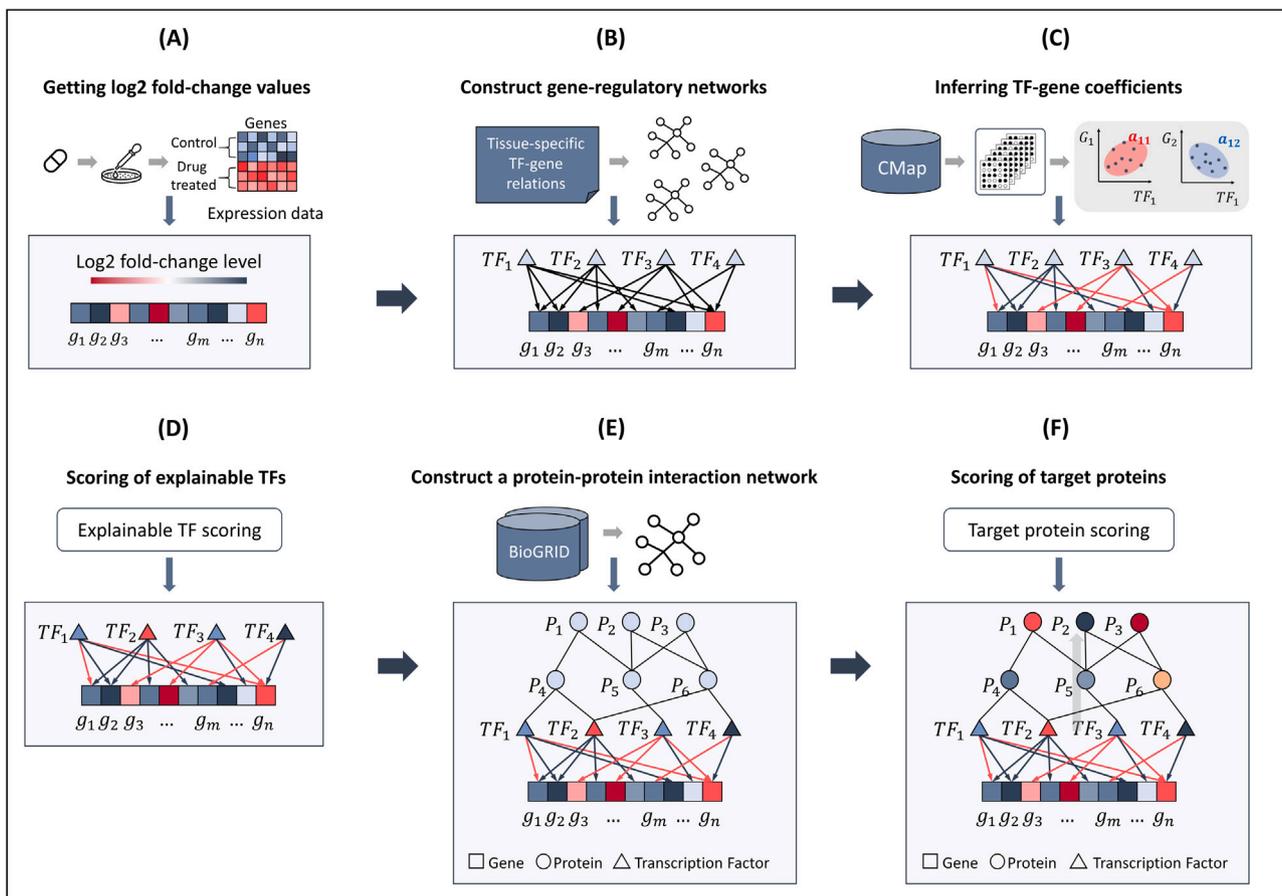
Fig. 1. Method overview. (A) From the expression data of a drug, get log_2 fold-change value of genes. **(B)** Construct a gene-regulatory interaction network using tissue-specific relations. **(C)** Using CMap LINCS data, calculate correlation coefficients of gene expressions between transcription factors and genes obtained from (B). From CMap LINCS, only signatures in the same cell or tissue type with the original dataset in (A) were selected. **(D)** Based on the suggested scoring method, transcription factors in the network were scored depending on how they are correlated with their coefficients and expression changes of their target genes. **(E)** Construct a protein–protein interaction network by integrating three types of public databases and connecting all proteins, including transcription factors. **(F)** Propagate transcription factor scores through the interactions in the PPI network.

that our method performs better in predicting drug targets compared to the previous state-of-the-art methods with three types of datasets: anticancer drugs, genotoxic drugs, and the GPCR targeting agents. We also suggest a potential molecular mechanism of drugs supported by literature evidence. Furthermore, we show that our method could be used to predict targets even when not having reliable structural information using the SARS-CoV-2 case as an example.

2. Materials and methods

2.1. Preprocessing of data

2.1.1. Preprocessing of input expression data

We used drug perturbation datasets from NCI-DREAM synergy challenge [18], a genotoxicity study [19], and a GPCR targeting agents study [20]. We obtained all datasets from Gene Expression Omnibus (GEO) database [21] (accession number: GSE51068, GSE28878, and GSE40017). Datasets from NCI-DREAM synergy challenge and genotoxicity study were also used in the previous work, ProTINA [17]. The dataset from GPCR targeting agents study was used to show an advantage of the transcriptome-based method. All datasets were quantile normalized and transformed into $log2$ scale. Datasets from NCI-DREAM synergy challenge and the genotoxicity study were preprocessed using refine.bio [22]. The dataset from the GPCR study was already normalized and into log_2 transformed. We only selected genes with changes that are statistically significant (adjusted p-values < 0.05)

2.1.2. Preprocessing of CMap data

We downloaded CMap LINCS [23] data, which contains the largest compound-induced gene expression profiles with the number of compounds and cell-lines. We used level 5 processed data which includes replicate-collapsed z-score profiles of 12,328 gene expression levels. Only the signatures that satisfied CMap quality control criteria [11] were selected.

2.2. Construction of multilayer network

We used the PPI network to utilize the relationships between drug targets and their interacting proteins and GRN to associate the TFs related to gene expression changes. BioGRID [24], KEGG [25], and EndoNet [26] were used to build the PPI network. For BioGRID, we downloaded a tab-separated text file from release 4.4.198 from the website (https://downloads.thebiogrid.org/BioGRID). We used only human PPIs whose taxonomy ID is '9606' (meaning Homo sapiens) and took 465,537 interactions between 18,897 entities from BioGRID. For KEGG, we downloaded text files in XML format using KEGG's official REST API (http://rest.kegg.jp/). We selected relations labeled 'PPrel' (meaning protein–protein interaction) and converted KEGG ID to Entrez ID using KEGG API. We took 47,550 interactions between 4,129 entities from KEGG. For EndoNet, we downloaded a text file in XML format (EndoNet.xgmml) from the EndoNet website (http://endonet.bioinf.med.uni-goettingen.de). We took 1,160 interactions between 781 entities from EndoNet by obtaining only relations whose both nodes have matched gene symbols. We used a mapping file between gene symbol
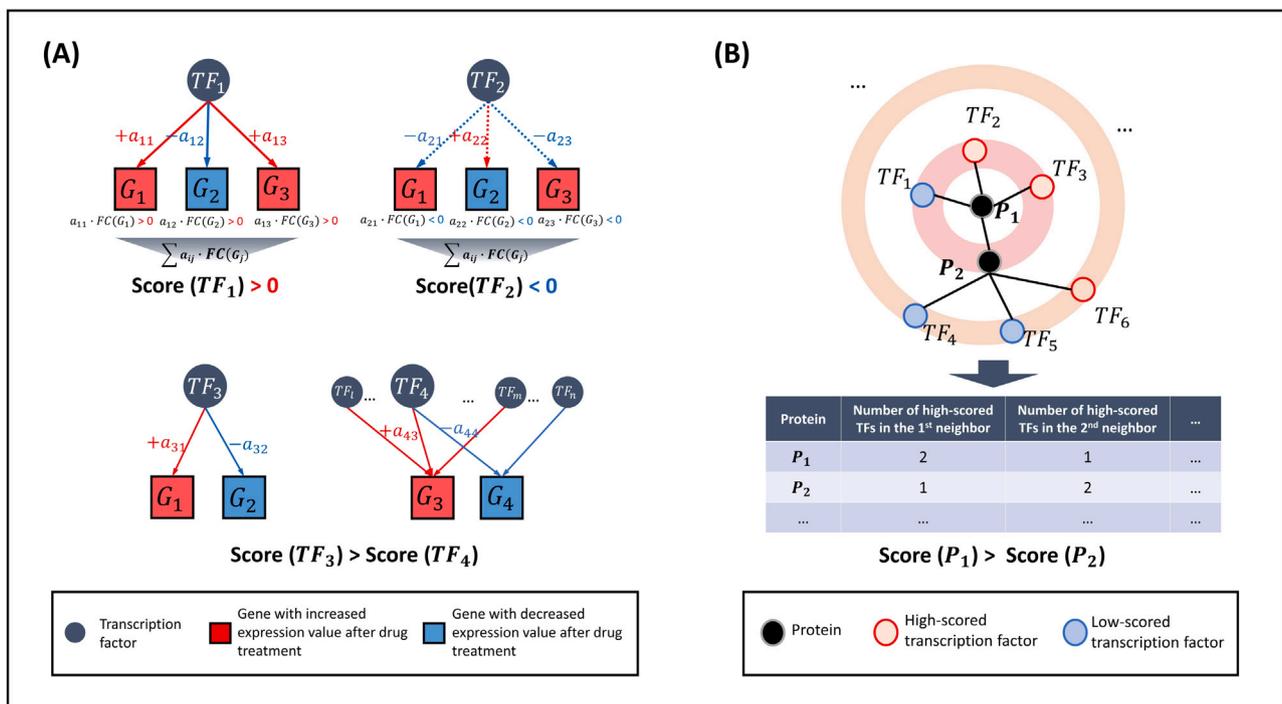
Fig. 2. The concept of scoring method. (A) shows how transcription factors are scored. a_{ij} represent a correlation coefficient value between transcription factor i and gene j and $FC(G_j)$ is the fold-change value of gene j. The red-colored arrow means a positive correlation; the blue-colored means a negative correlation between them; the dotted arrow means that the given pair is not correlated. At the top, the TF_1 is correlated with all its targeting genes while TF_2 is not correlated at all. The summation of each score of TF_1 becomes positive and is more explainable than TF_2. In the bottom, both TF_3 and TF_4 explain the gene expression changes well, but G_3 and G_4 are already regulated by other TFs. So, TF_3 become more explainable than TF_4. **(B)** shows how target proteins are scored. Protein(P_i)s and transcription factor(TF_j)s are located in the PPI network. P_1 is more closely located with the high-scored transcription factors, so the score also is higher than P_2.

and Entrez gene ID downloaded from the NCBI website (https://ftp.ncbi.nih.gov/gene/DATA/gene-info.gz). We converted entities represented as gene symbols into Entrez gene ID, and we used a union set of all PPIs obtained from the above databases. In total, the PPI network contains 514,247 interactions between 19,392 entities.

We acquired tissue-specific gene regulation data from the study [27] which provides TF-gene interactions of 41 types of tissues and cell-lines to build the GRN. Among 41 types of tissues and cell-lines, we only used TF-gene interactions from lymphoblastoid cell-line (as NCI-DREAM synergy challenge data uses OCI-LY3 cell-line), liver (for the genotoxicity study using HepG2 cell-line) and colon (for the GPCR study using SW480 cell-line). For the analysis, only TF-gene interactions with a prior score are 1 were selected. The prior score of 1 indicates that the interaction is already known and not inferred. Finally, interactions whose TFs were not in the nodes of constructed PPI network were excluded to combine two types of molecular networks. In total, we obtained tissue-specific TF-gene interactions for human lymphoma, hepatocellular carcinoma cell, and colon carcinoma cell, each with 7,849 interactions between 601 TFs and 2,238 genes, 2,765 interactions between 467 TFs and 1,204 genes, and 12,593 interactions between 583 TFs and 4,162 genes respectively.

2.3. Scoring method

Our scoring method considers the change in gene expression profile as a result of the perturbation of drug targets. TFs make gene regulation but the other proteins do not make and most known drug targets are not TFs. (See **Supplementary Figure 1**) So, we separated the roles of TFs and the other proteins(targets) in drug perturbation and scored them differently: TFs are scored only through the protein–gene interactions and proteins(including targets) are scored through the PPIs and from the scores of the TFs. This scores proteins based on how well the protein explains the outcome of perturbed gene expression through the constructed multilayer molecular network. To capture these perturbations

by drug targets, our scoring method consists of two steps (Fig. 2). First is getting TFs which can explain the change of gene expression well. The second is getting candidate target proteins that are closely located in TFs with high scores in the previous step.

2.3.1. Scoring of transcription factors

To know which TF can potentially explain the cause of perturbation in an input expression data well, we designed a score function to measure the explainability of a particular TF.

$$S(TF_i) = \sum_{j=1}^{n} \frac{r_{TF_i \cdot g_j} \cdot log_2 FC_{g_j}}{m_{g_j}} \qquad (1)$$

In the Eq. (1), g_j means targeted genes of TF_i, and $r_{TF_i \cdot g_j}$ means correlation coefficient between TF_i and g_j. Fold-change value of expression of g_j represent FC_{g_j}, and m_{g_j} means the number of transcription factors which regulate g_j. The higher the score of our score function is, the more explainable the transcription factor is to gene expression. Fig. 2(A) shows the process. Each TF is assigned a score based on how much each targeting gene's log_2 fold-change value matches the correlation coefficient calculated from CMap LINCS. If the log_2 fold-change value of a targeted gene and the correlation between the TF and the gene share the same direction (both positive or both negative), it can be reasonably assumed that the TF explains the gene expression [17]. We also consider the targeted gene itself, as the gene which is already regulated by many other TFs [28] will have lower explanatory power [29]. Therefore, we set a normalization term in our scoring function to reflect this assumption.

2.3.2. Correlation coefficients between a transcription factor and gene

We calculated the Pearson correlation coefficients by using the method from the study of Zaborowski et al. [30], which used expression values of genes and TFs from microarray data. We applied the same

method to CMap LINCS data and used only signatures from each of cell-line to get tissue-specific correlation coefficients. For example, when using input expression data from GSE51068, we used signatures that are only from the OCI-LY3 cell-line. We obtained 277 signatures for the OCI-LY3 cell-line, 4,259 signatures for the HepG2 cell-line, and 46 signatures for the SW480 cell-line, and each cell-line is matched to its corresponding input expression data. Each TF and gene in a relation is represented as a vector of gene expression values with a size of the number of signatures, and we calculated the Pearson correlation coefficients between the two vectors.

2.3.3. Scoring of target proteins

To score target proteins, we considered every protein in the PPI network as candidate targets of a drug. Each score of a candidate target protein is determined based on the previously calculated scores of TFs as well as its distance to the candidate protein in the PPI network (Fig. 2 **(B)**). For example, if a candidate target protein has highly scored TFs as its neighbor, the target protein also gives a higher score. We used the shortest path length in the PPI network as the distance between each protein and TF [31]. We designed a candidate target protein score like below.

$$S(P_i) = \sum_{r=1}^{d} \sum_{j=1}^{n_{TF,r}} \frac{S(TF_{r,j})}{r} \tag{2}$$

In the Eq. (2), $TF_{r,j}$ means transcription factor having r shortest path length from a protein P_i and $S(TF_j)$ means TF_j's score calculated in Eq. (1). r means the shortest path length from protein P_i and TF_j in the network and $n_{TF,r}$ means the number of transcription factors with r distance from a protein P_i. If we do not set the maximum distance, every entity in the network can be reached. Therefore, we limited the maximum length from a protein to TFs as 3 for the calculation because it is enough to reach almost every TF. (See **Supplementary Figure 2, 3.**)

2.4. Performance evaluation

We tested our method's performance in predicting already known drug targets with other methods. We used the same procedure described in ProTINA which compared high-scored proteins with drugs' already known targets by measuring the area under the receiver operating characteristic curve (AUROC) of each drug with other methods. For each method and drug treatment, we obtained a score of every protein. Then, we generated a ranked list in descending order of the protein scores from our method, descending order of the target protein scores from ProTINA, ascending order of the p-values from DeMAND, and descending order of log_2 fold-change gene expression value from differential expression analysis (DE). We used a binary vector for the label vector with 1 if the corresponding protein is the known target or 0 for each drug. Using each of the ranked list and the label vector, we calculated AUROC value for each drug. The higher the score of the already known target protein is, the higher the AUROC value of a given drug is. There exist replicates in each input drug perturbation profile, so we averaged AUROC values from the replicates.

We used three different drug perturbation datasets which contain different types of drugs to evaluate the overall performance of each method. The AUROC values are made as many as the number of drugs in each dataset. Then we plotted box plots to show distributions of AUROC values from every drug for each dataset. For the implementation of DeMAND and ProTINA, we used R source code from the original publication. In the case of DE analysis, we used the log_2 fold-changes of genes. We used the same tissue-specific multilayer networks as those in our method. We used the DrugBank[32] dataset for known targets of drugs. Only drugs which are used in our analysis were selected. In the case of the GSE40017 dataset, we were able to find targets from the literature [33–36]. The drugs and their targets used in our study are summarized in **Supplementary Table 1**.

For testing our scoring function in different conditions, we also compared our method with scoring methods using different approaches in calculating correlation coefficients and different neighboring distances in predicting known drug targets. We adopted random values and multivariate linear regression (MLR) for correlation coefficients between TFs and genes. The random coefficient values were each sampled 30 times from the distribution of each corresponding dataset, and we calculated AUROC of each drug using coefficients from the result of each random sample. Then, we averaged AUROC values of the 30 results for comparison. In the case of MLR, coefficients β_{ij} from (Eq. (3)) are used. For expression value of y_i of gene i, x_{ij} is expression value of TF j which regulates gene i. We also tested our scoring function by changing the maximum neighboring distances between proteins and TFs

$$y_i = \beta_{i0} + \beta_{i1}x_{i1} + \beta_{i2}x_{i2} + \cdots + \beta_{ij}x_{ij} + \epsilon \tag{3}$$

2.5. Data processing for case study

We processed every necessary data in the same pipeline that we mentioned above. Input expression data was obtained from GEO with the accession number GSE182297. The GSE182297 dataset contains gene-expression data collected from brain tissues of both healthy individuals and patients with coronavirus. We acquired tissue-specific TF-gene relations from the study [27] labeled as brain basal ganglia, brain cerebellum, and brain other. We obtained 66,035 interactions between 641 TFs and 9,542 genes. For TF-gene coefficients, we only used signatures from LN299 cell-line in CMap, and 4 signatures were selected. Following the same steps of the scoring function, every protein was calculated and measured AUROC value with a ranked list in descending order of the scores.

3. Results

3.1. Performance in predicting drug targets

We tested our method in predicting drug targets in three different sets of expression data: data from NCI-DREAM drug synergy study with anticancer drug agents in OCI-LY3 cell-line, data from the genotoxicity study in the HepG2 cell line, and data from the GPCR targeting study in the SW480 cell line. As mentioned in the method section, we measured the AUROC score for each drug using a ranked list of protein scores. Fig. 3 shows the distribution of the AUROC score of every drug used in each dataset. DE analysis shows average AUROC scores of 0.306, 0.440, and 0.501 in GSE51068, GSE28878, and GSE40017, respectively. DeMAND shows average AUROC scores of 0.520, 0.566, and 0.349. ProTINA shows average AUROC scores of 0.599, 0.495, and 0.324. Our method shows average AUROC scores of 0.809, 0.636, and 0.693. DE analysis shows the poorest AUROC values in the datasets of GSE51068 and GSE28878. The other previous methods show better results than DE analysis in the above dataset but also show worse results in the dataset of GSE40017. However, our method shows the highest performance regardless of the datasets. All AUROC scores from different datasets are summarized in **Supplementary Table 2.**

3.2. Performance in different scoring conditions

To evaluate our scoring function, we tested the scoring function in different conditions. We compared two other approaches in calculating coefficients to use in our scoring of TFs and genes: coefficients from random values following the same distribution of values from our method and coefficients from MLR.

Fig. 4**(A)** shows the results of using other measurements to get coefficients in the scoring of TFs. The average AUROC values using random coefficients are 0.427, 0.575, and 0.480 in GSE51068, GSE28878, and GSE40017, respectively. In the case of using MLR, the values are 0.772,
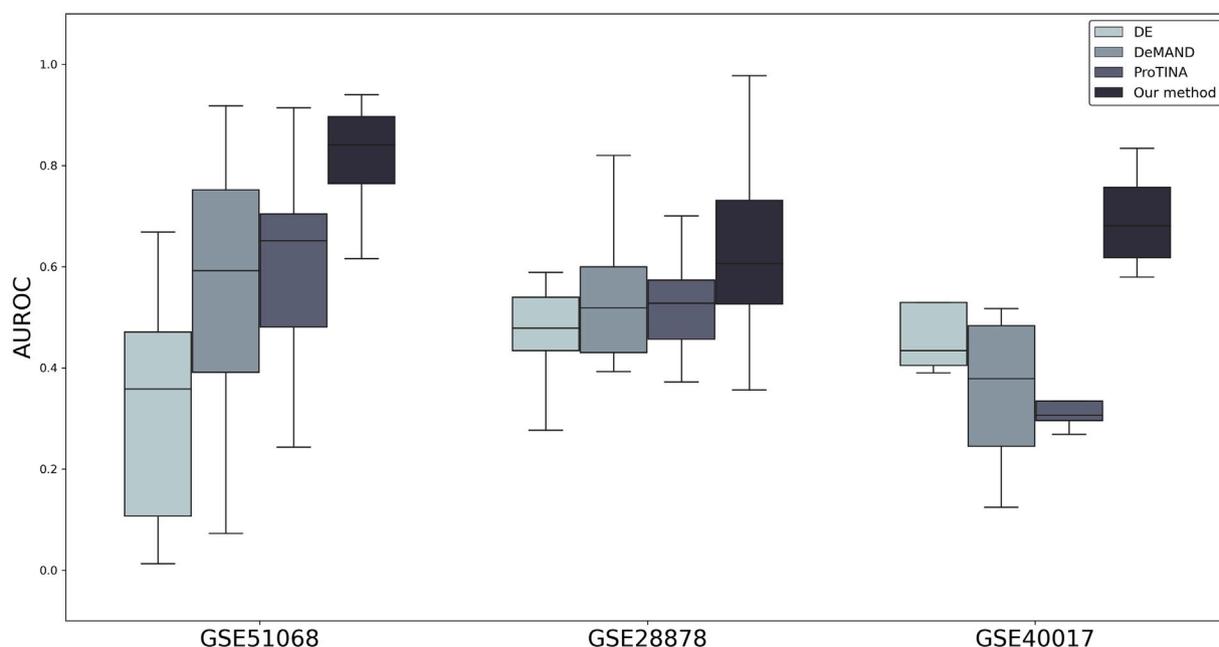
Fig. 3. Performance comparison with other methods in predicting drug targets. AUROC value of each drug is calculated in the three different datasets and other previous methods. The distribution of all AUROC values from each method is plotted as the format of the box-plot.
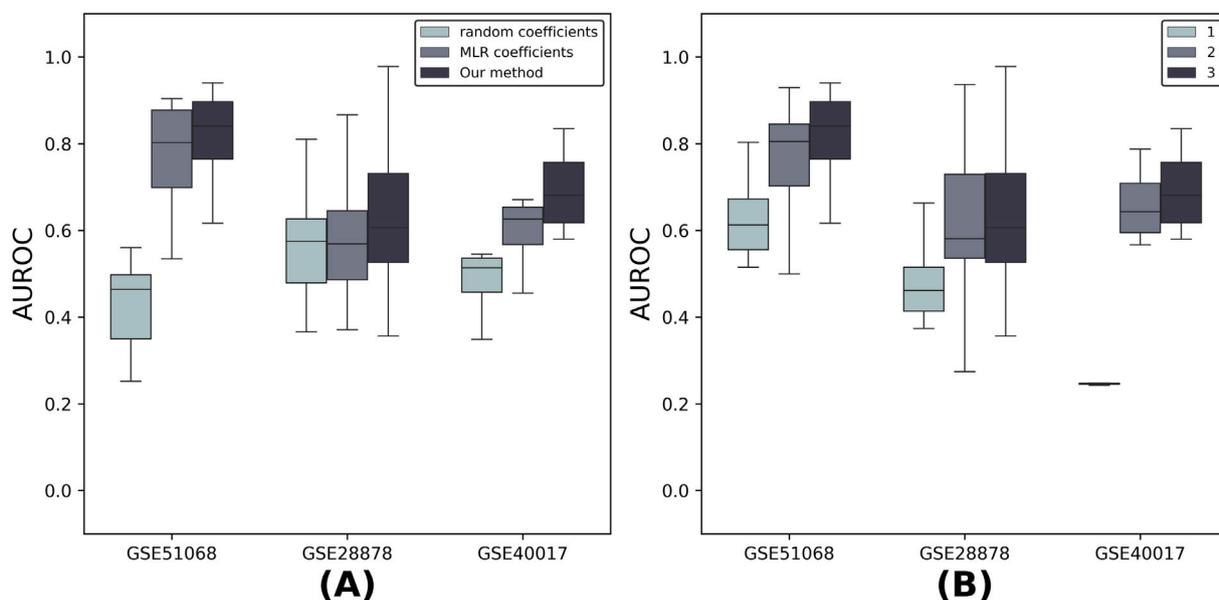

Fig. 4. AUROC values from different scoring conditions in our method. (A) shows the variation of performance depending on the way of getting correlation coefficients between TFs and genes when scoring of TFs. **(B)** shows the variation of performance depending on the neighboring distance between target proteins and TFs when scoring target proteins. Our method used a distance as 3.

0.581, and 0.594 for each dataset. For machine learning-based methods such as MLR to work well, the quantity and quality of the dataset are very important. By using CMap data, a sufficient number of expression data which is more than the number of variables in the equations consisting of the relationship between TFs and genes, was gathered. However, the quality of the gene expression profiles from CMap was relatively low. (See **Supplementary Figure 4, 5.**) Therefore, the performance of a relatively simple method such as using Pearson correlation between one-to-one (TF and gene) showed better performance than models using various input variables such as MLR. Fig. 4**(B)** shows the results of using different neighboring distances between proteins and TFs. When we used just TFs in the first neighbors of each protein, we were able to get average AUROC values of 0.634, 0.461, and 0.245 in

each dataset. The value increases to 0.759, 0.619, and 0.660 in each dataset when using TFs within the second neighbors, but still lower than the original method. All AUROC scores from different datasets and methods are summarized in **Supplementary Table 3**.

3.3. High-scored proteins with known mechanism of action of drugs

For further analysis, we ranked the scores of mechanism-related proteins to demonstrate our method. Some anticancer drugs in the dataset of the NCI-DREAM challenge (GSE51068) are known to relate to canonical p53 DNA damage pathway and eventually disrupt mitosis. For example, camptothecin (targets *TOP1*), doxorubicin (targets *TOP2A*), and etoposide (targets *TOP2A*) have DNA topoisomerases as
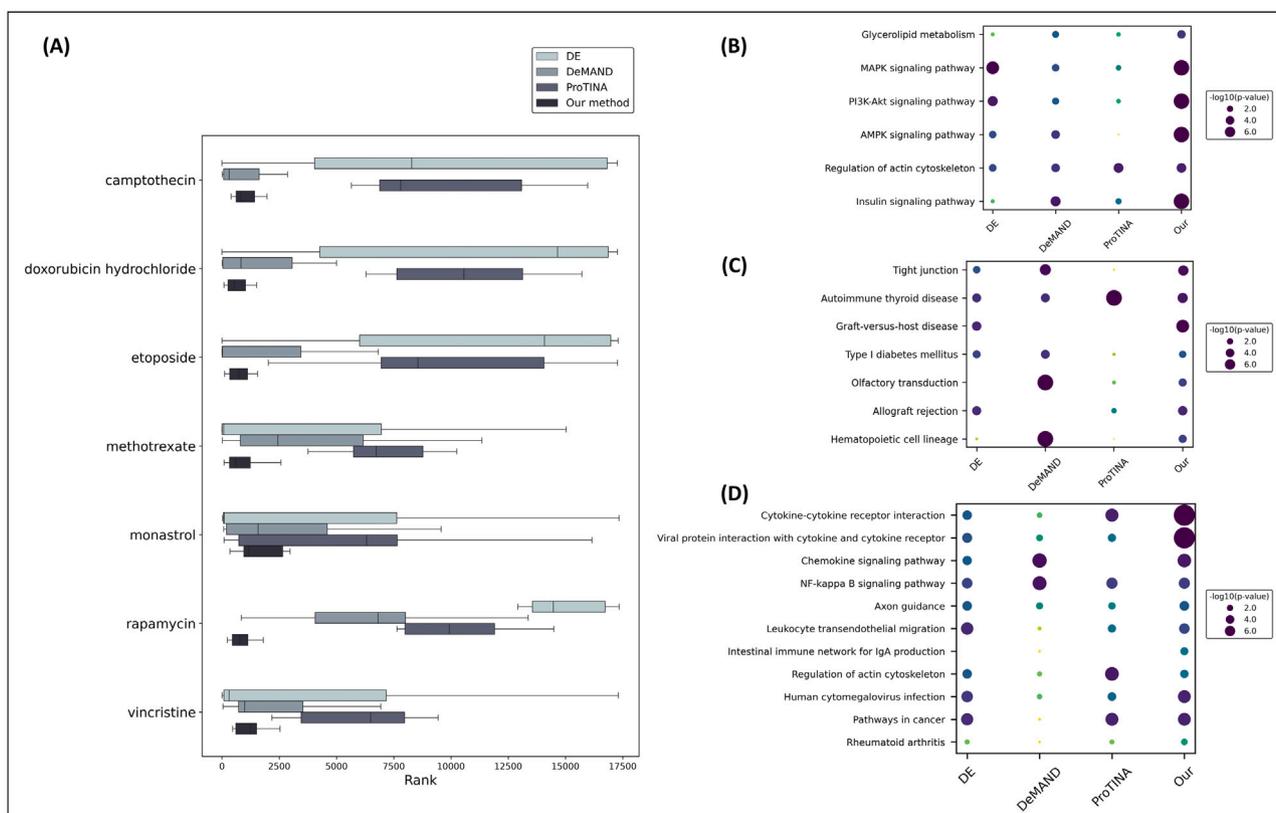
Fig. 5. High-scored proteins are associated with already known mechanism-related proteins. (A) The distribution of ranks of already reported proteins that are related to the mechanism of action of anticancer drugs in the dataset of NCI-DREAM challenge. **(B)** *P*-values of rapamycin-related KEGG pathways in each method. **(C)** *P*-values of azathioprine-related KEGG pathways in each method. **(D)** *P*-values of SDF-1 related KEGG pathways in each method.

their target proteins, and these proteins activate p53 protein [37–39]. Other drugs such as rapamycin [40], vincristine [41], trichostatin a [42], and methotrexate [43] are also known to regulate p53 protein. Monastrol (targets *KIF11*) was also reported that there is a correlation with p53 protein. Furthermore, it is also known that the p53 protein interacts with other proteins such as *CDKN1A* [44], *GADD45A* [45]. And these proteins interact with other proteins such as *AURKA* [46], *PCNA* [47], *PLK1* [48], and *CCNB1* [49] which are also in DNA damage response pathway [16]. To demonstrate the results are related to the already known mechanism of action (MoA), We measured the ranks of the seven proteins (*TP53, CDKN1A, GADD45A, AURKA, PCNA, PLK1* and *CCNB1*) in the scores of drug target prediction for each drug. Since replicates of a drug's expression profiles, an average value of rank in every profile was used. The distributions of the ranks of the mechanism-related proteins are illustrated in Fig. 5**(A)**. Some of the proteins ranked high even in other methods, but we confirmed that our method ranked the mechanism-related proteins higher (median rank: 906) than previous methods (DE median rank: 11,658, DeMAND median rank: 2,894, ProTINA median rank: 8,196). All mechanism-related proteins ranked within 3,000 in our method while there are proteins ranked more than 15,000 in other previous works.

We also tested the high-scored proteins by comparing the results of the gene set enrichment test (GSEA) of drug-related KEGG pathways. We selected a drug (rapamycin, azathioprine, and SDF-1) for each dataset and gathered related pathways of each drug from KEGG. High-scored proteins in the top 5 percent in each dataset and method were selected for the gene set enrichment test, and the *p*-values of corresponding drug-related KEGG pathways are illustrated in Fig. 5**(B–D)**. Previous methods show higher *p*-values and even fail to detect drug-related KEGG pathways. However, our method detects all the related pathways with lower *p*-values. Our method shows the lowest *p*-values for all related KEGG pathways in the case of rapamycin (Fig. 5-**B**).

DeMAND shows slightly better results of *p*-values in *olfactory transduction and hematopoietic cell lineage* in azathioprine (Fig. 5-**C**) but does not capture *graft-versus-host disease and allograft rejection*. However, our method can find all azathioprine-related KEGG pathways with reliable *p*-values. Our method also captures all related pathways of SDF-1 (Fig. 5-**D**) with lower *p*-values than previous methods.

3.4. Suggestion of mechanism of action of drugs

We identified that highly scored proteins are associated with already-known mechanisms. In this analysis, we can suggest potential molecular MoA of drugs from drug targets to genes through a multilayer molecular network by using the top results of our method. This analysis is possible only in our method because of the use of a multilayer molecular network. We selected high-scored proteins, TFs and relations between them to suggest a MoA of a given drug from targets to expressed genes. We chose one drug for each dataset and filtered proteins and TFs ranked in the top 1 percent, respectively. Singleton proteins that are not connected to other proteins in the networks were omitted from the analysis. We also selected genes connected to TFs with high scores. There exist so many numbers of connected genes, so genes were also filtered by their log_2 fold-change values. Then we found literature evidence supported by direct experimental results between drug and represented proteins. Fig. 6 shows the suggested molecular mechanisms. In the case of rapamycin, a molecular mechanism through the proteins such as *PPARG, PRKCA,* and *PML* [50–52] and transcription factors such as *EGR1, TFAP2A, SP1* and *FOS* [53–56] are represented. In the case of azathioprine, proteins such as *MAPK8, ACKT1* and *EGLN3* [57–59] and transcription factors such as *SPI1, JUN* and *NFKB1* [60–62] are represented. In the case of SDF-1, proteins such as *GNA13, GNB2, GNG12, GNAI2, PTPN6* and *PTPN11* [63–67] are represented. And transcription factors such as *AHR, MYC, SMAD3* and *E2F1* [68–71] are represented.
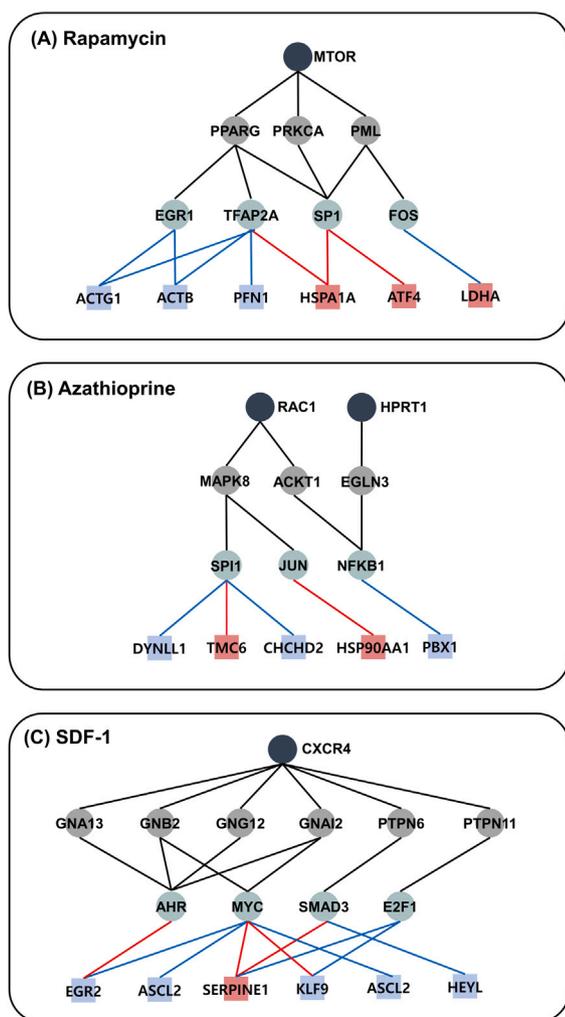
Fig. 6. Suggested molecular mechanism of actions of drugs. **(A)** is for the case of rapamycin from the dataset of NCI DREAM challenge, **(B)** is for the case of azathioprine from the dataset of genotoxicity study, and **(C)** is for the case of SDF-1 from the dataset of GPCR targeting study. Circle node means protein(including drug target and transcription factor), and rectangle node means gene with high log_2 fold-change. In circle nodes, protein targets are colored in navy, transcription factors are colored in green, and the others are in gray. In square nodes, highly up-regulated nodes are colored in red and highly down-regulated nodes are colored in blue. Colored lines between transcription factors and genes indicate the positive or negative value of the correlation coefficient between transcription factors and genes. The red means positive correlation and the blue means negative correlation.

3.5. Application to prediction of coronavirus targets

We applied our method to coronavirus (SARS-CoV-2) expression data obtained from the GSE182297 dataset to show an advantage in the case of predicting targets without solid structure information. It is hard to specify the exact structure of the virus because a virus is not a static chemical, and apply the previous work's method because it requires time-series data. We pre-processed data in the same way described in the Method section. We used thresholds from the top 10 to 300 in the protein score from the silver standard dataset [72] as it suggested 332 targets for the coronavirus. Then we calculated the AUROC value for each threshold count. The AUROC values depending on the threshold count are listed in Table 1. Although there was a slight difference depending on the threshold count, we could check that the performance was almost over 0.8 while differential expression analysis was almost under 0.5.

We could also find the related KEGG pathways. We performed a gene set enrichment test of high-scored proteins from our method and

differential expression analysis. We used three hundred proteins in the score for this analysis to match the number of the other work's suggestion. Top-5 KEGG pathway results from each method are listed in Table 2. In differential expression analysis, the most enriched pathway is viral protein interaction with cytokine and cytokine receptor, but the *p*-value was not significant because only proteins such as *CCL3*, *CCR5*, and *IL2* were included in the pathway. In our method, coronavirus disease ranked in top-3 and the most enriched pathway, ribosome, was also reported that the virus binds to the human 40S subunit and the non-translating 80S ribosome [73].

4. Discussion

We developed a method that can be applied to various datasets, effectively predict drug targets, and suggest the mechanism of actions of drugs. There have been many methods using structural information of chemicals and proteins. However, lack of data applicability and imbalance exist and even AlphaFold2 [74] predict structures well only for proteins with slight structural variation [75] but not for proteins with complex structures such as G-protein coupled receptor (GPCR) [76]. Considering that about one-third of FDA-approved drugs target GPCR [77,78], its use can be limited. Our method does not require structural information of drugs and proteins and can be used as long as gene expression is given. By using large-scale public databases and data from previous studies [27], we were able to construct tissue-specific networks. Using large-scale perturbation data, we also obtained correlation coefficients between TFs and genes. From the scoring function built from the above data, we were able to capture the more reliable molecular phenomenon of drug reaction in the body and have shown through performance comparisons in predicting already known drug targets in three different datasets. We also showed the capability of our method in suggesting drug's mechanism of action.

There have been several previous methods using gene transcriptional profiles and networks, but there were limitations such that they were not able to properly capture the phenomenon of drug response in network construction. Considering that the structure of GPCRs is unstable, a transcriptome-based approach seems appropriate to solve the problem. However, previous methods showed relatively low performance in the GPCR target study because GPCR regulates cellular processes at the top of the different signaling pathways and is far from the TFs in the PPI network. In the previous works, they tried to combine GRN and PPI but they simply stopped at building partners between proteins and genes only in the first neighbor. It is difficult to explain the process of a GPCR target that acts across multiple and serial proteins. In contrast, our network first separates TFs and proteins in the different layers and preserves the interactions between proteins to reflect more reliable biological phenomena. According to the network structure, our scoring method is also designed to consider the two features described above. First, our scoring method captures TFs which can explain drug-perturbed gene expression changes. Then, our scoring method finds which proteins can cover as many as and as close as the TFs in the PPI network. Our scoring method shows better performance in predicting known drug targets regardless of drug types (Fig. 3). Unlike previous studies in which the performance of GPCRs decreased, our method showed preserved performance. All factors in our scoring function are based on previous works, but we also tested the scoring function in different conditions (Fig. 4). For comparison, we first conducted random permutation and MLR to get correlation coefficients between proteins and genes. The regression models were generated as the same number of genes in the network, so they spent more computational resources and were more dependent on background data than our method (**Supplementary Figure 4, 5**). Scoring function using different neighboring distances also supports the importance of using deeper interactions between proteins.

We showed that our method not only outperformed in predicting drug targets in three independent GEO datasets but also suggested

Table 1

AUROC scores using the silver standard dataset of SARS-CoV-2 by threshold count in each method.

	Threshold count											
	10	20	30	40	50	60	70	80	90	100	200	300
Our method	0.855	0.910	0.934	0.856	0.863	0.856	0.860	0.828	0.816	0.825	0.810	0.761
DE	0.538	0.681	0.524	0.347	0.284	0.309	0.314	0.333	0.327	0.317	0.344	0.385

Table 2

Top-5 results from gene set enrichment test for KEGG pathways in each method.

	Order	KEGG Pathway	*P*-value	Adjusted *P*-value
Our Method	1	Ribosome	4.42e−08	1.79e−09
	2	Thyroid hormone signaling pathway	4.42e−08	4.25e−09
	3	Coronavirus disease	6.41e−06	2.54e−08
	4	Pathways in cancer	2.79e−05	1.06e−05
	5	Preteoglycans in cancer	1.24e−04	1.37e−05
	Order	KEGG Pathway	*P*-value	Adjusted *P*-value
DE	1	Viral protein interaction with cytokine and cytokine receptor	1.59e−1	9.95e−1
	2	Galactose metabolism	3.49e−1	9.95e−1
	3	Starch and sucrose metabolism	3.93e−1	9.95e−1
	4	Allograft rejection	4.09e−1	9.95e−1
	5	Chagas disease	4.10e−1	9.95e−1

the molecular mechanisms of action of drugs supported by literature evidence beyond simply providing insight. Because of the lack of interactions between proteins themselves, previous works have limitations in elucidating the mechanism of actions of drugs. However, our method can express interactions between proteins and is able to explain already known drug mechanisms as shown in Fig. 5. Our method has richer information of molecular interactions, so we can suggest the potential molecular mechanism of drugs as shown in Fig. 6.

Application to SARS-CoV-2 especially showed the advantages of our applicability even without structural information or time-series data. As far as we know, there is no *in-silico* based screening method for predicting targets of SARS-CoV-2 using the gene transcriptional profiles and network yet, so we have not further validated it. However, we confirmed the possibility of our method which showed quite a good performance in the AUROC score (Table 1) and finding related KEGG pathways (Table 2).

Like any other methods using any kind of network, the biological interactions including tissue-specific transcription and gene interactions are not complete. The database CMap gives a chance by offering numerous drug perturbation data, but it is still far from perfect. Our method also has the disadvantage of being dependent on these imperfect data. And the molecular interactions in our body have more dynamics and complexity between them than we tried to represent. Scoring function with a relatively simple format is also one of the limitations of our method. Current data quality seems hard to follow more elaborate methods. However, if more background data is supported, more sophisticated approaches such as using probabilistic models for getting explainable TFs and using the high-quality context-free network for network construction can elaborate our method to find targets of drugs. Nevertheless, we tried to derive a model to capture the biological phenomenon well within the solvable range and expect that the performance can be further improved as we accumulate knowledge about biological interactions. We expect that our method could help in the processes in drug screening and discovery by predicting targets with insights into the molecular mechanism.

CRediT authorship contribution statement

Gwangmin Kim: Conceptualization, Design of this study, Methodology, Implementation, Writing a manuscript. **Doheon Lee:** Conceptualization, Design of this study, Methodology, Writing a manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.compbiomed.2023.106881.

References

[1] David C. Swinney, Jason Anthony, How were new medicines discovered? Nat. Rev. Drug Discov. 10 (7) (2011) 507–519.

[2] Sayaka Mizutani, Edouard Pauwels, Véronique Stoven, Susumu Goto, Yoshihiro Yamanishi, Relating drug–protein interaction network with drug side effects, Bioinformatics 28 (18) (2012) i522–i528.

[3] M. Rami Reddy, Abby L. Parrill, Overview of Rational Drug Design, ACS Publications, 1999.

[4] Ellin Berman, Maria Nicolaides, Robert G. Maki, Martin Fleisher, Suzanne Chanel, Kelly Scheu, Bri-Anne Wilson, Glenn Heller, Nicholas P. Sauter, Altered bone and mineral metabolism in patients receiving imatinib mesylate, N. Engl. J. Med. 354 (19) (2006) 2006–2013.

[5] Kate Vandyke, Stephen Fitter, Andrea L. Dewar, Timothy P. Hughes, Andrew C.W. Zannettino, Dysregulation of bone remodeling by imatinib mesylate, Blood, J. Am. Soc. Hematol. 115 (4) (2010) 766–774.

[6] Michael Williams, Target validation, Curr. Opin. Pharmacol. 3 (5) (2003) 571–577.

[7] Steven M. Paul, Daniel S. Mytelka, Christopher T. Dunwiddie, Charles C. Persinger, Bernard H. Munos, Stacy R. Lindborg, Aaron L. Schacht, How to improve R&D productivity: the pharmaceutical industry's grand challenge, Nat. Rev. Drug Discov. 9 (3) (2010) 203–214.

[8] Honglin Li, Zhenting Gao, Ling Kang, Hailei Zhang, Kun Yang, Kunqian Yu, Xiaomin Luo, Weiliang Zhu, Kaixian Chen, Jianhua Shen, et al., TarFisDock: a web server for identifying drug targets with docking approach, Nucleic Acids Res. 34 (suppl_2) (2006) W219–W224.

[9] Hui Liu, Jianjiang Sun, Jihong Guan, Jie Zheng, Shuigeng Zhou, Improving compound–protein interaction prediction by building up highly credible negative samples, Bioinformatics 31 (12) (2015) i221–i229.

[10] Kyle Yingkai Gao, Achille Fokoue, Heng Luo, Arun Iyengar, Sanjoy Dey, Ping Zhang, et al., Interpretable drug target prediction using deep neural representation, in: IJCAI, 2018, pp. 3371–3377.

[11] Aravind Subramanian, Rajiv Narayan, Steven M. Corsello, David D. Peck, Ted E. Natoli, Xiaodong Lu, Joshua Gould, John F. Davis, Andrew A. Tubelli, Jacob K. Asiedu, et al., A next generation connectivity map: L1000 platform and the first 1,000,000 profiles, Cell 171 (6) (2017) 1437–1452.

[12] Mahmoud Ghandi, Franklin W. Huang, Judit Jané-Valbuena, Gregory V. Kryukov, Christopher C. Lo, E. Robert McDonald, Jordi Barretina, Ellen T. Gelfand, Craig M. Bielski, Haoxin Li, et al., Next-generation characterization of the cancer cell line encyclopedia, Nature 569 (7757) (2019) 503–508.

[13] Wanjuan Yang, Jorge Soares, Patricia Greninger, Elena J. Edelman, Howard Lightfoot, Simon Forbes, Nidhi Bindal, Dave Beare, James A. Smith, I. Richard Thompson, et al., Genomics of drug sensitivity in cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells, Nucleic Acids Res. 41 (D1) (2012) D955–D961.

[14] Francesco Iorio, Timothy Rittman, Hong Ge, Michael Menden, Julio Saez-Rodriguez, Transcriptional data: a new gateway to drug repositioning? Drug Discov. Today 18 (7–8) (2013) 350–357.

[15] Chanin Nantasenamat, Chartchalerm Isarankura-Na-Ayudhya, Virapong Prachay-asittikul, Advances in computational methods to predict the biological activity of compounds, Expert Opin. Drug Discovery 5 (7) (2010) 633–654.

[16] Jung Hoon Woo, Yishai Shimoni, Wan Seok Yang, Prem Subramaniam, Archana Iyer, Paola Nicoletti, María Rodríguez Martínez, Gonzalo López, Michela Mattioli, Ronald Realubit, et al., Elucidating compound mechanism of action by network perturbation analysis, Cell 162 (2) (2015) 441–451.

[17] Heeju Noh, Jason E. Shoemaker, Rudiyanto Gunawan, Network perturbation analysis of gene transcriptional profiles reveals protein targets and mechanism of action of drugs and influenza A viral infection, Nucleic Acids Res. 46 (6) (2018) e34.

[18] Mukesh Bansal, Jichen Yang, Charles Karan, Michael P. Menden, James C. Costello, Hao Tang, Guanghua Xiao, Yajuan Li, Jeffrey Allen, Rui Zhong, et al., A community computational challenge to predict the activity of pairs of compounds, Nature Biotechnol. 32 (12) (2014) 1213–1222.

[19] C. Magkoufopoulou, S.M.H. Claessen, M. Tsamou, D.G.J. Jennen, J.C.S. Kleinjans, J.H.M. van Delft, A transcriptomics-based in vitro assay for predicting chemical genotoxicity in vivo, Carcinogenesis 33 (7) (2012) 1421–1429.

[20] Doreen Heckmann, Patrick Maier, Stephanie Laufs, Li Li, Jonathan P. Sleeman, Marcus J. Trunk, Jörg H. Leupold, Frederik Wenz, W. Jens Zeller, Stefan Fruehauf, et al., The disparate twins: A comparative study of CXCR4 and CXCR7 in SDF-1α–Induced gene expression, invasion and chemosensitivity of colon CancerThe disparate twins: CXCR4 and CXCR7, Clin. Cancer Res. 20 (3) (2014) 604–616.

[21] Tanya Barrett, Dennis B. Troup, Stephen E. Wilhite, Pierre Ledoux, Carlos Evangelista, Irene F. Kim, Maxim Tomashevsky, Kimberly A. Marshall, Katherine H. Phillippy, Patti M. Sherman, et al., NCBI GEO: archive for functional genomics data sets—10 years on, Nucleic Acids Res. 39 (suppl_1) (2010) D1005–D1010.

[22] Richard W.W. Jones Casey S. Greene, et al., refine.bio: a resource of uniformly processed publicly available gene expression datasets, URL https://www.refine.bio.

[23] Broad Institute. Connectivity Map. 2020. URL https://clue.io.

[24] Rose Oughtred, Jennifer Rust, Christie Chang, Bobby-Joe Breitkreutz, Chris Stark, Andrew Willems, Lorrie Boucher, Genie Leung, Nadine Kolas, Frederick Zhang, et al., The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions, Prot. Sci. 30 (1) (2021) 187–200.

[25] Minoru Kanehisa, Susumu Goto, KEGG: kyoto encyclopedia of genes and genomes, Nucleic Acids Res. 28 (1) (2000) 27–30.

[26] Jürgen Dönitz, Björn Goemann, Muriel Lizé, Holger Michael, Nicole Sasse, Edgar Wingender, Anatolij P. Potapov, EndoNet: an information resource about regulatory networks of cell-to-cell communication, Nucleic Acids Res. 36 (suppl_1) (2007) D689–D694.

[27] Abhijeet Rajendra Sonawane, John Platig, Maud Fagny, Cho-Yi Chen, Joseph Nathaniel Paulson, Camila Miranda Lopes-Ramos, Dawn Lisa DeMeo, John Quackenbush, Kimberly Glass, Marieke Lydia Kuijjer, Understanding tissue-specific gene regulation, Cell Rep. 21 (4) (2017) 1077–1088.

[28] Andreas Wagner, Genes regulated cooperatively by one or more transcription factors and their identification in whole eukaryotic genomes, Bioinformatics (Oxford, England) 15 (10) (1999) 776–784.

[29] Lavisha Parab, Sampriti Pal, Riddhiman Dhar, Transcription factor binding process is the primary driver of noise in gene expression, PLoS Genet. 18 (12) (2022) e1010535.

[30] Adam B. Zaborowski, Dirk Walther, Determinants of correlated expression of transcription factors and their target genes, Nucleic Acids Res. 48 (20) (2020) 11347–11369.

[31] Emre Guney, Jörg Menche, Marc Vidal, Albert-László Barábasi, Network-based in silico drug efficacy screening, Nature Commun. 7 (1) (2016) 10331.

[32] David S. Wishart, Yannick D. Feunang, An C. Guo, Elvis J. Lo, Ana Marcu, Jason R. Grant, Tanvir Sajed, Daniel Johnson, Carin Li, Zinat Sayeeda, et al., DrugBank 5.0: a major update to the DrugBank database for 2018, Nucleic Acids Res. 46 (D1) (2018) D1074–D1082.

[33] Christopher T. Veldkamp, Christoph Seibert, Francis C. Peterson, Norberto B. De la Cruz, John C. Haugner III, Harihar Basnet, Thomas P. Sakmar, Brian F. Volkman, Structural basis of CXCR4 sulfotyrosine recognition by the chemokine SDF-1/CXCL12, Sci. Signaling 1 (37) (2008) ra4.

[34] Ting Ting Lau, Dong-An Wang, Stromal cell-derived factor-1 (SDF-1): homing factor for engineered regenerative medicine, Expert Opin. Biol. Therapy 11 (2) (2011) 189–197.

[35] Konstantinos Stellos, Harald Langer, Karin Daub, Tanja Schoenberger, Alexandra Gauss, Tobias Geisler, Boris Bigalke, Iris Mueller, Michael Schumm, Iris Schaefer, et al., Platelet-derived stromal cell–derived factor-1 regulates adhesion and promotes differentiation of human CD34+ cells to endothelial progenitor cells, Circulation 117 (2) (2008) 206–215.

[36] Yigang Wang, Kristin Luther, Genetically manipulated progenitor/stem cells restore function to the infarcted heart via the SDF-1α/CXCR4 signaling pathway, Prog. Mol. Biol. Transl. Sci. 111 (2012) 265–284.

[37] Ru-Wei Lin, Cheng-Jung Ho, Hsin-Wen Chen, Yu-Hsuan Pao, Li-En Chen, Min-Chi Yang, Shih-Bo Huang, Shiaw Wang, Chung-Hwan Chen, Chihuei Wang, P53 enhances apoptosis induced by doxorubicin only under conditions of severe DNA damage, Cell Cycle 17 (17) (2018) 2175–2186.

[38] E. Rudolf, K. Rudolf, M. Cervinka, Camptothecin induces p53-dependent and-independent apoptogenic signaling in melanoma cells, Apoptosis 16 (11) (2011) 1165–1176.

[39] Hyuk-Kwon Kwon, Hyeon-Jun Shin, Jae-Hyeok Lee, Seol-Hee Park, Min-Cheol Kwon, Suresh Panneerselvam, Chan Gyu Lee, Sang Geon Kim, Jae-Ho Kim, Sangdun Choi, Etoposide induces necrosis through p53-mediated antiapoptosis in human kidney proximal tubule cells, Toxicol. Sci. 148 (1) (2015) 204–219.

[40] Danrui Cui, Ruirui Qu, Dian Liu, Xiufang Xiong, Tingbo Liang, Yongchao Zhao, The cross talk between p53 and mTOR pathways in response to physiological and genotoxic stresses, Front. Cell Dev. Biol. (2021) 3320.

[41] Z.H. Li, Y.J. Zhu, L.S. Chao, X.T. Li, Wild-type p53 stimulates vincristine-induced apoptosis, Yao Xue Xue Bao=Acta Pharmaceutica Sinica 32 (8) (1997) 565–568.

[42] Limeng Dai, Gang He, Kun Zhang, Xingying Guan, Yan Wang, Bo Zhang, Trichostatin a induces p53-dependent endoplasmic reticulum stress in human colon cancer cells, Oncol. Lett. 17 (1) (2019) 660–667.

[43] Wen-Yu Huang, Pei-Ming Yang, Yu-Fan Chang, Victor E. Marquez, Ching-Chow Chen, Methotrexate induces apoptosis through p53/p21-dependent pathway and increases E-cadherin expression through downregulation of HDAC/EZH2, Biochem. Pharmacol. 81 (4) (2011) 510–517.

[44] Ornella Cazzalini, A. Ivana Scovassi, Monica Savio, Lucia A. Stivala, Ennio Prosperi, Multiple roles of the cell cycle inhibitor p21CDKN1A in the DNA damage response, Mutat. Res./Rev. Mutat. Res. 704 (1–3) (2010) 12–20.

[45] Qimin Zhan, Gadd45a, a p53-and BRCA1-regulated stress protein, in cellular response to DNA damage, Mutat. Res./Fund. Mol. Mech. Mutagenesis 569 (1–2) (2005) 133–143.

[46] Shujuan Shao, Yang Wang, Shunqian Jin, Yongmei Song, Xiaoxia Wang, Wenhong Fan, Zhiying Zhao, Ming Fu, Tong Tong, Lijia Dong, et al., Gadd45a interacts with aurora-A and inhibits its kinase activity, J. Biol. Chem. 281 (39) (2006) 28943–28950.

[47] Zvi Kelman, PCNA: structure, functions and interactions, Oncogene 14 (6) (1997) 629–640.

[48] Libor Macurek, Arne Lindqvist, Dan Lim, Michael A. Lampson, Rob Klompmaker, Raimundo Freire, Christophe Clouin, Stephen S. Taylor, Michael B. Yaffe, René H. Medema, Polo-like kinase-1 is activated by aurora a to promote checkpoint recovery, Nature 455 (7209) (2008) 119–123.

[49] Fumiko Toyoshima-Morimoto, Eri Taniguchi, Nobuko Shinya, Akihiro Iwamatsu, Eisuke Nishida, Polo-like kinase 1 phosphorylates cyclin B1 and targets it to the nucleus during prophase, Nature 410 (6825) (2001) 215–220.

[50] Mulki Angela, Yusuke Endo, Hikari K. Asou, Takeshi Yamamoto, Damon J. Tumes, Hirotake Tokuyama, Koutaro Yokote, Toshinori Nakayama, Fatty acid metabolic reprogramming via mTOR-mediated inductions of PPARγ directs early activation of T cells, Nature Commun. 7 (1) (2016) 1–15.

[51] Akio Iwanami, Beatrice Gini, Ciro Zanca, Tomoo Matsutani, Alvaro Assuncao, Ali Nael, Julie Dang, Huijun Yang, Shaojun Zhu, Jun Kohyama, et al., PML mediates glioblastoma resistance to mammalian target of rapamycin (mTOR)-targeted therapies, Proc. Natl. Acad. Sci. 110 (11) (2013) 4339–4344.

[52] Miguel Martin-Perez, Anthony S. Grillo, Takashi K. Ito, Anthony S. Valente, Jeehae Han, Samuel W. Entwisle, Heather Z. Huang, Dayae Kim, Masanao Yajima, Matt Kaeberlein, et al., PKC downregulation upon rapamycin treatment attenuates mitochondrial disease, Nat. Metabolism 2 (12) (2020) 1472–1481.

[53] David A. Fruman, mTOR signaling: new networks for ALL, Blood, J. Am. Soc. Hematol. 127 (22) (2016) 2658–2659.

[54] D. Vizza, A. Perri, G. Toteda, S. Lupinacci, I. Perrotta, D. Lofaro, F. Leone, P. Gigliotti, A. La Russa, R. Bonofiglio, Rapamycin-induced autophagy protects proximal tubular renal cells against proteinuric damage through the transcriptional activation of the nerve growth factor receptor NGFR, Autophagy 14 (6) (2018) 1028–1042.

[55] Aristotelis Astrinidis, Jiyoon Kim, Crystal M. Kelly, Beatrix A. Olofsson, Behzad Torabi, Elena M. Sorokina, Jane Azizkhan-Clifford, The transcription factor SP1 regulates centriole function and chromosomal stability through a functional interaction with the mammalian target of rapamycin/raptor complex, Genes Chromosom. Cancer 49 (3) (2010) 282–297.

[56] Juan-juan Sun, Xiao-wei Yin, Hui-hui Liu, Wen-xiu Du, Lu-yao Shi, Ya-bo Huang, Fen Wang, Chun-feng Liu, Yong-jun Cao, Yan-lin Zhang, Rapamycin inhibits ox-LDL-induced inflammation in human endothelial cells in vitro by inhibiting the mTORC2/PKC/c-Fos pathway, Acta Pharmacol. Sin. 39 (3) (2018) 336–344.

[57] Young-Eun Cho, Pyong-Gon Moon, Moon-Chang Baek, An integrated proteomic and transcriptomic approach to understanding azathioprine-induced hepatotoxicity in rat primary hepatocytes, Electrophoresis 35 (6) (2014) 911–922.

[58] Alessio Reggio, Filomena Spada, Marco Rosina, Giorgia Massacci, Alessandro Zuccotti, Claudia Fuoco, Cesare Gargioli, Luisa Castagnoli, Gianni Cesareni, The immunosuppressant drug azathioprine restrains adipogenesis of muscle Fibro/Adipogenic Progenitors from dystrophic mice by affecting AKT signaling, Sci. Rep. 9 (1) (2019) 1–13.

[59] Jian Fu, Mark B. Taubman, Prolyl hydroxylase EGLN3 regulates skeletal myoblast differentiation through an NF-κB-dependent pathway, J. Biol. Chem. 285 (12) (2010) 8927–8935.

[60] Jan K. Nowak, Alex T. Adams, Rahul Kalla, Jonas C. Lindstrøm, Simen Vatn, Daniel Bergemalm, Åsa V. Keita, Fernando Gomollón, Jørgen Jahnsen, Morten H. Vatn, et al., Characterisation of the circulating transcriptomic landscape in inflammatory bowel disease provides evidence for dysregulation of multiple transcription factors including NFE2, SPI1, CEBPB, and IRF2, J. Crohn's Colitis 16 (8) (2022) 1255–1268.

[61] Goran Marinković, Stijntje Hibender, Mark Hoogenboezem, Amber van Broekhoven, Arginell F. Girigorie, Natascha Bleeker, Anouk A.J. Hamers, Jan Stap, Jaap D. van Buul, Carlie J.M. de Vries, et al., Immunosuppressive drug azathioprine reduces aneurysm progression through inhibition of Rac1 and c-Jun-terminal-N-kinase in endothelial cells, Arterioscler. Thromb. Vasc. Biol. 33 (10) (2013) 2380–2388.

[62] Terezinha M. Souza, Jos C.S. Kleinjans, Danyel G.J. Jennen, Dose and time dependencies in stress pathway responses during chemical exposure: novel insights from gene regulatory networks, Front. Genet. 8 (2017) 142.

[63] Kisha A. Scarlett, El-Shaddai Z. White, Christopher J. Coke, Jada R. Carter, Latoya K. Bryant, Cimona V. Hinton, Agonist-induced CXCR4 and CB2 heterodimerization inhibits Gα13/RhoA-mediated migration, Mol. Cancer Res. 16 (4) (2018) 728–739.

[64] Jane D. Holland, Marina Kochetkova, Chareeporn Akekawatchai, Mara Dottore, Angel Lopez, Shaun R. McColl, Differential functional activation of chemokine receptor CXCR4 is mediated by G proteins in breast cancer cells, Cancer Res. 66 (8) (2006) 4117–4124.

[65] Lei Dong, Wen-Mei Yu, Hong Zheng, Mignon L. Loh, Silvia T. Bunting, Melinda Pauly, Gang Huang, Muxiang Zhou, Hal E. Broxmeyer, David T. Scadden, et al., Leukaemogenic effects of Ptpn11 activating mutations in the stem cell microenvironment, Nature 539 (7628) (2016) 304–308.

[66] Skye C. McIver, Kate L. Loveland, Shaun Daryl Roman, Brett Nixon, Riko Kitazawa, Eileen Anne McLaughlin, The chemokine CXCL 12 and its receptor CXCR 4 are implicated in human seminoma metastasis, Andrology 1 (3) (2013) 517–529.

[67] Zubair A. Karim, Fatima Z. Alshbool, Hari Priya Vemana, Christine Conlon, Kirk M. Druey, Fadi T. Khasawneh, CXCL12 regulates platelet activation via the regulator of G-protein signaling 16, Biochim. Biophys. Acta 1863 (2) (2016) 314–321.

[68] A. Dubrovska, A. Hartung, L.C. Bouchez, J.R. Walker, VA Reddy, C.Y. Cho, P.G. Schultz, CXCR4 activation maintains a stem cell population in tamoxifen-resistant breast cancer cells through AhR signalling, Br. J. Cancer 107 (1) (2012) 43–52.

[69] Achinto Saha, Songyeon Ahn, Jorge Blando, Fei Su, Mikhail G. Kolonin, John DiGiovanni, Proinflammatory CXCL12–CXCR4/CXCR7 signaling axis drives myc-induced prostate cancer in obese MiceCXCL12–CXCR4/CXCR7 signaling and prostate cancer progression, Cancer Res. 77 (18) (2017) 5158–5168.

[70] Chien-Huang Lin, Chung-Huang Shih, Yu-Chang Lin, You-Lan Yang, Bing-Chang Chen, MEKK1, JNK, and SMAD3 mediate CXCL12-stimulated connective tissue growth factor expression in human lung fibroblasts, J. Biomed. Sci. 25 (1) (2018) 1–11.

[71] Muhammad Zafrullah Khan, Renato Brandimarti, Brian Joseph Musser, Danielle Marie Resue, Alessandro Fatatis, Olimpia Meucci, The chemokine receptor CXCR4 regulates cell-cycle proteins in neurons, J. Neurovirol. 9 (3) (2003) 300–314.

[72] David E Gordon, Gwendolyn M Jang, Mehdi Bouhaddou, Jiewei Xu, Kirsten Obernier, Kris M White, Matthew J O'Meara, Veronica V Rezelj, Jeffrey Z Guo, Danielle L Swaney, et al., A sars-cov-2 protein interaction map reveals targets for drug repurposing, Nature 583 (7816) (2020) 459–468.

[73] Katharina Schubert, Evangelos D. Karousis, Ahmad Jomaa, Alain Scaiola, Blanca Echeverria, Lukas-Adrian Gurzeler, Marc Leibundgut, Volker Thiel, Oliver Mühlemann, Nenad Ban, SARS-CoV-2 Nsp1 binds the ribosomal mRNA channel to inhibit translation, Nat. Struct. Mol. Biol. 27 (10) (2020) 959–966.

[74] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al., Highly accurate protein structure prediction with AlphaFold, Nature 596 (7873) (2021) 583–589.

[75] Lim Heo, Michael Feig, Multi-state modeling of G-protein coupled receptors at experimental accuracy, Proteins: Struct. Funct. Bioinform. (2022).

[76] Xin-heng He, Chong-zhao You, Hua-liang Jiang, Yi Jiang, H. Eric Xu, Xi Cheng, AlphaFold2 versus experimental structures: evaluation on G protein-coupled receptors, Acta Pharmacol. Sin. (2022) 1–7.

[77] Alexander S. Hauser, Misty M. Attwood, Mathias Rask-Andersen, Helgi B. Schiöth, David E. Gloriam, Trends in GPCR drug discovery: new agents, targets and indications, Nat. Rev. Drug Discov. 16 (12) (2017) 829–842.

[78] Tudor I. Oprea, Cristian G. Bologa, Søren Brunak, Allen Campbell, Gregory N. Gan, Anna Gaulton, Shawn M. Gomez, Rajarshi Guha, Anne Hersey, Jayme Holmes, et al., Unexplored therapeutic opportunities in the human genome, Nat. Rev. Drug Discov. 17 (5) (2018) 317–332.